1 **An angiosperm *NLR* atlas reveals that *NLR* gene reduction is associated with**
2 **ecological specialization and signal transduction component deletion**

3
4 **Short title:** Drivers of *NLR* contraction

5
6 Yang Liu¹, Zhen Zeng¹, Qian Li¹, Xing-Mei Jiang¹, Zhen Jiang¹, Ji-Hong Tang¹,
7 Yan-Mei Zhang², Dijun Chen¹, Qiang Wang¹, Jian-Qun Chen^{1,*}, Zhu-Qing Shao^{1,*}

8
9 ¹ State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences,
10 Nanjing University, Nanjing, 210023, China

11 ² Jiangsu Key Laboratory for the Research and Utilization of Plant Resources,
12 Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing,
13 210014, China

14
15 * Corresponding authors:

16 zhuqingshao@nju.edu.cn (Z.Q.S.) and chenjq@nju.edu.cn (J.Q.C.)

Abstract

Nucleotide-binding site-leucine-rich repeat receptor (*NLR*) genes comprise the largest family of plant disease resistance genes. *NLR* genes are phylogenetically divided into the *TNL*, *CNL*, and *RNL* subclasses. *NLR* copy numbers and subclass composition vary tremendously across angiosperm genomes. However, the evolutionary associations between genomic *NLR* content and plant lifestyle, or between *NLR* content and signal transduction components, are poorly characterized due to limited genome availability. Here, we established an angiosperm *NLR* atlas (ANNA, <http://compbio.nju.edu.cn/app/ANNA/>), which includes *NLR* genes from over 300 angiosperm genomes. Using ANNA, we revealed that *NLR* copy numbers differ up to 66-fold among closely related species due to rapid gene loss and gain. Interestingly, *NLR* contraction was associated with adaptations to aquatic, parasitic, and carnivorous lifestyles. The convergent *NLR* reduction in aquatic plants resembles the long-term evolutionary silence of *NLR* genes in green algae before the colonization of land. A co-evolutionary pattern between *NLR* subclasses and plant immune-pathway components was also identified, suggesting that immune pathway deficiencies may drive *TNL* loss. Finally, we recovered a conserved *TNL* lineage that may function independently of the *RNL* pathway. Our findings provide new insights into the evolution of *NLR* genes in the context of plant lifestyles and genome content variation.

Keywords: *NLR* genes, plant disease resistance, *EDSI* gene family, gene family evolution, angiosperms

1 Introduction

2 Plants rely on a two-layered immune system to defend against various pathogens: the
 3 first immune-system layer, termed pathogen-associated molecular pattern
 4 (PAMP)-triggered immunity (PTI), recognizes PAMPs using cell surface-localized
 5 receptors; the second immune-system layer, known as effector-triggered immunity
 6 (ETI), detects effectors released by pathogens using intracellular disease resistance
 7 genes (*R* genes) (Wang et al., 2020). Several different types of *R* genes have been
 8 identified over the past twenty years (Kourelis and van der Hoorn, 2018). The largest
 9 *R* gene family is comprised of genes encoding nucleotide-binding site (NBS) and
 10 leucine-rich repeat (LRR) domain receptors, which are known as *NLR* or *NBS-LRR*
 11 genes (Kourelis and van der Hoorn, 2018). The plant-specific *NLR* gene family
 12 originated and diverged in the common ancestor of green plants (Shao et al., 2019).
 13 Three *NLR* gene subclasses, *TIR-NBS-LRR* (*TNL*), *CC-NBS-LRR* (*CNL*), and
 14 *RPW8-NBS-LRR* (*RNL*), have been characterized based on the N-terminal domains of
 15 the encoded NLR proteins: the Toll/Interleukin-1 receptor (TIR) domain, the
 16 coiled-coil (CC) domain, and the resistance to powdery mildew8 (RPW8) domain,
 17 respectively (Pan et al., 2000; Parker et al., 1997; Shao et al., 2014). Most CNL and
 18 TNL proteins function as pathogen detectors, either directly interacting with pathogen
 19 effectors or monitoring the state alteration of host proteins targeted by the effectors
 20 (Kourelis and van der Hoorn, 2018). The proteins encoded by the *RNL* genes are
 21 “helper” NLRs, which are involved in the downstream signal transduction of CNL
 22 and TNL proteins (Kourelis and van der Hoorn, 2018; Wang et al., 2020).

23 Over much of evolutionary time, *NLR* genes have been silent and have indeed been
 24 frequently lost in the green algae (Shao et al., 2019). However, after land colonization,
 25 the *NLR* genes expanded rapidly in green-plant genomes: angiosperm genomes often
 26 possess dozens to hundreds of *NLR* genes, most of which belong to the *TNL* and *CNL*
 27 subfamilies (Shao et al., 2016). Numbers of *NLR* genes may vary widely among
 28 genomes within the same family: two- to six-fold differences in gene number have
 29 frequently been reported (Luo et al., 2012; Qian et al., 2017; Shao et al., 2014; Tirnaz
 30 et al., 2020; Zhang et al., 2016). Indeed, a recent study revealed interspecies

1 differences in *NLR* gene number as high as 20-fold among species in the Orchidaceae
2 (Xue et al., 2020). This suggests that dramatic *NLR* gain or loss might occur rapidly
3 after speciation. Pathogenic selection pressure may be the primary driving force
4 underlying these *NLR* gene expansions. However, possible associations between *NLR*
5 gene number and adaptations to specific lifestyles remain largely unexplored,
6 primarily because few angiosperm genomes have been sequenced.

7 The absence of *TNL* genes from monocot genomes and several dicot genomes is
8 another long-standing puzzle in *NLR* gene evolution (Bai et al., 2002; Collier et al.,
9 2011). Several recent studies investigated the involvement of proteins from the two
10 RNL lineages (*ADRI* and *NRG1*) in TNL-triggered immunity. These studies revealed
11 that nearly all tested TNL proteins were functionally dependent on RNL proteins,
12 predominantly from the *NRG1* lineage, to confer full resistance against pathogens
13 (Castel et al., 2019; Qi et al., 2018; Saile et al., 2020; Wu et al., 2019). Moreover, a
14 pattern of co-absence was observed for *TNL* and *NRG1* genes in one monocot and two
15 dicot species (Collier et al., 2011). The functional and evolutionary relationships
16 among *TNL* and *RNL* genes provide intriguing clues concerning the mechanism of
17 *TNL* loss.

18 In addition to RNL proteins, other key signaling elements from the ENHANCED
19 DISEASE SUSCEPTIBILITY (*EDS1*) family are required for *TNL*-triggered disease
20 resistance. The *EDS1* family includes *EDS1*, PHYTOALEXIN DEFICIENT 4
21 (*PAD4*), and SENESCENCE ASSOCIATED GENE 101 (*SAG101*); in *Arabidopsis*
22 *thaliana*, *EDS1* forms heterodimers with *SAG101* and with *PAD4* (Lapin et al., 2020).
23 When pathogens invade and TNL proteins are activated, the resulting signaling
24 involves the enzymatic catalysis of NAD^+ , which is transmitted downward to *EDS1*
25 via an unknown mechanism. *EDS1* then further transmits the signal to downstream
26 RNL proteins to induce local cell necrosis and transcriptional responses (Lapin et al.,
27 2019; Wan et al., 2019).

28 Interestingly, several plant species have been shown to lack both *SAG101* and *TNL*
29 (Lapin et al., 2019). The concomitant loss of the *TNL* gene and its signal transduction
30 partners suggests that genomic context alteration may lead to variations in *NLR* gene

1 copy number. However, this pattern of *TNL* and *NRG1/SAG101* co-absence has been
2 observed in only a few angiosperm genomes, corresponding to three independent
3 losses (Collier et al., 2011; Lapin et al., 2019; Shao et al., 2016). It remains unclear
4 how many other angiosperm genomes lack *TNL* genes, and whether these losses are
5 related to similar genome content alterations.

6 Due to recent advances in sequencing technologies, hundreds of genomes,
7 representing many of the major angiosperm clades, have been sequenced. This
8 provides a unique opportunity to explore *NLR*-associated questions using a large
9 dataset. Thus, in this study, we establish an angiosperm *NLR* atlas (ANNA) by
10 identifying and incorporating over 90,000 *NLR* genes from more than 300 angiosperm
11 genomes. Using this huge dataset, we determined that plants that adapted to aquatic,
12 parasitic, and carnivorous lifestyles tended to lose *NLR* genes. In addition, numbers of
13 *TNL* and *RNL* genes were highly correlated within a genome, and *TNL* loss tended to
14 coincide with the loss of the *SAG101* gene, as well as genes in the *RNL-NRG1* lineage.
15 This suggested that *NRG1/SAG101* pathway deficiency may potentially drive *TNL*
16 loss. These findings provide new insights into genome context variation and the
17 adaptive evolution of disease resistance genes to plant lifestyles.

18 **Results**

19 **Construction of the ANNA, including over 300 genomes**

20 We identified 91,291 *NLR* genes in 305 publicly available angiosperm genomes,
21 including four basal angiosperms, 65 monocots, 230 dicots, five magnoliids, and one
22 Ceratophyllales (Table S1, S2). These species were distributed across the major
23 angiosperm clades, representing 85 families in 37 of the 64 angiosperm orders (Figure
24 1A). The 65 monocot genomes represented nine families in six orders, providing a
25 large and diverse group of taxa with which to test the hypothesis of ancestral loss of
26 *TNL* genes in this clade. The 230 dicot taxa comprised seven species representing two
27 basal lineages (Proteales and Ranunculales); 74 superasterid species, representing 32
28 families in 11 orders, including 21 species from the basal lineages Caryophyllales,
29 Santalales, and Ericales; and 149 superrosid species, representing 32 families in 12
30 orders, including four species from two basal superrosid lineages (Saxifragales and

1 Vitales).

Figure 1. Construction of an angiosperm NLR atlas (ANNA) covering 91,291 NLR genes in 304 angiosperm genomes. A) Phylogenetic relationships among angiosperm orders, constructed based on the APG IV system of flowering plant classification (Byng et al., 2016). Circle size corresponds to the number of species in each order; circle segments reflect the distribution of species among families. B) Representative screenshots of the ANNA interface, showing the major functions of ANNA: browsing, downloading, and primary comparative analysis of NLR datasets of interest.

2
3

4 We analyzed these genomes to clarify whether *TNL* gene loss occurs rarely or
5 frequently in dicot genomes. Notably, we identified *NLR* genes in five magnoliids, a
6 major angiosperm clade that has to date been poorly investigated due to a lack of
7 genomic data. Exploration of these data might reveal whether the patterns of *NLR*
8 evolution in this clade differ from those recovered in the monocots and dicots. The
9 91,291 *NLR* genes were unevenly distributed across the 305 genomes examined: the
10 number of *NLR* genes per genome ranged from 5 to more than 2000, except for one

1 genome (*Utricularia gibba*) that contained no detectable *NLR* genes. *U. gibba*
2 represents the only known land plant genome completely lacking *NLR* genes.

3 All identified *NLR* genes were combined to form ANNA, a publicly available free
4 website that allows users to browse, download, and perform primary comparative
5 analyses of *NLR* datasets of interest (Figure 1B). In total, the ANNA dataset is
6 comprised of 3.6 species per family; 16 families are represented by more than five
7 genomes. ANNA thus provides a uniquely large dataset for use by the *R* gene research
8 community, allowing users to explore various aspects of *NLR* gene evolution. The
9 *NLR* data in ANNA were used for all subsequent analyses in this study.

10 Dramatic variations in *NLR* gene numbers among angiosperms are associated **11 with adoptions to different lifestyles**

12 Evaluation of the 305 angiosperm genomes showed that *NLR* copy number varied
13 tremendously among genomes, even within the same clade (Figure 2A). Ten genomes
14 included more than 1000 *NLR* genes each (Table S2): three species in one monocot
15 order (Poales) and seven species in five dicot orders (Gentianales, one species;
16 Rosales, two species; Malpighiales, one species; Fagales, two species; and Myrtales,
17 one species). This suggested several instances of lineage- or species-specific gene
18 expansion. Consistent with this, CD-hit analysis (Huang et al., 2010) showed a high
19 degree of sequence similarity among many of the *NLR* genes in these ten genomes
20 (Figure S1), suggesting intense recent gene duplications. By contrast, we identified 30
21 genomes that possessed fewer than 50 *NLR* genes each, distributed among 21 families
22 in 17 orders. CD-hit analysis showed that few of these genomes contained large gene
23 clusters with high sequence similarity, indicating rare recent gene duplications (Figure
24 S1).

Figure 2. Variations in *NLR* gene number among angiosperm genomes. A) Phylogenetic relationships among the 305 angiosperm genomes included in this study, constructed following the APG IV system (Byng et al., 2016). The numbers of *NLR* genes in each genome are shown by the blue bars. B) Maximum fold-differences in *NLR* gene copy number within each angiosperm family. C) Boxplots showing *NLR* gene numbers in the genomes of the basal angiosperms (n=4), magnoliids (n=5), monocots (n=65), dicots (n=230), Ceratophyllales (n=1). Gene numbers did not differ significantly among groups. D) Boxplots showing *NLR* gene numbers in the genomes of special lifestyle species (SL species, n=17) and non-special lifestyle species (NSL species, n=288). SL species includes aquatic (14 genomes), parasitic (3 genomes), and carnivorous (4 genomes, 2 of which are also aquatic) species.

1
2

1 The scattered distribution of species with unusually high or low numbers of *NLR*
2 genes across the angiosperms suggested rapid *NLR* expansion and/or contraction
3 during speciation. To test the extent to which the *NLR* gene number varied among
4 closely related species, we calculated the fold-difference in *NLR* gene number
5 between the species with the most *NLR* genes and the species with the least *NLR*
6 genes within each of the 85 families. Within the 42 families for which multiple
7 genomes were available, the fold-difference in *NLR* gene number ranged from 1.2 to
8 more than 66 (Figure 2B). In 17 families, there was a greater than five-fold difference
9 in interspecific *NLR* gene numbers (Figure 2B). These data suggested that *NLR* gene
10 expansion and/or contraction can occur rapidly during speciation and may facilitate
11 the rapid response of plants to changeable pathogenic environments. Importantly,
12 however, there were no significant differences in *NLR* gene numbers among the five
13 major angiosperm clades (Figure 2C), suggesting that evolutionary forces do not favor
14 maintaining stable gene numbers in specific plant clades over evolutionary
15 time-scales.

16 Across all genomes investigated, aquatic, parasitic, and carnivorous plants had
17 significantly fewer *NLR* genes on average than other angiosperm genomes (Figure
18 2D). For example, of the 12 aquatic plant genomes included in this study, seven had
19 fewer than 100 *NLR* genes, and an additional four had no more than 200 *NLR* genes
20 (Table S2). These species included three basal angiosperms, four monocots, four
21 dicots, and one Ceratophyllales, suggesting a convergent contraction of *NLR* genes
22 when adapting to an aquatic lifestyle. Two aquatic dicots in the Lentibulariaceae
23 (superasterid lineage), *Genlisea aurea* and *Utricularia gibba*, had fewer than 50 *NLR*
24 genes, consistent with drastically shortened genomes of these species (Lan et al., 2017;
25 Leushkin et al., 2013). However, a third Lentibulariaceae species, *U. reniformis*,
26 which is terrestrial and has a large genome (300 Mb; (Silva et al., 2019), also had
27 fewer than 50 *NLR* genes. Thus, the low *NLR* gene numbers in the *G. aurea* and *U.*
28 *gibba* genomes might not be associated with genome size reduction and an adaptation
29 to an aquatic lifestyle. Instead, these low numbers might be linked to a lifestyle shared
30 by all three species: carnivory (Figure 2A). Consistent with this speculation, another

1 carnivorous plant in the rosoid lineage, *Cephalotus follicularis*, also had an unusually
2 small number of *NLR* genes (33 genes). *NLR* copy reductions were also observed in
3 several parasitic species (Figure 2A). The two parasitic species in the
4 Convolvulaceae included in ANNA, *Cuscuta australis* and *C. campestris*, possessed
5 only 15 and 16 *NLR* genes, respectively, although both have large genomes including
6 more than 40,000 protein coding genes each (Sun et al., 2018; Vogel et al., 2018).
7 Likewise, the genome of the parasitic dicot *Striga asiatica* contained only 89 *NLR*
8 genes. These data indicated that convergent reductions in *NLR* genes were also
9 associated with adaptations to a parasitic lifestyle. These examples suggested that
10 adaptations to specific lifestyles or environments may have shaped the *NLR* profiles
11 of the angiosperms. Additionally, the utter absence of *NLR* genes from the *U. gibba*
12 genome shows that at least some angiosperms can survive without *NLR* genes.

13 **Evolutionary associations among *NLR* subclasses**

14 Angiosperms inherited three *NLR* subclasses from their common ancestor (Shao et al.,
15 2016). Our data revealed that the *TNL* and *CNL* subclasses together account for over
16 90% of all *NLR* genes in 291 out of the 305 genomes (Table S2). The remaining 14
17 genomes have extremely low numbers of *NLR* genes overall. *RNL* genes represented
18 no more than 10% of all *NLR* genes in most angiosperms, reflecting the conserved
19 signal transduction functions of these genes. *CNL* genes were the most common *NLR*
20 genes in 268 angiosperm genomes, while *TNL* genes were the most common *NLR*
21 genes in only 36 genomes (Figure 3A). This might be because the ancestral *CNL*
22 lineage has gradually expanded in the angiosperms over evolutionary time, while the
23 *TNL* lineage has remained relatively stable (Shao et al., 2016). Of the genomes in
24 which the *TNL* genes were most common, half of them were species in the
25 Brassicaceae. This was consistent with the ancestral expansion of the *TNL* genes in
26 the Brassicaceae lineage (Zhang et al., 2016). The remaining 18 species in which *TNL*
27 genes dominate were scattered among families that tend to be *CNL*-dominant,
28 suggesting species-specific *TNL* to *CNL* ratio alterations.

29

30

Figure 3. Variations in *NLR* subclass abundance among angiosperms. A) Relative proportions of the three *NLR* subclasses in the angiosperm genomes included in this study. B) Spearman correlation analysis for gene numbers among different *NLR* subclasses. The correlation coefficient is shown at top right of each graph. **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

1

2

3 The variations in the relative abundances of the three *NLR* subclasses among the

angiosperm genomes raised two interesting questions: first, does the plant immune system require the synchronous expansion or contraction of certain *NLR* subclasses in order to respond to altered pathogenic environments; and second, have *NLR* detectors co-evolved with *NLR* helpers. Analysis of *NLR*, *CNL*, *TNL*, and *RNL* gene numbers identified a robust, significant correlation between total *NLR* gene number and *CNL* gene number (Figure 3B), probably due to the high relative abundance of *CNL* genes in most angiosperm genomes. The numbers of *TNL* and *RNL* genes were also significantly correlated with total *NLR* gene number, although both of these correlations were weaker than that between *NLR* gene number and *CNL* gene number. These weak correlations might be associated with the inconsistent variations in *RNL* and *TNL* gene numbers with respect to variations in *CNL* gene numbers across lineages, and the fact that neither subclass predominated in a majority of the angiosperm lineages. Consistent with this, we identified a poor correlation between *TNL* and *CNL* gene numbers, suggesting that these two *NLR* subclasses might evolve independently in response to pathogenic selection in some species. Interestingly, we observed a strong correlation between *TNL* and *RNL* copy numbers, but not between *CNL* and *RNL* copy numbers. Further dividing the *RNL* subclass into the *ADR1* and *NRG1* lineages revealed that the correlation between *RNL* and *TNL* gene copy numbers was primarily driven by the *NRG1* lineage. These results were consistent with previous functional studies, which have shown that nearly all investigated *TNL* proteins are functionally dependent on *RNL* proteins, particularly the *NRG1* lineage (Castel et al., 2019; Qi et al., 2018; Saile et al., 2020; Wu et al., 2019). In contrast, some *CNL* proteins could confer resistance when the *RNL* genes are deleted (Saile et al., 2020).

Ancient and ongoing complete loss of *TNL* genes from angiosperm genomes

TNL loss in the common ancestor of the monocots is a familiar, long-standing hypothesis. However, genome-wide data supporting this notion have been restricted to the Poaceae and a few non-Poaceae species. Here, the data in ANNA (65 monocot genomes across nine families in six orders) provide substantial support for the hypothesis of ancestral *TNL* loss in the monocots (Figure 4A and Table S2).

Figure 4. Absence of TNL genes and TNL signal transfer components in angiosperms. A) A heat map showing gene copy numbers for TNL and TNL signal transfer components (RNL and EDS1 family members). Phylogenetic relationships among the 305 angiosperm genomes resolved to the order level were constructed following the APG IV system (Byng et al., 2016). Green circles indicate simultaneous losses of TNL and all five signal transduction components. Red stars indicate simultaneous losses of TNL, NRG1, and SAG101. B-D) Predicted histories of loss of TNL and TNL signal transfer components in (B) the Magnoliales, (C) the Lamiales, and (D) the Ranunculales. Phylogenetic relationships among and within the three clades were constructed based on (Byng et al., 2016; Song et al., 2017; Mint-Evolutionary-Genomics-Consortium, 2018; Zumajo-Cardona et al., 2018; Xue et al., 2021). E) Numbers of species lacking TNL and/or various TNL signal transfer components. Black dots indicate gene loss, and gray dots indicate gene presence.

Recent studies have also reported the absence of TNL genes in a basal dicot species (*Aquilegia coerulea*) and two Lamiales species (*Sesamum indicum* and *Mimulus*

1 *guttatus*), suggesting that *TNL* loss has also occurred in the dicots (Collier et al., 2011;
2 Shao et al., 2016). *TNL* genes in angiosperms descended from only a few ancestral
3 genes, which did not expand during long-term evolution (Shao et al., 2016). Therefore,
4 when many genomes are surveyed, it is possible to detect additional *TNL*-lacking
5 species among dicots that have not experienced substantial *TNL* expansion. Our data
6 revealed that *TNL* loss indeed occurred frequently in the dicots. Of the 230 surveyed
7 dicot species, 23 exhibited a complete loss of the *TNL* genes (Figure 4A and Table S2).
8 These species were scattered across 18 families in nine orders and included
9 superrosids, superasterids, and basal dicots. Moreover, of the five magnoliid genomes
10 investigated, *TNL* genes were absent in the four more basal species, but three *TNL*
11 genes were detected in the most derived genome, *Cinnamomum kanehirae* (Figure
12 4B). These results suggested that independent *TNL* losses may have occurred
13 repeatedly throughout dicot and magnoliid diversification.

14 Previous studies have shown that the genomes of *S. indicum* and *M. guttatus* lack *NLR*
15 genes, suggesting that *NLR* genes may have been lost in the common ancestor of the
16 Lamiales (Collier et al., 2011; Shao et al., 2016). Interestingly, our data indicated an
17 extensive but not exhaustive absence of *TNL* genes in the Lamiales order (Figure 4C).
18 Of the 13 Lamiales species investigated, nine were lacking *TNL* genes. The remaining
19 four species did not form a single basal clade but were instead scattered across the
20 Lamiales phylogeny (Figure 4C). If we define a *TNL* loss event based on the presence
21 of *TNL* genes in the sister lineage, our phylogeny suggested at least five independent
22 *TNL* loss events throughout the diversification of the Lamiales. A similar pattern of
23 independent *TNL* loss was also observed in the basal dicot lineage Ranunculales
24 (Figure 4D), where three predicted *TNL* loss events resulted in the absence of *TNL*
25 genes in four of the five genomes examined. Nine additional independent *TNL* loss
26 events were identified: in the dicots *Carnegieia gigantea*, *Gelsemium sempervirens*,
27 *Begonia fuchsioides*, *Cladopus chinensis*, *Juglans regia*, *Ostrya chinensis*, *Citrus*
28 *grandis*, and the genus *Cuscuta*, as well as the Ceratophyllales *Ceratophyllum*
29 *demersum* (Figure 4A). In total, we identified at least 23 independent *TNL* loss events
30 across the angiosperm genomes examined, suggesting ancient, recent, and ongoing

1 genome-wide *TNL* erasure.

2 ***TNL* loss is tightly associated with absence of *TNL* signal transduction**
3 **components**

4 The absence of *TNL* genes in monocots has long puzzled researchers. However, recent
5 studies have revealed the co-absence of the *RNL-NRG1* lineage and *TNL* genes in
6 monocots, two Lamiales species, and *A. coerulea* (Collier et al., 2011; Shao et al.,
7 2016). In addition to *NRG1*, various other proteins play a role in *TNL* signal
8 transduction, including *EDS1*, *PAD4*, *SAG101*, and *ADR1* (Castel et al., 2019; Lapin
9 et al., 2019; Qi et al., 2018; Saile et al., 2020; Wu et al., 2019). Across all 305
10 angiosperm genomes, we identified 93, 109, 29, 23, 40, and 110 lacking *TNL*, *NRG1*,
11 *ADR1*, *EDS1*, *PAD4*, and *SAG101*, respectively (Table S2). *NRG1* and *SAG101* were
12 the most frequently lost genes; these genes were both absent in 104 genomes. *NRG1*
13 and *SAG101* were also the most frequently absent genes in genomes lacking *TNL*
14 genes: of the 93 genomes lacking *TNL*, 89 had simultaneously lost either *NRG1* or
15 *SAG101*, and 63 of these had lost both genes (Figure 4E). As expected, most of the
16 genomes lacking *TNL* as well as *NRG1* and/or *SAG101* were monocots (65 in total).
17 In monocots, the absence of these genes is likely the ancestral state. By contrast, in
18 the dicots and magnoliids, the scattered losses of *TNL* as well as *NRG1* and/or
19 *SAG101* across the phylogeny might suggest multiple independent loss events (Figure
20 4A). The strong correlation between *TNL* loss and the loss of the signal component(s)
21 *NRG1* and/or *SAG101* further emphasizes the functional dependence of *TNL* proteins
22 on the *NRG1*-*SAG101* pathway.

23 In addition to the 89 genomes lacking *TNL* as well as *NRG1* and/or *SAG101*, more
24 than 20 genomes carried *TNL* but lacked *NRG1* and/or *SAG101*. Conversely, only four
25 genomes carrying *NRG1* and *SAG101* lacked *TNL* (Figure 4B). This suggested that
26 *NRG1* and/or *SAG101* loss often preceded *TNL* loss in angiosperm genomes. This
27 possibility was consistent with the pattern of gene loss events mapped onto
28 species-level phylogenies for several taxa. For example, two species in the Lamiales
29 (*Antirrhinum majus* and *Tectona grandis*) carried *TNL* as well as all five signal
30 transduction components (Figure 4C), suggesting that the ancestral Lamiales carried

all of these genes. Of the remaining 11 species analyzed in this group, nine lacked *TNL*, *NRG1*, and *SAG101*; however, one species (*Handroanthus impetiginosus*) lacked *NRG1* and *SAG101* (as well as *PAD4*), but not *TNL* (Figure 4C). This suggested that, in the Lamiales, *TNL* loss followed the loss of *NRG1* and *SAG101*. A similar pattern of gene loss was also recovered in the magnoliids (Figure 4B): our phylogeny suggested that *NRG1* and *SAG101* were absent in the ancestral magnoliid, and that the *TNL* gene was independently lost four times during subsequent diversification. One species, *Persea americana*, lost *PAD4* and *ADR1* in addition to *TNL*. Thus, in both the magnoliids and the Lamiales, our character trait mapping indicated that *TNL* was lost after *NRG1* and/or *SAG101*. Notably, although *EDS1*, *PAD4*, and *ADR1* were also occasionally lost with *TNL*, these genes were absent less frequently than *TNL*, *NRG1*, and *SAG101* (Figure 4E). This might be because *EDS1*, *PAD4*, and *ADR1* are required for *CNL* function and basal immunity in addition to *TNL* signal transduction, resulting in a tendency for conservation (Lapin et al., 2020; Saile et al., 2020). Overall, our results suggested that the loss of *NRG1* and/or *SAG101* might drive the erasure of *TNL* genes from angiosperm genomes.

A previous study revealed that independent small-scale deletions involving a few genes contributed to the deletion of *EDS1*, *PAD4*, and *ADR1* genes from angiosperm genomes (Baggs et al., 2020). To better understand how *SAG101* and *NRG1* were lost over evolutionary time, pairwise genomic synteny analyses were performed using both dicot and monocot species. Synteny analysis identified homologous regions surrounding *SAG101* and *NRG1* in both the monocot and the dicot species examined (Figure S2). Therefore, the repeated loss of *SAG101* and *NRG1* might be a result of small-scale or single-gene loss.

Convergent retention of a conserved *TNL* lineage in species lacking *NRG1* and *SAG101*

Above, we reported that 27 analyzed genomes lacked *NRG1* and/or *SAG101* but retained one or more *TNL* genes (Figure 4E). Two different mechanisms may have led to the preservation of *TNL* genes in these *NRG1*/*SAG101*-deficient species. First, if the loss of *NRG1* and/or *SAG101* occurred relatively recently,

1 *NRG1/SAG101*-dependent *TNL* genes may have not yet been pseudogenized or
2 deleted from the genome. Second, some *TNL* genes may be functionally independent
3 of the *NRG1-SAG101* pathway, and therefore might be preserved despite *NRG1*
4 and/or *SAG101* loss. To better understand how often each of these possible
5 mechanisms are implicated in *TNL* preservation, we constructed a phylogeny of all
6 available *TNL* genes from genomes lacking *NRG1* and/or *SAG101*. *TNL* sequences
7 from *A. thaliana*, *Solanum lycopersicum*, and *Amborella trichopoda* were included to
8 help better resolve relationships of the phylogeny (Figure 5A). If we assume that all
9 *TNL* genes depend on the *NRG1-SAG101* pathway, we would expect to observe the
10 random loss of various *TNL* lineages in multiple angiosperm genomes after the
11 *NRG1-SAG101* pathway is destroyed. Surprisingly, one *TNL* lineage was conserved
12 across 23 of the 27 genomes examined (Figure 5B). This *TNL* lineage also included
13 genes from *A. thaliana*, *S. lycopersicum*, and *A. trichopoda*, suggesting that it
14 diverged prior to angiosperm radiation. In addition, each genome within in this
15 lineage had less than three *TNL* genes, reflecting a conserved pattern of evolution.
16 To confirm that the *TNL* genes within the conserved *TNL* lineage were orthologous,
17 we performed a syntenic analysis of the genes in this lineage from four species: the
18 superrosid *A. thaliana*, the superrosid *Vigna radiata*, the superasterid *Actinidia*
19 *chinensis*, and the superasterid *S. lycopersicum*. We found that the *TNL* genes in the
20 conserved lineage formed well-aligned chromosomal blocks in these four species
21 (Figure 5C), supporting an orthologous relationship among these *TNL* genes. The
22 wide distribution of these genes, in conjunction with their highly conserved
23 chromosomal context, suggest that this may be most conserved lineage of *TNL* genes
24 identified in angiosperms to date. Moreover, the convergent conservation of this *TNL*
25 lineage in species lacking *NRG1* and/or *SAG101* that diverged over 100 million years
26 ago suggested that these *TNL* genes remain functional and may be functionally
27 independent of the *NRG1-SAG101* pathway (Figure 5B).

Figure 5. Convergent conservation of a *TNL* lineage in angiosperm species lacking the *NRG1-SAG101* pathway. A) Phylogenetic analysis of *TNL* genes in 27 species carrying *TNL* but not *NRG1* and/or *SAG101*. Branch support values from SH-aLRT and UFBoot2 are indicated at each node. The orange box indicates a conserved *TNL* lineage. B) Expanded view of the conserved lineage highlighted in panel (A). C) Syntenic analysis of the conserved *TNL* lineage in *A. thaliana*, *S. lycopersicum*, *A. trichopoda*, and *V. radiata*.

Identification of candidate *TNL* signal transfer components based on evolutionary patterns

The convergent loss of *TNL* plus *NRG1* and/or *SAG101* in 89 species hinted that there might be other key signal components for *TNL* proteins, as yet unknown, that have been convergently lost. To test this hypothesis, we first identified 23 orthologous

1 groups present in five species carrying *TNL*, *NRG1*, and *SAG101* (*A. thaliana*, *S.*
2 *lycopersicum*, *Eschscholzia californica*, *Glycine max*, and *Olea europaea*) but absent
3 in five species lacking all three genes (*A. coerulea*, *Musa acuminata*, *Persea*
4 *americana*, *Salvia miltiorrhiza* and *Zea mays*; Figure 6A) using the online
5 OrthoVenn2 server (Xu et al., 2019). Importantly, this analysis recovered the *NRG1*,
6 *SAG101*, and *TNL* genes, supporting its reliability (Figure 6A, Table S5). To further
7 investigate the loss of these 23 orthogroups in the species that had also lost *NRG1*,
8 *SAG101*, and *TNL*, we performed a BLASTp search, querying the 20 *A. thaliana*
9 protein sequences (representing the 20 remaining orthogroups excluding *NRG1*,
10 *SAG101*, and *TNL*) against all 63 angiosperm species lacking *NRG1*, *SAG101*, and
11 *TNL*. Five of the 20 orthogroups (clusters 4, 6, 8, 10, and 12) were also lost in over
12 50% of the 63 genomes lacking *NRG1*, *SAG101*, and *TNL* (Figure 6B). The GO
13 annotations generated by the OrthoVenn2 server indicated that the *A. thaliana* genes
14 in these five orthogroups were associated with signal transduction, transcription, the
15 wound response, transmembrane transport activity, and the oxidative stress response
16 (Table S5).

17 To explore the potential involvement of these genes in *TNL*-triggered immunity, we
18 performed a meta-analysis of their expression patterns in *A. thaliana* after infection
19 with *TNL*-recognized pathogens (Asai et al., 2014b; Yang et al., 2017). We found that
20 *A. thaliana* genes in clusters 4, 6, and 8 (AT2G33720, AT2G04795, and AT2G47270,
21 respectively) were upregulated after infection with *Pseudomonas syringae* DC3000
22 (Figure 6C), which is recognized by the *TNL* protein RPS4 (Narusaka et al., 2009).
23 AT2G47270 (cluster 4) was also upregulated when *A. thaliana* was challenged with
24 the oomycete *Hyaloperonospora arabidopsidis* (Figure 6C), which is recognized by
25 the *TNL* protein RPP4 (van der Biezen et al., 2002). AT2G47270 encodes the
26 UPBEAT1 (UPB1) transcription factor, which harbors a bHLH domain (Li et al.,
27 2019). A previous study showed that UPB1 regulates the expression of a set of
28 peroxidases that modulate the balance of reactive oxygen species (ROS) (Li et al.,
29 2019). ROS inducement is common in ETI, suggesting that UPB1 potentially acts as a
30 signal transduction component in *TNL*-triggered immunity (Littlejohn et al., 2020).

1 However, the exact functions of UPB1 in plant disease resistance remain to be
2 explored.

Discussion

Our understanding of plant *NLR* gene evolution is continuously extended as plant genomes become increasingly available (Bai et al., 2002; Meyers et al., 2003; Shao et al., 2019). Here, we developed ANNA, a database including over 90,000 *NLR* genes from more than 300 angiosperm genomes, and used this large dataset to explore the co-evolution of *NLR* subclasses, and their adaptive evolution leading to plant ecological specialization.

NLR contraction tends to occur in aquatic, parasitic, and carnivorous species

Variations in *NLR* gene number among species have been widely reported. Indeed, these variations may be even more extreme than previously thought: studies of the Fabaceae, Solanaceae, Poaceae, and Brassicaceae have identified only two- to six-fold differences in *NLR* number among confamilial species, but a 20-fold difference in *NLR* number was recovered in the Orchidaceae (Luo et al., 2012; Qian et al., 2017; Shao et al., 2014; Tirnaz et al., 2020; Xue et al., 2020; Zhang et al., 2016). Here, our survey of 305 angiosperm genomes identified the greatest within-family difference in *NLR* number in the Poaceae: a 66-fold difference in *NLR* number between *T. aestivum* (2298 *NLR* genes) and *Oropetium thomaeum* (35 *NLR* genes). This pattern of extreme variation among closely related species indicates that *NLR* genes may undergo rapid and dramatic expansion and/or contraction after speciation.

It is difficult to determine the exact forces driving these species-specific *NLR* expansions and contractions because the precise qualities of the pathogenic environment during speciation are unknown. However, we identified a convergent reduction in *NLR* genes in plants with three distinct, specialized lifestyles: aquatic, parasitic, and carnivorous. That is, 11 of the 12 aquatic plants surveyed had no more than 200 *NLR* genes, and nine of these had fewer than 100 *NLR* genes. This is consistent with the findings of a previous study, which showed convergent reductions of *NLR* genes among four aquatic plants in two orders (Baggs et al., 2020). Furthermore, all carnivorous and parasitic plants surveyed exhibited convergent reductions in *NLR* gene number. These results suggested that adaptations to certain specialized lifestyles may have contributed to *NLR* reduction. Notably, no *NLR* genes

1 were detected in the carnivorous aquatic plant *U. gibba*, demonstrating that at least
2 some angiosperms survive without *NLR* genes.

3 Associations between *NLR* reduction and aquatic habitats have been observed in fern
4 genomes (Li et al., 2018) as well as angiosperm genomes. For example, the genomes
5 of the aquatic ferns *Azolla filiculoides* and *Salvinia cucullata* contain only one and
6 eight *NLR* genes, respectively (Table S7). Plant-pathogen interactions are generally
7 divided into stomatal immunity to leaf-associated pathogens and rhizospheric
8 immunity to root-associated pathogens (Zhang et al., 2020). Although there is no
9 empirical evidence to suggest that fewer pathogens invade aquatic plants, the
10 anaerobic environment surrounding the roots of aquatic plants suggests that
11 rhizospheric microbe composition might differ between terrestrial and aquatic plants.
12 Indeed, the roots of some aquatic plants are free-floating, and thus have hardly any
13 interaction with soil microbes. The altered or reduced interactions between roots and
14 microbes in aquatic plants might imply that aquatic plants are less dependent on
15 *NLR*-triggered immunity. Consistent with this, although *NLR* genes originated in the
16 common ancestor of green plants, *NLR* genes are absent from most sequenced green
17 algae genomes, and only began to expand after plants colonized the terrestrial
18 environment (Shao et al., 2019). The silence of the *NLR* family in algal species during
19 throughout several hundred million years of evolution, followed by the rapid
20 expansion of these genes after the invasion of the terrestrial environment might reflect
21 the reduced dependence of aquatic plants on *NLR* triggered immunity (Figure 7A).

22

Figure 7. Schematic diagram showing the adaptive evolution of the *NLR* gene leading to ecological specialization and the deletion of signal transduction components. A) A reduction in *NLR* gene number is usually coupled with adaptations to aquatic (blue branches), parasitic (yellow branch), and carnivorous (green branch) lifestyles. *NLR* gene numbers in non-vascular plants are from Shao et al (2019). B) A summary of the dependence of CNL- and TNL-triggered immunity on the EDS1 and RNL proteins. C) A hypothesized mechanism of TNL loss in angiosperms. EDS1 and RNL proteins are known to be important signal transduction components for most TNL genes. The frequent loss of *SAG101* and *NRG1* may drive TNL reduction and loss. However, a few TNL genes may be functionally independent of *SAG101* and *NRG1*, and thus may be preserved in some species lacking a *SAG101*/*NRG1* pathway. Additive loss of *EDS1* and *RNL* genes may cause total *NLR* number reduction.

- 1
- 2 The altered microbial environment of the roots may also explain the scarcity of *NLR*
- 3 genes in parasitic and carnivorous plants. That is, parasitic plants often lack roots;
- 4 some, such as *Cuscuta* spp., even lack leaves (Kaiser et al., 2015). Thus, these plants
- 5 have little interaction with root (and even leaf) pathogens. Similarly, aquatic
- 6 carnivorous plants are often rootless, and many hygrophytic and epiphytic
- 7 carnivorous species have only weakly developed root systems (Adlassnig et al., 2005).
- 8 Furthermore, most carnivorous plants live in wet soil environments that contain very

1 limited nitrogenous material due to acidic or other unfavorable soil conditions (Xu et
2 al., 2020). Such soil conditions may also restrict microbial growth. In addition, the
3 leaves of most carnivorous plants form specialized trap structures that contain
4 digestive enzymes; these enzymes may help to defend the plant from microbial
5 invasion (Renner and Specht, 2013; Thorogood et al., 2018). Finally, carnivorous
6 plants produce a large number of diverse secondary metabolites, many of which
7 function as antifeedant and antifungal agents (Pavlovic and Mithofer, 2019). Thus,
8 several characters of carnivorous plants, including their specialized root and leaf
9 structures, low-nutrient habitat, and secondary-metabolite content, may together tend
10 to relax the selective pressures on *NLR* genes in these taxa. In general, morphological
11 adaptations to specialized lifestyles may have reduced the interactions between plants
12 and microbes, leading to the contraction or elimination of *NLR* genes (Figure 7A).

13 **Signal transduction pathway deficiencies drive *NLR* contraction**

14 From their common ancestor, the angiosperms inherited three *NLR* subclasses, each of
15 which plays a unique role in PTI (Shao et al., 2016). Nearly all characterized CNL
16 and TNL proteins participate in PTI, either by directly interacting with effector
17 proteins, or via indirect recognition using guard or decoy mechanisms (Kourelis and
18 van der Hoorn, 2018). This long-term arms-race drove convergent *TNL* and *CNL*
19 expansion in the angiosperms at the Cretaceous-Paleogene (K-P) boundary (Shao et
20 al., 2016). Interestingly, in this study, *TNL* genes were absent from the genomes of the
21 entire monocot clade, the majority of the Magnoliids, and some of the dicots due to
22 repeated gene loss events.

23 Recent studies have shown that most TNL proteins depend on RNL to confer full
24 resistance (Castel et al., 2019; Qi et al., 2018; Saile et al., 2020; Wu et al., 2019). The
25 dimerization or tetramerization of TNL proteins upon effector recognition activates
26 NADase in the TIR domain, which subsequently transfers the defense signal to the
27 downstream EDS1/SAG101 or EDS1/PAD4 complex (Ma et al., 2020; Wan et al.,
28 2019; Zhou and Zhang, 2020). In *Arabidopsis*, the signal from EDS1 to the
29 downstream components travels via two different pathways: one signal is passed to
30 NRG1 to induce a hypersensitivity reaction (HR), whereas the other is passed to the

1 ADR1 to restrict bacterial invasion (Lapin et al., 2020). Consistent with this
2 functional analysis, our results showed that *TNL* copy number was highly correlated
3 with *NRG1* copy number across the 305 surveyed angiosperm genomes. Furthermore,
4 we observed the co-absence of *NRG1/SAG101* and *TNL* in several dozens of species.
5 These results strongly suggested that the frequent co-absence of *NRG1/SAG101* and
6 *TNL* genes is not accidental. Furthermore, the deduced prior loss of *NRG1/SAG101* in
7 *TNL*-absent species suggested that the loss of *TNL* signal transfer components might
8 be one of the driving forces underlying *TNL* loss. However, we identified a conserved
9 *TNL* lineage in over 20 angiosperm genomes lacking *NRG1/SAG101*. These genomes
10 were widely scattered throughout the angiosperm phylogeny. The convergent
11 preservation of this anciently diverged *TNL* lineage in *NRG1/SAG101*-absent species
12 hints that *TNL* genes might have functions independent of the *NRG1-SAG101*
13 pathway, and that these genes may trigger immunity through a different signal
14 transduction pathway (Figure 7B). Therefore, our results suggested that
15 *NRG1/SAG101* loss is the major, but not the only, factor driving *TNL* loss (Figure 7C).
16 However, the pattern of co-absence between *TNL* and *TNL* signal pathway
17 components helped us to identify UPB1 as a novel candidate *TNL* signal transfer
18 component. Additional studies of this transcription factor may help to further clarify
19 the evolution and functional mechanisms of the *TNL* genes.

20 While investigating the association between *TNL* loss and the loss of *TNL* signal
21 transduction components, we identified 16 genomes (six monocots and ten dicots)
22 lacking all five *RNL* and *EDS1* family genes (Figure 4B). It was unsurprising that
23 species that had lost *EDS1* had also lost *TNL*, due to the critical role played by the
24 *EDS1* protein in *TNL* signal transduction (Lapin et al., 2020). In conjunction with the
25 loss of *SAG101* and *NRG1*, the loss of *EDS1* would increase the likelihood of *TNL*
26 loss. Interestingly, all 16 of these species had few *NLR* genes: 11 of these genomes
27 had less than 50 *NLR* genes, while an additional three had no more than 100 *NLR*
28 genes. However, *TNL* loss driven by *NRG1/SAG101* deficiency did not necessarily
29 lead to a reduction in *NLR* gene number, as *CNL* gene expansion may occur to
30 compensate. This pattern was observed in many monocot species, which possess

1 many *NLR* genes but have lost *TNL* genes. Therefore, it is possible that the reduction
2 in *NLR* number seen in the species lacking *RNL* and *EDSI* was due to a reduction in
3 *CNL* gene number concurrent with *TNL* loss. Recent studies have shown that CNL
4 proteins adopt at least two different mechanisms to trigger immunity (as summarized
5 in Figure 7B). The CNL protein ZAR1 induced HR and pathogen resistance
6 independent of both RNL lineages by assembling a polymer-protein resistosome and
7 forming pores on the cell membrane (Saile et al., 2020; Wang et al., 2019). In addition,
8 a conserved MADA motif was found in the CC domain of ZAR1 and other CNL
9 proteins (Adachi et al., 2019). The motif is essential for the induction of cell death by
10 the CNL protein in an RNL-independent manner. However, some CNL proteins do not
11 have the MADA motif at the CC domain (Adachi et al., 2019). These CNL proteins
12 likely do not form pores at the cell membrane, and may induce cell death by
13 transferring the defense signal to downstream signal components. For example, the
14 RPS2 protein, which does not harbor the MADA motif, requires ADR1 to confer full
15 resistance (Saile et al., 2020). Several other CNL proteins are also functionally
16 dependent on RNL (Castel et al., 2018). Therefore, the deletion of *RNL* genes from
17 angiosperm genomes may also drive *CNL* contraction by disrupting the functions of
18 RNL-dependent CNLs. Moreover, the convergent deletion of both *RNL* and *EDSI*
19 family genes in certain taxa may suggest that ETI has become less important in these
20 lineages over evolutionary time. Therefore, the reductions in *NLR* gene number
21 observed in these species might result from the combined losses of all five component
22 genes, as well as of other undefined immune-system components (Figure 7C).

23 Baggs et al. (2020) reported the convergent loss of *EDSI*, *PAD4*, and *ADR1* in four
24 aquatic species, and hypothesized that the plant immune system co-evolved with the
25 drought response system. Of the 16 species lacking the *RNL* and *EDSI* gene families
26 identified in the present study, eight are indeed aquatic species: *Spirodela polyrhiza*,
27 *Lemna minor*, *Zostera marina*, *Zostera muelleri*, *U. gibba*, *G. aurea*, *Cladopus*
28 *chinensis*, and *Ceratophyllum demersum*. Although the remaining eight species
29 (*Asparagus officinalis*, *Begonia fuchsioides*, *C. campestris*, *C. australis*, *G. elata*,
30 *Kingdonia uniflora*, *Papaver somniferum*, and *U. reniformis*) are not aquatic, two are

1 parasitic and one is carnivorous, while the remaining species have diverse lifestyles.
 2 These results emphasized that plants with aquatic and parasitic lifestyles are apt to
 3 convergently lose disease resistance signal transduction components (e.g., *RNL* and
 4 *EDSI* family genes) convergently, which may reduce the number of *NLR* genes in the
 5 genome. However, other mechanisms, yet to be characterized, apparently also result
 6 in the convergent loss of *RNL* and *EDSI* family gene as well as *NLR* contraction, as
 7 observed in the four non-aquatic, non-parasitic, and non-carnivorous species.
 8 In conclusion, we herein established ANNA, a dataset including the *NLR* genes from
 9 more than 300 angiosperm genomes and covering most major angiosperm clades.
 10 This large dataset may serve as an essential resource for exploring various aspects of
 11 *NLR* gene evolution. Using this dataset, we revealed dramatic variations in *NLR* copy
 12 numbers among closely related species and identified notable *NLR* reductions in
 13 species adapted to aquatic, parasitic, and/or carnivorous lifestyles. The high
 14 correlation between *TNL* and *RNL* gene numbers, in addition to the high frequency of
 15 *TNL* and *NRG1* co-absence, suggested that NRG1-SAG101 pathway deficiency may
 16 drive repeated independent *TNL* losses. Our findings provide new insights into the
 17 evolution of *NLR* genes in the context of plant lifestyles and genome content
 18 variation.

19

20 **Materials and methods**

21 **Data used in this study**

22 The genomic sequences, annotations, and gene models for the 305 angiosperm
 23 genomes and two fern genomes were downloaded from public databases (listed in
 24 Table S1). The transcriptome data for *A. thaliana* were retrieved from previously
 25 published works (Asai et al., 2014b; Yang et al., 2017).

26 **Identification of *NLR* and *EDSI* family genes**

27 *NLR* genes were identified in the 305 angiosperm genomes as described previously
 28 (Shao et al., 2015). Briefly, we first screened all protein sequences in each genome for
 29 the NBS domain (NB-ARC, Pfam: PF00931) using an HMM search, as implemented
 30 in hmmer3.0 (Johnson et al., 2010) with default parameters. The amino acid

1 sequences of obtained *NLR* genes were then used to run a genome-wide BLASTp
2 analysis for each genome (E-value = 1.0). All hits were further analyzed using
3 hmmscan in hmmer3.0 against a local Pfam-A database (E-value = 10^{-4}) to confirm a
4 detectable NBS domain in each sequence. When two or more transcripts were
5 annotated for a single gene, the longest form was selected. All obtained *NLR*
6 candidates were classified into the *TNL*, *CNL*, and *RNL* subclasses based on a
7 BLASTp search of their amino acid sequences against a local database containing
8 *NLR* subclass genes representing species across the green plants (Shao et al., 2019). A
9 similar strategy was adopted to identify EDS1 proteins. The obtained EDS1 family
10 members were classified based on BLASTp alignments against EDS1, SAG101, and
11 PAD4 protein sequences from *A. thaliana*, *Solanum tuberosum*, *S. lycopersicum*,
12 *Solanum tuberosum*, and *Nicotiana benthamiana* (E-value = $1e-10$).

13 **Sequence alignment and phylogenetic analysis**

14 Sequence alignment and phylogenetic analysis of *TNL* genes were performed as
15 previously described (Shao et al., 2019), with *Arabidopsis thaliana* *RNL* sequences as
16 outgroups. Amino acid sequences of the NBS domain were aligned using ClustalW
17 (Edgar, 2004) with default options (Thompson et al., 1994), and then manually
18 corrected in MEGA 7.0 (Kumar et al., 2016). Genes with very short or divergent NBS
19 domains were eliminated from the matrix because these interfered with fine alignment
20 and phylogenetic analysis. Phylogenetic analyses were performed using IQ-TREE
21 with the maximum likelihood algorithm (Nguyen et al., 2015). ModelFinder was used
22 to estimate the best-fit model of nucleotide substitution (Kalyaanamoorthy et al.,
23 2017). Branch support values were calculated using SH-aLRT (Anisimova et al., 2011)
24 and UFBoot2 (Minh et al., 2013) with 1000 bootstrap replicates.

25 **Synteny analyses**

26 The synteny of the conserved *TNL* lineage was analyzed in four species: *Actinidia*
27 *chinensis*, *Amaranthus hypochondriacus*, *Arabidopsis thaliana*, and *Solanum*
28 *lycopersicum*. The synteny of the *NRG1* and *SAG101* proteins was analyzed in five
29 species: *Antirrhinum majus*, *Cladopus chinensis*, *Sesamum indicum*, *Spirodela*
30 *polyrhiza*, and *Theobroma cacao*. Genome-wide pair-wise BLASTp analyses were

performed between all protein sequences of the target species (E-value = 1e-10). Synteny analysis was performed with the MCScanX module of TBtools v0.6669 (Chen et al., 2020; Wang et al., 2012). Syntenic relationships were then drawn using TBtools.

CD-hit analyses

The *NLR* protein sequences of species with more than 1000 or less than 50 *NLR* genes were analyzed using CD-HIT version 4.6 (Huang et al., 2010), with identity cutoffs of 0.8 and 0.9, respectively.

Orthologous group identification

The online OrthoVenn2 server (<https://orthovenn2.bioinfotoolkits.net>, v2.0.9) (Xu et al., 2019) was used to identify orthologous gene groups in five species carrying *TNL*, *NRG1*, and *SAG101*, and in five species lacking *TNL*, *NRG1*, and *SAG101* (E-value = 1e-10). *A. thaliana* proteins from the obtained orthologous gene groups specific to species carrying *TNL*, *NRG1*, and *SAG101* were extracted to form a genomic annotation file. These sequences were used as BLASTp queries against the genomic protein sequences of all 63 species that had simultaneously lost *TNL*, *NRG1*, and *SAG101* (E-value = 1e-10).

Acknowledgment

This work was supported by the National Natural Science Foundation of China (32070243 to Z.Q.S. and 31770245 to J.Q.C.). We thank LetPub (www.letpub.com) for its linguistic assistance and scientific consultation during the preparation of this manuscript.

Author Contributions

Z.Q.S. and J.Q.C. conceived and designed the research. Y.L. and Z.Q.S. obtained, analyzed and interpreted the data; Z.Z., Q.L., X.M.J., J.H.T. and Y.M.Z. participated in data analysis; Z.J. and D.C. constructed the database; Y.M.Z. and Q.W. participated in data discussion, Y.L. and Z.Q.S. drafted the manuscript; Z.Q.S. and J.Q.C. revised the manuscript.

References

- Adachi, H., et al. (2019). An N-terminal motif in NLR immune receptors is functionally conserved across distantly related plant species. *Elife* 8.
- Adlassnig, W., Peroutka, M., Lambers, H., and Lichtscheidl, I.K. (2005). The roots of carnivorous plants. *Plant Soil* 274:127-140.
- Anisimova, M., Gil, M., Dufayard, J.F., Dessimoz, C., and Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic biology* 60:685-699.
- Asai, S., et al. (2014a). Expression profiling during arabidopsis/downy mildew interaction reveals a highly-expressed effector that attenuates responses to salicylic acid. *PLoS Pathog* 10:e1004443.
- Asai, S., et al. (2014b). Expression Profiling during Arabidopsis/Downy Mildew Interaction Reveals a Highly-Expressed Effector That Attenuates Responses to Salicylic Acid. *Plos Pathogens* 10.
- Baggs, E.L., et al. (2020). Convergent Loss of an EDS1/PAD4 Signaling Pathway in Several Plant Lineages Reveals Coevolved Components of Plant Immunity and Drought Response. *Plant Cell* 32:2158-2177.
- Bai, J., et al. (2002). Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. *Genome Res* 12:1871-1884.
- Byng, J.W., et al. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc* 181:1-20.
- Castel, B., et al. (2018). Diverse NLR immune receptors activate defence via the RPW8-NLR NRG1. *New Phytol* 222:966-980.
- Castel, B., et al. (2019). Diverse NLR immune receptors activate defence via the RPW8-NLR NRG1. *New Phytol* 222:966-980.
- Chen, C., et al. (2020). TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol Plant* 13:1194-1202.
- Collier, S.M., Hamel, L.P., and Moffett, P. (2011). Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. *Mol Plant Microbe Interact* 24:918-931.
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-1797.
- Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26:680-682.
- Johnson, L.S., Eddy, S.R., and Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC bioinformatics* 11:431.
- Kaiser, B., Vogg, G., Furst, U.B., and Albert, M. (2015). Parasitic plants of the genus *Cuscuta* and their interaction with susceptible and resistant host plants. *Frontiers in plant science* 6.
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermini, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587-589.
- Kourelis, J., and van der Hoorn, R.A.L. (2018). Defended to the Nines: 25 Years of Resistance Gene Cloning Identifies Nine Mechanisms for R Protein Function. *Plant Cell* 30:285-299.
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution* 33:1870-1874.
- Lan, T., et al. (2017). Long-read sequencing uncovers the adaptive topography of a carnivorous plant genome. *Proc Natl Acad Sci U S A* 114:E4435-E4441.

- 1 Lapin, D., Bhandari, D.D., and Parker, J.E. (2020). Origins and Immunity Networking Functions of EDS1
- 2 Family Proteins. *Annual review of phytopathology* 58:253-276.
- 3 Lapin, D., et al. (2019). A Coevolved EDS1-SAG101-NRG1 Module Mediates Cell Death Signaling by
- 4 TIR-Domain Immune Receptors. *Plant Cell* 31:2430-2455.
- 5 Leushkin, E.V., Sutormin, R.A., Nabieva, E.R., Penin, A.A., Kondrashov, A.S., and Logacheva, M.D. (2013).
- 6 The miniature genome of a carnivorous plant *Genlisea aurea* contains a low number of genes
- 7 and short non-coding sequences. *BMC genomics* 14:476.
- 8 Li, F.W., et al. (2018). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat*
- 9 *Plants* 4:460-472.
- 10 Li, T., et al. (2019). The bHLH transcription factor gene *AtUPB1* regulates growth by mediating cell
- 11 cycle progression in *Arabidopsis*. *Biochem Biophys Res Commun* 518:565-572.
- 12 Littlejohn, G.R., Breen, S., Smirnoff, N., and Grant, M. (2020). Chloroplast immunity illuminated. *New*
- 13 *Phytol.*
- 14 Luo, S., et al. (2012). Dynamic Nucleotide-Binding Site and Leucine-Rich Repeat-Encoding Genes in the
- 15 Grass Family. *Plant Physiology* 159:197-210.
- 16 Ma, S., et al. (2020). Direct pathogen-induced assembly of an NLR immune receptor complex to form a
- 17 holoenzyme. *Science* 370.
- 18 Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W. (2003). Genome-wide analysis of
- 19 NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* 15:809-834.
- 20 Minh, B.Q., Nguyen, M.A., and von Haeseler, A. (2013). Ultrafast approximation for phylogenetic
- 21 bootstrap. *Mol Biol Evol* 30:1188-1195.
- 22 Mint-Evolutionary-Genomics-Consortium. (2018). Phylogenomic Mining of the Mints Reveals Multiple
- 23 Mechanisms Contributing to the Evolution of Chemical Diversity in Lamiaceae. *Mol Plant*
- 24 11:1084-1096.
- 25 Narusaka, M., et al. (2009). RRS1 and RPS4 provide a dual Resistance-gene system against fungal and
- 26 bacterial pathogens. *The Plant journal : for cell and molecular biology* 60:218-226.
- 27 Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective
- 28 stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and*
- 29 *evolution* 32:268-274.
- 30 Pan, Q., Wendel, J., and Fluhr, R. (2000). Divergent evolution of plant NBS-LRR resistance gene
- 31 homologues in dicot and cereal genomes. *Journal of molecular evolution* 50:203-213.
- 32 Parker, J.E., et al. (1997). The *Arabidopsis* downy mildew resistance gene *RPP5* shares similarity to the
- 33 toll and interleukin-1 receptors with N and L6. *Plant Cell* 9:879-894.
- 34 Pavlovic, A., and Mithofer, A. (2019). Jasmonate signalling in carnivorous plants: copycat of plant
- 35 defence mechanisms. *Journal of experimental botany* 70:3379-3389.
- 36 Qi, T., et al. (2018). NRG1 functions downstream of EDS1 to regulate TIR-NLR-mediated plant
- 37 immunity in *Nicotiana benthamiana*. *Proc Natl Acad Sci U S A* 115:E10979-E10987.
- 38 Qian, L.H., et al. (2017). Distinct Patterns of Gene Gain and Loss: Diverse Evolutionary Modes of
- 39 NBS-Encoding Genes in Three Solanaceae Crop Species. *G3-Genes Genom Genet*
- 40 7:1577-1585.
- 41 Renner, T., and Specht, C.D. (2013). Inside the trap: gland morphologies, digestive enzymes, and the
- 42 evolution of plant carnivory in the Caryophyllales. *Current opinion in plant biology*
- 43 16:436-442.
- 44 Saile, S.C., et al. (2020). Two unequally redundant "helper" immune receptor families mediate

- 1 Arabidopsis thaliana intracellular "sensor" immune receptor functions. PLoS biology
- 2 18:e3000783.
- 3 Shao, Z.Q., Xue, J.Y., Wang, Q., Wang, B., and Chen, J.Q. (2019). Revisiting the Origin of Plant NBS-LRR
- 4 Genes. Trends in plant science 24:9-12.
- 5 Shao, Z.Q., et al. (2016). Large-Scale Analyses of Angiosperm Nucleotide-Binding Site-Leucine-Rich
- 6 Repeat Genes Reveal Three Anciently Diverged Classes with Distinct Evolutionary Patterns.
- 7 Plant Physiol 170:2095-2109.
- 8 Shao, Z.Q., et al. (2014). Long-term evolution of nucleotide-binding site-leucine-rich repeat genes:
- 9 understanding gained from and beyond the legume family. Plant Physiol 166:217-234.
- 10 Shao, Z.Q., Zhang, Y.M., Wang, B., and Chen, J.Q. (2015). Computational Identification of
- 11 MicroRNA-targeted Nucleotide-Binding Site-Leucine-Rich Repeat Genes in Plants.
- 12 Bio-protocol 5.
- 13 Silva, S.R., et al. (2019). The Terrestrial Carnivorous Plant Utricularia reniformis Sheds Light on
- 14 Environmental and Life-Form Genome Plasticity. International journal of molecular sciences
- 15 21.
- 16 Song, Y., et al. (2017). Evolutionary Comparisons of the Chloroplast Genome in Lauraceae and Insights
- 17 into Loss Events in the Magnoliids. Genome Biol Evol 9:2354-2364.
- 18 Sun, G., et al. (2018). Large-scale gene losses underlie the genome evolution of parasitic plant Cuscuta
- 19 australis. Nature communications 9:2683.
- 20 Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of
- 21 progressive multiple sequence alignment through sequence weighting, position-specific gap
- 22 penalties and weight matrix choice. Nucleic Acids Res 22:4673-4680.
- 23 Thorogood, C.J., Bauer, U., and Hiscock, S.J. (2018). Convergent and divergent evolution in carnivorous
- 24 pitcher plant traps. New Phytol 217:1035-1041.
- 25 Tirnaz, S., et al. (2020). Resistance Gene Analogs in the Brassicaceae: Identification, Characterization,
- 26 Distribution, and Evolution. Plant Physiol 184:909-922.
- 27 van der Biezen, E.A., Freddie, C.T., Kahn, K., Parker, J.E., and Jones, J.D.G. (2002). Arabidopsis RPP4 is a
- 28 member of the RPP5 multigene family of TIR-NB-LRR genes and confers downy mildew
- 29 resistance through multiple signalling components. Plant Journal 29:439-451.
- 30 Vogel, A., et al. (2018). Footprints of parasitism in the genome of the parasitic flowering plant Cuscuta
- 31 campestris. Nature communications 9:2515.
- 32 Wan, L., et al. (2019). TIR domains of plant immune receptors are NAD(+)-cleaving enzymes that
- 33 promote cell death. Science 365:799-+.
- 34 Wang, J., et al. (2019). Reconstitution and structure of a plant NLR resistosome conferring immunity.
- 35 Science 364.
- 36 Wang, W., Feng, B., Zhou, J.M., and Tang, D. (2020). Plant immune signaling: Advancing on two
- 37 frontiers. J Integr Plant Biol 62:2-24.
- 38 Wang, Y., et al. (2012). MCSanX: a toolkit for detection and evolutionary analysis of gene synteny and
- 39 collinearity. Nucleic Acids Res 40:e49.
- 40 Wu, Z., et al. (2019). Differential regulation of TNL-mediated immune signaling by redundant helper
- 41 CNLs. New Phytol 222:938-953.
- 42 Xu, L., et al. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of
- 43 orthologous clusters across multiple species. Nucleic Acids Res 47:W52-W58.
- 44 Xu, S., Wang, J., Guo, Z., He, Z., and Shi, S. (2020). Genomic Convergence in the Adaptation to Extreme

1 Environments. Plant communications 1:100117.
2 Xue, J.Y., et al. (2021). Mitochondrial genes from 18 angiosperms fill sampling gaps for phylogenomic
3 inferences of the early diversification of flowering plants. Journal of Systematics and
4 Evolution.
5 Xue, J.Y., et al. (2020). Genome- Wide Analysis of the Nucleotide Binding Site Leucine-Rich Repeat
6 Genes of Four Orchids Revealed Extremely Low Numbers of Disease Resistance Genes. Front
7 Genet 10.
8 Yang, L., et al. (2017). Pseudomonas syringae Type III Effector HopBB1 Promotes Host Transcriptional
9 Repressor Degradation to Regulate Phytohormone Responses and Virulence. Cell Host
10 Microbe 21:156-168.
11 Zhang, J., Coaker, G., Zhou, J.M., and Dong, X.N. (2020). Plant Immune Mechanisms: From
12 Reductionistic to Holistic Points of View. Molecular Plant 13:1358-1378.
13 Zhang, Y.M., et al. (2016). Uncovering the dynamic evolution of nucleotide-binding site-leucine-rich
14 repeat (NBS-LRR) genes in Brassicaceae. J Integr Plant Biol 58:13.
15 Zhou, J.M., and Zhang, Y. (2020). Plant Immunity: Danger Perception and Signaling. Cell 181:978-989.
16 Zumajo-Cardona, C., Pabon-Mora, N., and Ambrose, B.A. (2018). Duplication and Diversification of
17 REPLUMLESS - A Case Study in the Papaveraceae. Frontiers in plant science 9:1833.
18
19
20

Figure legends

Figure 1. Construction of an *angiosperm NLR atlas* (ANNA) covering 91,291 *NLR* genes in 304 angiosperm genomes. A) Phylogenetic relationships among angiosperm orders, constructed based on the APG IV system of flowering plant classification (Byng et al., 2016). Circle size corresponds to the number of species in each order; circle segments reflect the distribution of species among families. B) Representative screenshots of the ANNA interface, showing the major functions of ANNA: browsing, downloading, and primary comparative analysis of *NLR* datasets of interest.

Figure 2. Variations in *NLR* gene number among angiosperm genomes. A) Phylogenetic relationships among the 305 angiosperm genomes included in this study, constructed following the APG IV system (Byng et al., 2016). The numbers of *NLR* genes in each genome are shown by the blue bars. B) Maximum fold-differences in *NLR* gene copy number within each angiosperm family. C) Boxplots showing *NLR* gene numbers in the genomes of the basal angiosperms (n=4), magnoliids (n=5), monocots (n=65), dicots (n=230), and Ceratophyllales (n=1). Gene numbers did not differ significantly among groups. D) Boxplots showing *NLR* gene numbers in the genomes of special lifestyle species (SL species, n=17) and non-special lifestyle species (NSL species, n=288). SL species includes aquatic (14 genomes), parasitic (3 genomes), and carnivorous (4 genomes, 2 of which are also aquatic) species.

Figure 3. Variations in *NLR* subclass abundance among angiosperms. A) Relative proportions of the three *NLR* subclasses in the angiosperm genomes included in this study. B) Spearman correlation analysis for gene numbers among different *NLR* subclasses. The correlation coefficient is shown at top right of each graph. **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

Figure 4. Absence of *TNL* genes and *TNL* signal transfer components in angiosperms. A) Heat map showing gene copy numbers for *TNL* and *TNL* signal transfer components (*RNL* and *EDSI* family members). Phylogenetic relationships among the 305 angiosperm genomes resolved to the order level were constructed following the APG IV system (Byng et al., 2016). Green circles indicate simultaneous

1 losses of *TNL* and all five signal transduction components. Red stars indicate
2 simultaneous losses of *TNL*, *NRG1*, and *SAG101*. B-D) Predicted histories of loss of
3 *TNL* and *TNL* signal transfer components in (B) the Magnolias, (C) the Lamiales, and
4 (D) the Ranunculales. Phylogenetic relationships among and within the three clades
5 were constructed based on (Byng et al., 2016;
6 Mint-Evolutionary-Genomics-Consortium, 2018; Song et al., 2017; Xue et al., 2021;
7 Zumajo-Cardona et al., 2018). E) Numbers of species lacking *TNL* and/or various
8 *TNL* signal transfer components. Black dots indicate gene loss, and gray dots indicate
9 gene presence.

10 **Figure 5. Convergent conservation of a *TNL* lineage in angiosperm species**
11 **lacking the *NRG1*-*SAG101* pathway.** A) Phylogenetic analysis of *TNL* genes in 27
12 species carrying *TNL* but not *NRG1* and/or *SAG101*. The orange box indicates a
13 conserved *TNL* lineage. Branch support values from SH-aLRT and UFBoot2 are
14 indicated at each node. B) Expanded view of the conserved lineage highlighted in
15 panel (A). C) Syntenic analysis of the conserved *TNL* lineage in *A. thaliana*, *S.*
16 *lycopersicum*, *A. trichopoda*, and *V. radiata*.

17 **Figure 6. Identification of novel *TNL* signal transduction components.** A) Gene
18 clusters present in five species carrying *TNL*, *NRG1* and *SAG101* but absent in five
19 species lacking *TNL*, *NRG1* and *SAG101*. B) Five of the 23 gene clusters were absent
20 in more than 50% of the 63 angiosperm species lacking *TNL*, *NRG1* and *SAG101*.
21 Empty circles indicate gene loss, while filled circles indicate gene presence. The bar
22 graph shows the percentages of genomes lacking each gene cluster. C) The bar plot
23 shows fold-changes in the expression levels of *A. thaliana* genes in the five clusters
24 after *Pseudomonas syringae* DC3000 infection. Data shown are from (Asai et al.,
25 2014a). Heat map, showing expression levels (Transcripts Per Million, TPM) of these
26 genes after *Hyaloperonospora arabidopsidis* infection. Data shown are from (Yang et
27 al., 2017).

28 **Figure 7. Schematic diagram showing the adaptive evolution of the *NLR* gene,**
29 **leading to ecological specialization and the deletion of signal transduction**
30 **components.** A) A reduction in *NLR* gene number is usually coupled with adaptations

1 to aquatic (blue branches), parasitic (yellow branch), and carnivorous (green branch)
2 lifestyles. *NLR* gene numbers in non-vascular plants are from Shao et al (2019). B) A
3 summary of the dependence of CNL- and TNL-triggered immunity on the EDS1 and
4 RNL proteins. C) A hypothesized mechanism of *TNL* loss in angiosperms. EDS1 and
5 RNL proteins are known to be important signal transduction components for most
6 *TNL* genes. The frequent loss of *SAG101* and *NRG1* may drive *TNL* reduction and
7 loss. However, a few *TNL* genes may be functionally independent of *SAG101* and
8 *NRG1*, and thus may be preserved in some species lacking a *SAG101*/*NRG1* pathway.
9 Additive loss of *EDS1* and *RNL* genes may cause total *NLR* number reduction.

10 Supplemental Information

11 **Figure S1.** Clustering of *NLR* genes from angiosperm species with A) large (> 1, 000)
12 or B) small (< 50) number of *NLR* genes by CD-hit. The sequence identity cut-off was
13 set as 0.8 and 0.9 respectively.

14 **Figure S2. Pairwise synteny analyses of five angiosperm species, showing that**
15 **small-scale gene losses contribute to the independent losses of *NRG1* and *SAG101***
16 **in angiosperms.** Genes are shown as pointed colored blocks, with the point indicating
17 gene direction, on the specified chromosome. Orthologs between two species are
18 linked by black lines. Black arrowheads show the numbers of genes not displayed in
19 the indicated intervals.

20 **Table S1.** Genomes used in this study.

21 **Table S2.** The gene numbers of *NLR* and *EDS1* families identified from the 305
22 angiosperms.

23 **Table S3.** Intraspecies syntenic blocks containing *NRG1* and *SAG101* genes.

24 **Table S4.** Intraspecies syntenic blocks containing genes from the conserved *TNL*
25 lineage.

26 **Table S5.** The 23 orthogroups specific to five *TNL*-absent species identified by
27 OrthoVenn2.

28 **Table S6.** Expression change of potential *TNL* signal transduction components in *A.*
29 *thaliana* after infection by *TNL*-recognized pathogens.

30 **Table S7.** *NLR* genes identified from two ferns' genome.

A

B

C

D

Figure 2. Variations in NLR gene number among angiosperm genomes. A) Phylogenetic relationships among the 305 angiosperm genomes included in this study, constructed following the APG IV system (Byng et al., 2016). The numbers of NLR genes in each genome are shown by the blue bars. B) Maximum fold-differences in NLR gene copy number within each angiosperm family. C) Boxplots showing NLR gene numbers in the genomes of the basal angiosperms (n=4), magnoliids (n=5), monocots (n=65), dicots (n=230), Ceratophyllales (n=1). Gene numbers did not differ significantly among groups. D) Boxplots showing NLR gene numbers in the genomes of special lifestyle species (SL species, n=17) and non-special lifestyle species (NSL species, n=288). SL species includes aquatic (14 genomes), parasitic (3 genomes), and carnivorous (4 genomes, 2 of which are also aquatic) species.

Figure 3. Variations in *NLR* subclass abundance among angiosperms. A) Relative proportions of the three *NLR* subclasses in the angiosperm genomes included in this study. B) Spearman correlation analysis for gene numbers among different *NLR* subclasses. The correlation coefficient is shown at top right of each graph. **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

Figure 4. Absence of TNL genes and TNL signal transfer components in angiosperms. A) A heat map showing gene copy numbers for TNL and TNL signal transfer components (NRG1 and EDS1 family members). Phylogenetic relationships among the 305 angiosperm genomes resolved to the order level were constructed following the APG IV system (Byng et al., 2016). Green circles indicate simultaneous losses of TNL and all five signal transduction components. Red stars indicate simultaneous losses of TNL, NRG1, and SAG101. B-D) Predicted histories of loss of TNL and TNL signal transfer components in (B), the Magnolias, (C) the Lamiales, and (D) the Ranunculales. Phylogenetic relationships among and within the three clades were constructed based on (Byng et al., 2016; Song et al., 2017; Mint-Evolutionary-Genomics-Consortium, 2018; Zumajo-Cardona et al., 2018; Xue et al., 2021). E) Numbers of species lacking TNL and/or various TNL signal transfer components. Black dots indicate gene loss, and gray dots indicate gene presence.

A

C

Figure 5. Convergent conservation of a *TNL* lineage in angiosperm species lacking the *NRG1*-*SAG101* pathway. A) Phylogenetic analysis of *TNL* genes in 27 species carrying *TNL* but not *NRG1* and/or *SAG101*. Branch support values from SH-aLRT and UFBoot2 are indicated at each node. The orange box indicates a conserved *TNL* lineage. B) Expanded view of the conserved lineage highlighted in panel (A). C) Syntenic analysis of the conserved *TNL* lineage in *A. thaliana*, *S. lycopersicum*, *A. trichopoda*, and *V. radiata*.

A

C

Figure 6. Identification of novel TNL signal transduction components. A) Gene clusters present in five species carrying TNL, NRG1 and SAG101 but absent in five species lacking TNL, NRG1 and SAG101. B) Five of the 23 gene clusters were absent in more than 50% of the 63 angiosperm species lacking TNL, NRG1 and SAG101. Empty circles indicate gene loss, while filled circles indicate gene presence. The bar graph shows the percentages of genomes lacking each gene cluster. C) The bar plot shows fold-changes in the expression levels of *A. thaliana* genes in the five clusters after *Pseudomonas syringae* DC3000 infection. Data shown are from (Asai et al., 2014). Heat map showing expression levels (Transcripts Per Million, TPM) of these genes after *Hyaloperonospora arabidopsidis* infection. Data shown are from (Yang et al., 2017).

A

B

C

Figure 7. Schematic diagram showing the adaptive evolution of the NLR gene leading to ecological specialization and the deletion of signal transduction components. A) A reduction in NLR gene number is usually coupled with adaptations to aquatic (blue branches), parasitic (yellow branch), and carnivorous (green branch) lifestyles. NLR gene numbers in non -vascular plants are from Shao et al (2019). B) A summary of the dependence of CNL- and TNL-triggered immunity on the EDS1 and RNL proteins. C) A hypothesized mechanism of TNL loss in angiosperms. EDS1 and RNL proteins are known to be important signal transduction components for most TNL genes. The frequent loss of SAG101 and NRG1 may drive TNL reduction and loss. However, a few TNL genes may be functionally independent of SAG101 and NRG1, and thus may be preserved in some species lacking a SAG101/NRG1 pathway. Additive loss of EDS1 and RNL genes may cause total NLR number reduction.