Evolutionary Patterns in Plant NLR Immune Receptor Structure and Function

AmirAli Toghani

Thesis submitted for the Degree of Doctor of Philosophy

The Sainsbury Laboratory

University of East Anglia

September 2026

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

Plant nucleotide-binding leucine-rich repeat (NLR) immune receptors detect pathogen effectors in the cell and trigger immune responses that lead to cell death. Within the superasterid NRC (NLR Required for Cell death) network, sensor NLRs detect pathogens but need NRC helpers to mediate immune signaling and confer disease resistance. NLR evolution has been extensively studied in the context of co-evolution with pathogen effector. However, the degree to which plant NLR evolution is shaped by other selection pressures remains poorly understood.

This thesis studies the relationships between NLRs themselves and the selective pressures that have shaped these NLR–NLR interactions. I investigated the evolutionary dynamics of the NRC network and found that the interfaces mediating helper homodimerization in the resting state have diverged between paralogous clades but remain conserved within lineages. This pattern is consistent with selection acting to keep the signaling pathways of paralogous helpers insulated from one another.

Through comparative evolutionary analysis, I found that sensor NLR diversification rate correlates with helper dependency: sensors dependent on recently derived helpers have diversified more rapidly than those reliant on ancient helpers. These findings indicate that evolutionary pressures beyond direct co-evolution with pathogen effectors act on NLRs to maintain immune network dynamics and robust signaling.

I also showed that AlphaFold 3 can model NLR resistosomes with high confidence and accuracy. The models reproduced experimentally determined structures and resolved the membrane-inserting N-terminal α1 helix, a region frequently missing from cryo-EM structures. Helper NLRs were modeled with higher confidence than their cognate sensors across several sensor–helper systems, including monocot NLR pairs and the NRC network. This supports the hypothesis that sensor NLR subfunctionalization involves loss of the capacity to oligomerize, and provides a structural criterion, complementary to sequence-based methods, for classifying sensor and helper NLRs to help determine genetic dependencies and assist crop breeding.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate ‘take down’ action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Acknowledgments

I would like to start by thanking my supervisor, Sophien, for his continued support and encouragement. Throughout my time at TSL, you always provided a sea of ideas and were never shy about discussing science with me, even if it meant me dropping by your office literally every day. Thank you for being the best mentor I could have ask for and for nurturing an environment that allowed me to do my research with the utmost joy. I would also like to thank my secondary supervisor, Phil Carella. Thank you, Phil, for your insights and support before and during my PhD. My sincere thanks also go to my undergraduate supervisors and mentors, Professors Nematollah Etemadi, Mahdi Gheysari, Aghafakhr Mirlohi, and Rahim Mehrabi. Without your encouragement and guidance, I could not have reached this point. I also want to thank Sebastian Schornak for his support and career advice over the years.

Being part of the "Kamounity" has been a delightful experience and a true privilege. Throughout the years, I worked with many talented colleagues and friends whom I learned so much from; each of you shaped my PhD and predoc journey. Thank you to everyone who makes this place so magical. I would also like to thank all my collaborators over the years, including Vivianne Vleeshouwers , Tolga Bozkurt, Chih-Hang Wu, Wenbo Ma, Hiroaki Adachi, and Suomeng Dong, for the countless opportunities to work on exciting topics, learn, and contribute to their research.

I would like to thank my first mentors at TSL: thank you, Daniel and Jiorgos, for teaching me the basics and helping me find my feet in the lab. Thank you, Mau, for welcoming me to Norwich and for being such a close friend and collaborator throughout. Andy, Yu, and Pai, many thanks for several years of friendship and teamwork. I have always treasured your input and support.

My work would not have been possible without Adeline and Joe. Thank you for keeping the lab running and keeping us sane. I also greatly appreciate the assistance from all the TSL support teams, especially the bioinformatics team and Dan MacLean.

Diana, thank you for passing the torch after your viva and for being my pal throughout my PhD. Josh and Alex, I will always cherish our time together and having an "elite spine" for competitiveness, from football and bowling to CCBL. Thank you, Sara, for making me feel so welcome in Norwich and for being there whenever I needed. Thank you, Andy, for being my PhD buddy since day one. Thank you, Jennie, for being the best company during my thesis writing period.

I am so grateful for the Room 200 student office and the TSLytherins. I could not imagine

a better student community. I thoroughly enjoyed spending time with you all: Mauricio Contreras, Andres Posbeyikian, Sara Dorhmi, Alexander McClelland, Josh Bennett, Yufei Li, Jennie Yang, Lena Knorr, Attilio Pascucci, Diana Gómez De La Cruz, Barbara Schrammel, Juan Azcue, Maián Garro, Yordan Dolaptchiev, Camilla Molinari, and Matthew Wengler. A special thank you to the ICPP gang. Thank you, Nef, Lauren, and Diana, for being part of our 100%.

Thank you, Emma, Petra, and Rozhan for all the weekend teas and walks. Thank you to my housemates and friends throughout this journey, from Gloucester St. with Josh, Maialen, and Yu, to Whitehall Rd. with James. Thank you for all the nighttime science discussions and banter.

Thank you again, Daniel, Jennie, Alex, and Sophien, for reading through my thesis and providing invaluable feedback.

I would also like to extend this moment and my joy to my friends who could not be here, whom I have not seen for more than five years: thank you, Maziar and Roozbeh. I miss you both and wish you were here so we could celebrate together.

Lastly, Maman, Baba, I love you and miss you. This has all been possible because of your sacrifices and love. This is for you. Shirin and Shekoofeh, thank you for always having my back and being the best sisters ever. AmirHossein, Mohammad, Mahour, and Mayha, I miss you all so much too. I hope to see you very soon.

Contents

Abstract 1

Acknowledgments 2

1 Chapter 1: General Introduction 8

1.1 An overview of plant immunity 8

1.2 NLR proteins: architecture and function 9

1.2.1 Domain architecture and the STAND superfamily 9

1.2.2 NLRs as molecular switches 9

1.2.3 Classes of plant NLRs defined by the N-terminal domain 10

1.2.4 Integrated domains and the integrated decoy model 11

1.3 NLR singletons, pairs, and networks 12

1.3.1 Functional singletons 12

1.3.2 Paired NLRs 13

1.3.3 Immune receptor networks 14

1.3.4 Modulators and atypical NLRs 15

1.4 The N-terminal α1 helix and the MADA motif 15

1.5 Mechanisms of NLR activation 17

1.5.1 The ZAR1 resistosome and the pentameric paradigm 17

1.5.2 Sr35: the second pentameric resistosome 18

1.5.3 Activation in the NRC network 19

1.5.4 CC_{G10}-NLRs and the octameric resistosome 20

1.5.5 TIR-NLR and CC_R-NLR signaling 21

1.6 Oligomerization and cell death across the domains of life 24

1.6.1 Nucleotide-dependent oligomerization as a shared signaling principle 24

1.6.2 Executing cell death: convergent routes to membrane permeabilization 26

1.7 The evolution of NLR immune receptors 27

1.7.1 The classical view: co-evolution with pathogen effectors 27
1.7.2 Emergence of networks from singleton ancestors 27
1.7.3 Genomic organization and the mechanics of NLR diversification 28
1.7.4 Asymmetric evolution of sensors and helpers 29
1.7.5 Evolutionary pressures beyond R–AVR co-evolution 29
1.8 Genome-scale resources for studying NLR evolution 30
1.8.1 The expansion of sequenced plant genomes 30
1.8.2 Why NLRs are particularly affected 30
1.8.3 Deep-learning gene annotation 31
1.9 Structural biology and computational modeling of NLRs 32
1.9.1 What experimental structures have revealed 32
1.9.2 Protein structure prediction and its extension to complexes . . 32
1.9.3 Applications to plant immunity and their limits 33
1.10 Aims and thesis outline 34

2 Chapter 2: Curation of harmonized plant NLRome datasets on multiple taxonomic resolutions 36

2.1 Introduction 36
2.2 Functional annotation of 180 RefSeq reference plant proteomes reveals a dataset of 113,684 NLR proteins 39
2.3 Deep-learning-based annotation of 230 superasterid genomes reveals a harmonized dataset of 91,366 NLRs 41
2.4 Harmonized genome annotation of 346 Solanaceae genomes reveals 197,834 NLR immune receptors 44
2.5 Conclusions and Discussion 48

3 Chapter 3: Functional and molecular evolution of the NLR proteins beyond R-AVR co-evolution 49

3.1 Introduction 49
3.2 The N-terminal MADA motif has diversified in NRC helpers 52
3.2.1 NRCX carries a signature threonine at the fourth position of its MADA motif 53
3.2.2 A subset of NRC4 sequences share the NRCX N-terminal signature 53

3.2.3 NRC4t proteins induce cell death as autoactive mutants 54
3.3 Resting-state NRC2 forms a homodimer 55
3.4 Evolution of homodimerization in NRCs 57
3.4.1 The dimerization interfaces have diverged between NRC clades 59
3.4.2 Entropy analysis reveals clade-specific variation in NRC homodimerization stretches 59
3.5 Conclusions and Discussion 61
3.6 Supplementary figures 64
3.7 Tables 66

4 Chapter 4: Contrasting patterns of evolution in sensor and helper NLRs 67

4.1 Introduction 67
4.2 The NRC network has expanded less in Asterales relative to Solanales . . 69
4.3 Asterales NRC networks show lineage-specific expansion and variable genomic clustering 70
4.4 LsatNRC0 and Ast-LsatNRC1 signal for distinct sensor groups 72
4.5 NRC-S dependent on Ast-LsatNRC1 are under stronger positive selection 74
4.6 Conclusions and discussion 76
4.7 Supplementary figures 82

5 Chapter 5: Computational structural modeling of NLR resistosomes reveals structural diversity in plant CC-type NLRs 94

5.1 Introduction 94
5.2 AlphaFold 3 predicts a lipid-bound NRC2 resistosome with high confidence 95
5.3 Model confidence is highest for the hexameric stoichiometry 96
5.4 AtZAR1 and TmSr35 are predicted with higher confidence as pentamers 98
5.5 Modeling across the NRC helper clade 98
5.6 CC-NLR resistosomes show a diversity of predicted stoichiometries and N-terminal pores 100
5.7 Conclusions and discussion 101
5.8 Supplementary figures 103

6 Chapter 6: Using AI-based modeling to distinguish between Sensor and Helper NLRs 117

6.1 Introduction 117

6.2 Helper NLRs produce higher confidence scores than their paired sensors when modeled as resistosomes 118

6.3 Structure-based classification outperforms the MADA motif as a determinant of sensor/helper NLR predictions in paired NLRs 119

6.4 Structure-based classification can be used to predict putative paired NLRs in rice 119

6.5 Validation on an experimentally characterized wheat NLR pair 121

6.6 Extension to a dicot NLR network 121

6.7 Conclusions and discussion 123

6.8 Supplementary figures 126

7 Chapter 7: General Discussion 136

7.1 Diversification structured by internal partner compatibility 136

7.2 Selection pressures beyond R-AVR co-evolution 138

7.3 Subfunctionalization and the loss of the ability to oligomerize 139

7.4 How many protomers in a resistosome? 140

7.5 Outlook 141

8 Chapter 8: Materials and Methods 143

Abbreviations 153

List of Figures 158

List of Tables 161

References 162

Appendix I: Publications and datasets 198

Chapter 1: General Introduction

1.1 An overview of plant immunity

Plants are constantly engaged in biotic interactions with their environment, many of which involve pathogenic and parasitic organisms such as fungi, bacteria, oomycetes, and nematodes. To counter these threats, plants have evolved a multi-tiered immune system that protects them against most potential pathogens. This immune system is traditionally conceptualized as two layers of defense. First, upon pathogen contact, plant cell-surface receptors recognize conserved microbe- or pathogen-associated molecular patterns (MAMPs/PAMPs) to mediate an initial defense response named pattern-triggered immunity (PTI). These plasma membrane-associated pattern recognition receptors (PRRs) belong to distinct protein families, most notably leucine-rich repeat receptor-like kinases (LRR-RLKs) and leucine-rich repeat receptor-like proteins (LRR-RLPs) (Dievart et al., 2020; Ngou et al., 2024).

Successful pathogens, however, have evolved to evade or suppress this first line of defense by delivering effector proteins into host cells, leading to effector-triggered susceptibility (ETS) (Ngou et al., 2022). In response, plants evolved intracellular receptors like nucleotide-binding leucine-rich repeat receptors (NLRs) to detect these pathogen-derived effectors (Contreras et al., 2023a). Effector recognition by NLRs activates effector-triggered immunity (ETI), which is often accompanied by a localized form of programmed cell death known as the hypersensitive response (HR) (Jones et al., 2016). NLRs constitute this second layer of plant immunity, and most cloned disease resistance (R) genes in crops belong to this superfamily (Kourelis and Van Der Hoorn, 2018).

Historically, PTI and ETI were viewed as distinct, sequentially acting tiers functioning in a dynamic "zig-zag" model (Jones and Dangl, 2006). Recent studies, however, demonstrate that these two pathways are intimately connected, mutually potentiating one another and sharing overlapping signaling networks (Ngou et al., 2021; Yuan et al., 2021; Pruitt et al., 2021; Tian et al., 2021). In addition, PRRs that cause cell death appear to execute this response via NLRs (Kourelis et al., 2022). In this thesis, I focus primarily on intracellular NLR immune receptors, reviewing our current understanding of their function, structural diversity, and evolution.

1.2 NLR proteins: architecture and function

1.2.1 Domain architecture and the STAND superfamily

NLR proteins are found across all kingdoms of life and share a tripartite domain architecture: an N-terminal variable signaling domain, a central nucleotide-binding and oligomerization domain (NOD), and C-terminal superstructure-forming repeats (SSFRs) (Duxbury et al., 2021; Uehling et al., 2017; Dyrka et al., 2020). NLRs belong to the STAND (signal transduction ATPases with numerous domains) superfamily, a distinct lineage of P-loop NTPases that includes key animal and plant regulators of programmed cell death (Leipe et al., 2004). The N-terminal domains typically mediate downstream signaling; in plants, this often culminates in a localized cell death response upon receptor activation (Dyrka et al., 2020; Kourelis et al., 2021). In plant NLRs, the central NOD is specifically an NB-ARC (nucleotide-binding adaptor shared by APAF-1, certain R gene products, and CED-4) domain, while the C-terminal SSFRs are predominantly leucine-rich repeat (LRR) modules (Duxbury et al., 2021; Kourelis et al., 2021).

Although this modular architecture is conserved across kingdoms, the family has diversified far more extensively in plants than in other lineages. Animal NLRs constitute a comparatively small, conserved set of cytosolic sensors that typically assemble into inflammasomes (Duxbury et al., 2021). NLR proteins functioning via analogous signaling logic have also been characterized in fungi and bacteria, where they frequently link to amyloid-forming motifs and cell death-associated domains (Uehling et al., 2017; Dyrka et al., 2020; Daskalov et al., 2016). Plant genomes, by contrast, frequently encode hundreds of diverse NLRs, with repertoire sizes varying by more than an order of magnitude across species (Barragan and Weigel, 2021; Shao et al., 2016; Toghani and Kamoun, 2024; Toghani et al., 2025b; Toghani et al., 2026c). This massive lineage-specific expansion, coupled with high sequence divergence and domain modularity, makes plant NLRs both a compelling model for evolutionary studies and a notoriously difficult target for automated annotation, a challenge addressed in detail in Section 1.8.

1.2.2 NLRs as molecular switches

Like most STAND proteins, plant NLRs function as molecular switches (Takken et al., 2006). In the absence of a pathogen, they are held in an inactive, ADP-bound state, activating immune signaling upon detection of foreign or altered-self elements. The central NB-ARC domain plays a crucial role in this process, facilitating

the transition between inactive and active states through the exchange of ADP for ATP (Wang et al., 2019b). While the C-terminal LRR sometimes determines pathogen perception, it also contributes to maintaining the receptor in its inactive state through autoinhibitory interactions. Once the receptor is activated, these inhibitory contacts are released, allowing the N-terminal domain to initiate downstream immune signaling (Förderer et al., 2022; Takken and Goverse, 2012).

The nucleotide-binding pocket that governs this switch is defined by a set of conserved motifs across the NB-ARC domain. These include the P-loop, which coordinates the bound nucleotide, and the MHD motif, which its aspartate makes contacts the ADP β-phosphate in the resting state (Takken et al., 2006; Takken and Goverse, 2012). These motifs have proven both biologically important and experimentally useful. Substitution of the P-loop lysine abolishes nucleotide binding and renders most NLRs inactive, whereas substitution of the MHD aspartate destabilizes the autoinhibited conformation, yielding autoactive receptors that trigger cell death in the absence of pathogen recognition (Wang et al., 2019b; Adachi et al., 2019a). Autoactive MHD mutants are used throughout this thesis to test whether a given receptor is competent to execute cell death independently of upstream activation partners.

1.2.3 Classes of plant NLRs defined by the N-terminal domain

Plant NLRs are classified into distinct groups based on their N-terminal signaling domains, which align with the phylogeny of the NB-ARC domain and point to a deep evolutionary origin for this partitioning. In angiosperms, four main types of N-terminal signaling domains have been described: coiled-coil (CC), RESISTANCE TO POWDERY MILDEW 8 (RPW8)-type (CC_R), G10-type CC (CC_{G10}), and Toll/interleukin-1 receptor-type (TIR) (Kourelis et al., 2021). Certain plant lineages like non-flowering plants appear to possess additional types, such as α/β hydrolase and kinase domains (Chia et al., 2022; Andolfo et al., 2019). The N-terminal domain is thought to determine both the nature of the downstream signaling pathway and the specific execution events that follow NLR activation. Throughout this thesis, I refer to CC, CC_R, and CC_{G10} NLRs collectively as CC-type NLRs.

The four classes are unevenly distributed across the plant phylogeny and differ in their functional roles. CC-NLRs are the most phylogenetically widespread and numerically abundant class in angiosperm genomes, comprising both functional singletons and the helper receptors that execute cell death in NLR networks (Shao et al., 2019; Kourelis et al., 2021). TIR-NLRs are abundant in eudicots but have been lost from monocot lineages, with their signaling routed through a dedicated downstream module rather than direct membrane insertion (Lapin et al., 2019; Liu et al., 2023b).

CC_R-NLRs form a small, highly conserved group that functions almost exclusively as helpers (Saile et al., 2020; Jacob et al., 2021), while CC_{G10}-NLRs are a comparatively poorly characterized class that forms a distinct branch in NB-ARC phylogenies (Kourelis et al., 2021). Because these assignments rest on phylogeny and on the detection of N-terminal domains that are often weakly conserved at sequence level, automated classification can be ambiguous; consequently, a proportion of receptors in any large-scale annotation remain unresolved among the CC-type classes.

1.2.4 Integrated domains and the integrated decoy model

In many plant species, around 3.5% to 10% of NLR proteins deviate from the canonical tripartite architecture by carrying additional integrated domains (IDs), as exemplified by the rice receptors Pik-1 and RGA5 (Cesari et al., 2014; Kroj et al., 2016; César et al., 2014). These IDs often resemble host effector targets and play a central role in pathogen detection, either through direct or indirect interactions (Maqbool et al., 2015; Fujisaki et al., 2015; Gu et al., 2023; De La Concepcion et al., 2018). This concept was formalized as the integrated decoy hypothesis, which proposes that plants have genetically incorporated effector targets within NLRs over evolution to detect pathogens and activate host immunity efficiently (Cesari et al., 2014).

Systematic surveys have since shown that the repertoire of integrated domains is broad and evolutionarily labile. Hundreds of distinct domain families have fused to NLRs across the angiosperms, including heavy metal-associated (HMA), WRKY, BED, kinase, and exocyst subunit domains, with identical domain families being integrated independently across divergent lineages (Sarris et al., 2016; Kroj et al., 2016; Marchal et al., 2018; Shao et al., 2024). The presence of an ID is informative regarding both function and recognition mechanism. Because IDs act in perception rather than downstream execution, they are mostly found in sensor NLRs and are typically absent from helpers. Detecting an ID therefore provides one of the few sequence-based criteria available for hypothesizing sensor and helper functions (Białas et al., 2018; Marchal et al., 2018). However, this criterion remains incomplete, as many sensors carry no integrated domain at all. Moreover, it is not clear whether all NLRs with IDs are sensors and rely on other components for signaling. These diagnostic limitations motivate the structural and predictive approaches developed in Chapter 6.

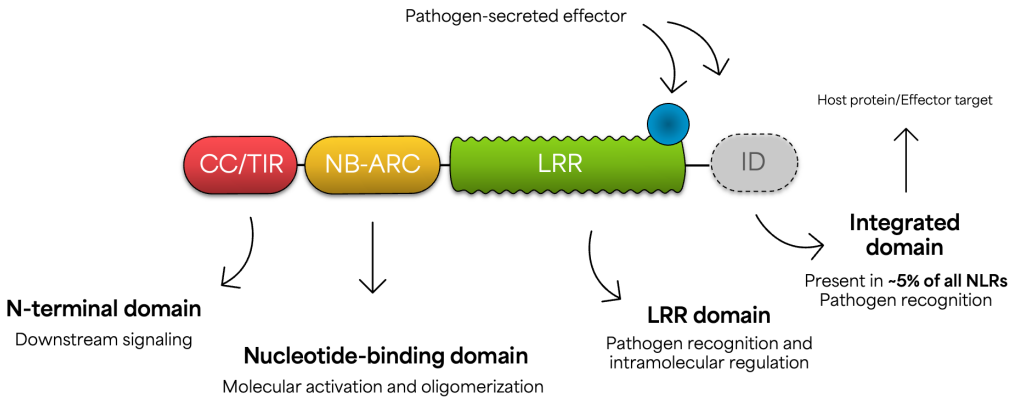


Figure 1.1: Domain architecture of plant NLR immune receptors.
Canonical plant NLRs have a tripartite architecture comprising an N-terminal signaling domain, most commonly a coiled-coil (CC) or Toll/interleukin-1 receptor (TIR) domain, which mediates downstream signaling; a central NB-ARC domain, which acts as a molecular switch governing activation and oligomerization; and a C-terminal leucine-rich repeat (LRR) domain, which contributes both to pathogen recognition and to intramolecular regulation of the resting state. Pathogen-secreted effectors may be detected at the LRR, either directly or indirectly. Around 3.5% to 10% of the NLR family, carry an additional integrated domain (ID). IDs typically resemble host proteins targeted by pathogen effectors and function in recognition.

1.3 NLR singletons, pairs, and networks

1.3.1 Functional singletons

NLRs can function in distinct configurations: as functional singleton NLRs, paired NLRs, or complex NLR networks (Figure 1.2A) (Adachi et al., 2019b; Feehan et al., 2020; Gong et al., 2023). Functional singleton NLRs operate as single genetic units, detecting pathogen effectors and initiating immune signaling without the requirement of additional NLR partners (Adachi et al., 2019b; Adachi et al., 2019a). Well-studied examples include ZAR1 from *Arabidopsis thaliana* and Sr35 from wheat. ZAR1 does not recognize pathogen effectors directly; instead, it detects them indirectly through its associated receptor-like cytoplasmic kinases (RLCKs), which act as guardees or decoys (Wang et al., 2019a; Adachi et al., 2023b). In contrast, Sr35 binds its cognate effector AvrSr35 directly via its LRR domain (Förderer et al., 2022). Additional NLRs, such as Sr50 (Chen et al., 2017), various members of the barley MLA allelic series (Brabham et al., 2024; Maekawa et al., 2019; Saur et al., 2019;

Gómez De La Cruz et al., 2026), and RPS5 (Qi et al., 2012; Pottinger and Innes, 2020), are likewise considered singletons based on their ability to sense effectors and trigger hypersensitive cell death across diverse plant species.

A widely used method to characterize these NLRs is agroinfiltration, in which immune receptors are transiently expressed in the leaves of *Nicotiana benthamiana* or other amenable plant species to assess their function (Derevnina et al., 2019). This assay is particularly valuable for testing whether an NLR from another plant species can autonomously trigger cell death in *N. benthamiana* (Derevnina et al., 2019), which provides initial evidence for singleton or helper activity. However, this assay is not definitive on its own, as the outcome depends on the conservation of downstream signaling partners as well as adequate expression and correct subcellular localization. A receptor that fails to induce cell death in this heterologous system may be a dedicated sensor lacking its partner helper, a helper whose downstream components have diverged, or simply a poorly expressed protein. For this reason, the assay is most informative when combined with "autoactive" MHD mutants to test intrinsic cell death execution, and especially when performed in genetic backgrounds where candidate downstream potential helpers have been knocked out.

1.3.2 Paired NLRs

Paired NLR systems operate through division of labor, comprising a sensor NLR dedicated to effector detection and a helper NLR that executes downstream immune signaling. Notable examples of paired NLRs include the rice NLR pairs Pik-1/Pik-2 and RGA5/RGA4 (Maqbool et al., 2015; Gong et al., 2023; Zdrzałek et al., 2020; De La Concepcion et al., 2021), alongside the *Arabidopsis* TIR-NLR pair RPS4/RRS1 (Sarris et al., 2015; Saucet et al., 2015). In these systems, pathogen recognition by the sensor NLR is communicated to the helper NLR, which then activates immune responses.

The two genes in a pair are typically encoded adjacent to each other in a head-to-head orientation, sharing a common promoter region, with the sensor frequently carrying an integrated domain (Kroj et al., 2016; Sarris et al., 2016). The regulatory mechanism transmitting signals between them varies across pairs. In the RGA5/RGA4 pair, the helper RGA4 is constitutively active and held in an autoinhibited state by the sensor RGA5, such that effector binding relieves negative regulation (Césari et al., 2014). For Pikp-1/Pikp-2, by contrast, the evidence supports cooperative activation of the helper by the sensor upon effector binding (Zdrzałek et al., 2020; Dong et al., 2018; Pai et al., 2025a). Because paired receptors often originate from distinct phylogenetic clades while functioning as a unit, they are more difficult to classify on purely phylogenetic criteria than the known NLR networks, making the identification of sensor and helper partners in candidate pairs a

persistent challenge (Kourelis et al., 2021; Contreras et al., 2023a).

1.3.3 Immune receptor networks

Plant NLRs can also organize into interconnected networks that extend beyond isolated sensor–helper pairs (Adachi et al., 2019b; Wu et al., 2018). These networks encompass multiple, often genetically unlinked NLRs, featuring flexible architectures with "one-to-many" or "many-to-one" connections. Multiple sensor NLRs can converge onto a single downstream helper, while individual sensors may also signal through multiple helpers. This architecture provides evolutionary flexibility and robustness to the plant immune system, with helpers exhibiting functional specialization in sensor compatibility and execution dynamics (Adachi et al., 2019b; Wu et al., 2018).

A prominent example of an NLR network occurs in superasterid plants, where NRCs (NLRs required for cell death) serve as central helper nodes (NRC-H). NRC helpers mediate immune responses against a diverse spectrum of pathogens and pests, including oomycetes, fungi, nematodes, insects, viruses, and bacteria (Wu et al., 2017; Derevnina et al., 2021). The sensors that depend on these helpers (NRC-S) fall into two phylogenetic subtypes. Rx-type NRC-S, which include the potato virus X resistance protein Rx, the pepper bacterial spot resistance protein Bs2, and Rpi-amr3, retain the canonical tripartite NLR architecture with occasional integration of IDs in their scaffold (Contreras et al., 2022; Ahn et al., 2023; Lin et al., 2021). The second subclass, named SD-type NRC-S such as Prf, Sw5-b, Hero, and Rpi-blb2 typically carry an additional N-terminal extension termed the Solanaceous domain (SD) (Seong et al., 2020; De Oliveira et al., 2016). The function of the SD is not fully resolved, although in the tomato sensor Prf it provides binding sites for the Pto kinase as well as mediating homotypic interactions (Saur et al., 2015). SD-type NRC-S can also carry integrated domains such as HMA IDs in potato R1 clade sensors (Wang et al., 2026).

The network features functional redundancy, as several sensors can signal through more than one helper. While some helpers can act upstream of numerous sensors, others maintain restricted compatibility (Wu et al., 2017; Goh et al., 2024; Lüdke et al., 2025). This redundancy provides an adaptive defense advantage: when a pathogen effector suppresses a specific helper node, sensors capable of signaling through alternative uninhibited helpers preserve immune output (Derevnina et al., 2021). Similarly, in an independent network, the CC_R-NLRs NRG1 and ADR1 function as helpers that execute immune responses downstream of small-molecule signals generated by diverse TIR-type sensor NLRs (Castel et al., 2019; Feehan et al., 2023; Qi et al., 2018).

1.3.4 Modulators and atypical NLRs

Beyond standard sensor and helper categories, NLRs can also modulate the activity of other immune receptors. This was first recognized in genetically linked sensor–helper pairs where the sensor suppresses an otherwise autoactive helper, as seen with RGA5/RGA4. More recent studies have identified atypical NLRs that function as dedicated modulators of immune signaling outside of canonical sensor or helper roles (Cesari et al., 2014; Ma et al., 2018).

NRCX, a full-length CC-NLR closely related to the helpers NRC2 and NRC3, is a representative example. Its N-terminal MADA motif is non-functional, and silencing NRCX in *N. benthamiana* triggers spontaneous autoimmunity that depends specifically on NRC2 and NRC3, but not NRC4. NRCX therefore acts as a negative regulator of NRC2 and NRC3, likely having evolved to prevent inappropriate activation of these helper pathways in the absence of pathogens (Adachi et al., 2023a). NRCX highlights how atypical NLRs can act as network modulators (Adachi et al., 2023a; Wu et al., 2022). Another example is truncated NLR variants such as NRG1C, which antagonizes immunity mediated by its full-length counterparts NRG1A and NRG1B (Wu et al., 2022). The existence of such modulators indicates that complex network configurations evolved regulatory mechanisms to balance immune output and prevent autoimmunity.

The known network configurations are non-randomly distributed across the plant NLR phylogeny. Within a given network, NRC-S and NRC-H belong to distinct, monophyletic clades. In the superasterids, the NRC superclade, comprising the NRC-H alongside their dependent Rx-type and SD-type NRC-S, forms a well-supported lineage nested within the broader CC-NLR clade (Figure 1.2B).

1.4 The N-terminal α1 helix and the MADA motif

The MADA motif, located in the N-terminal α1 helix of about 20% of all angiosperm CC-NLRs, is required for the induction of cell death (Adachi et al., 2019a). This motif, along with related sequences such as the MAEPL motif found in non-flowering plants, is functionally conserved across a wide range of evolutionarily distant NLRs capable of triggering cell death (Chia et al., 2022; Chia and Carella, 2023). NLRs carrying an N-terminal MADA motif include diverse receptors, from the ancient singleton ZAR1 to helper NLRs like Pik-2 and NRC4 (Adachi et al., 2019a). This points to a shared origin for these CC-NLRs and implies that the MADA-type α1 helix has been a feature of the family since the emergence of land plants. However, as sensor NLRs specialized in detecting pathogens, they appear to have frequently lost or degenerated their MADA motif, rendering them

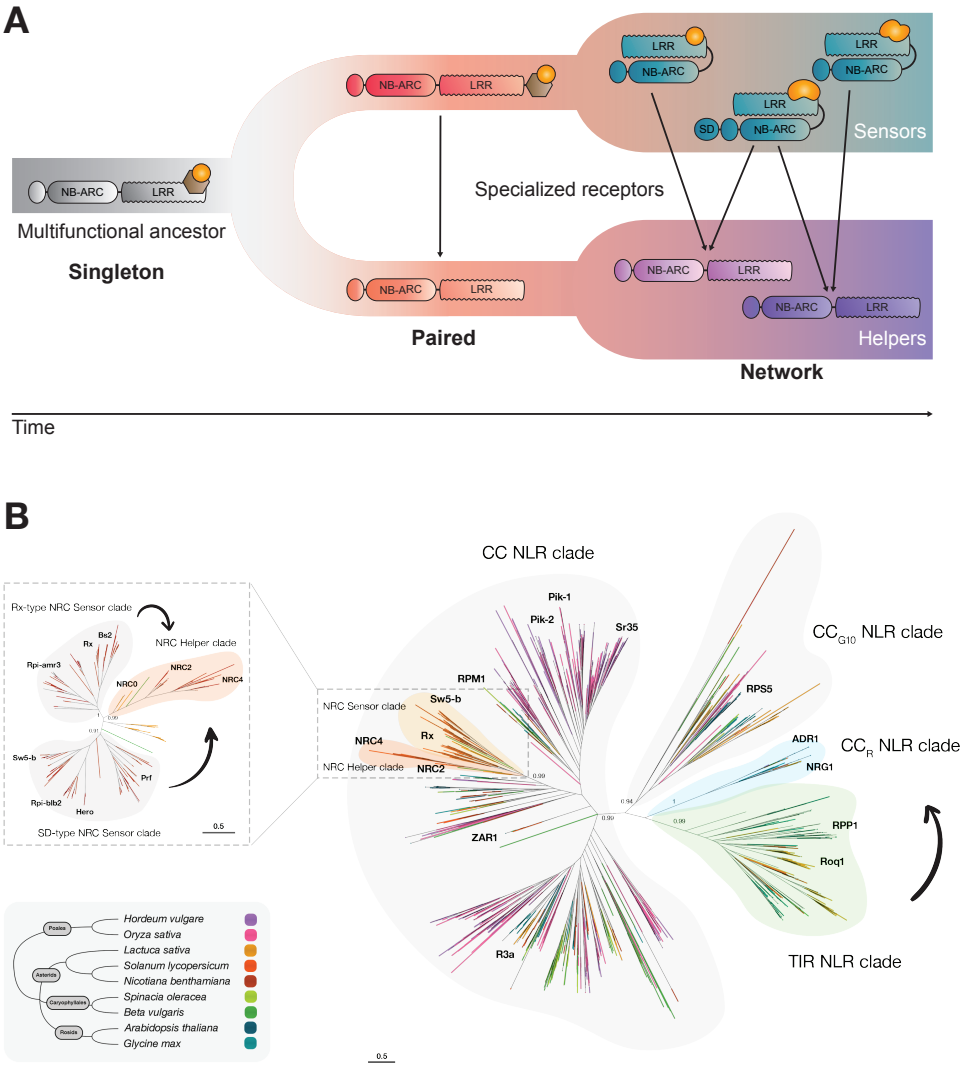


Figure 1.2: Functional configurations of NLR immune receptors and their phylogenetic organization.

(A) Schematic of the proposed transition from a multifunctional singleton ancestor, capable of both pathogen perception and immune signaling, to specialized receptors. Duplication and subfunctionalization gave rise to dedicated sensors, which detect effectors and may acquire integrated domains or a Solanaceous domain (SD), and helpers, which execute immune signaling. These specialized receptors form connections ranging from genetically linked pairs to networks in which several sensors converge on a single helper and individual sensors signal through more than one helper. (B) Maximum likelihood phylogenetic tree of NLRs from nine plant species, with branches colored by species and the CC-NLR, CC$_{G10}$-NLR, CC$_R$-NLR, and TIR-NLR clades indicated. Selected node support values are shown. The inset (left) shows the NRC superclade present in superasterids in detail, comprising the NRC helper clade and the Rx-type and SD-type NRC sensor clades. This figure has been adapted from Contreras et al. (Contreras et al., 2023a) and appears here with permission.

unable to initiate immunity independently (Adachi et al., 2019a).

For instance, while NRC helpers carry a MADA motif, the NRC-S that depend on them lack it. Moreover, the N-termini of NLRs containing MADA or MAEPL motifs can functionally replace the N-terminus of the helper NRC4, whereas the N-termini of sensor CC-NLRs cannot (Chia et al., 2022; Adachi et al., 2019a). These observations are consistent with an evolutionary "use-it-or-lose-it" model. Under this model, as sensor NLRs came to rely on helpers for executing cell death, the conserved motifs in their N-termini degenerated into sequences that no longer function for independently activating immunity or causing cell death.

The functional interchangeability of the motif also makes it a useful diagnostic. Because the MADA motif is largely confined to receptors that execute cell death, its presence has been used as a sequence-based criterion to identify candidate helpers and singletons in genomes where functional data are lacking (Adachi et al., 2019a; Kourelis et al., 2021). Like the integrated domain criterion discussed above. However, this more complicated as: 1/ the motif is detected in only around a fifth of CC-NLRs and many non-MADA are capable to cause cell death, 2/ its degeneration in sensors is gradual rather than absolute, and 3/ receptors carrying an apparently intact motif are not always competent to induce cell death.

1.5 Mechanisms of NLR activation

1.5.1 The ZAR1 resistosome and the pentameric paradigm

Recent studies have provided substantial insight into the activation mechanisms of CC-NLRs and of TIR-NLR/CC_R-NLR pathways (Wang et al., 2019b; Förderer et al., 2022; Wang et al., 2019a; Ma et al., 2020; Martin et al., 2020; Zhao et al., 2022). In functionally singleton CC-NLRs such as ***Arabidopsis thanliana*** ZAR1, pathogen detection triggers a cascade of conformational changes leading to the assembly of a wheel-shaped pentameric complex termed the "resistosome". This activation is driven primarily by the NB-ARC domain: uridylylation of the receptor-like cytoplasmic kinase PBL2 by the bacterial effector AvrAC is sensed through the partner pseudokinase RKS1, which primes the ADP-bound ZAR1 monomer for nucleotide exchange and subsequent oligomerization (Wang et al., 2019b; Wang et al., 2019a).

The resulting pentamer inserts into the plasma membrane through its N-terminal CC domain. In this assembly, the $\alpha 1$ helices of the five protomers, which are buried against the NB-ARC domain in the resting state, swing outward to form a

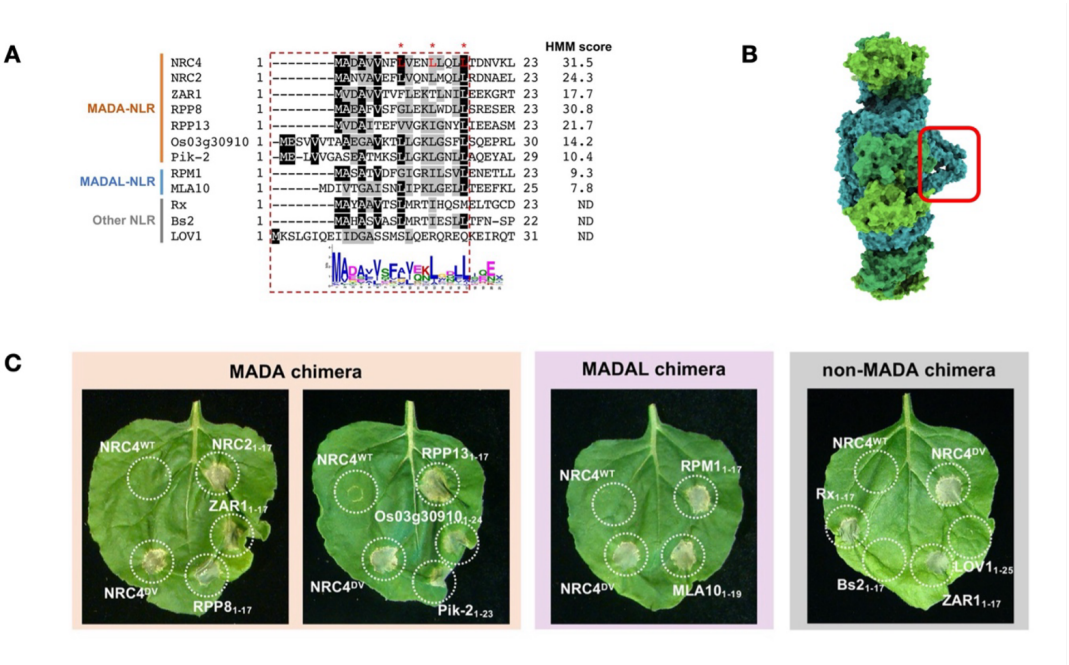


Figure 1.3: The MADA motif is a conserved, functionally replaceable unit at the very N-terminus of various CC-NLRs.

(A) Alignment of the N-terminal region of MADA and MADA-like (MADAL) CC-NLRs, together with other NLRs that lack the motif. Key residues for cell death activity are marked as red characters with asterisks, and the HMM score for the MADA motif is indicated for each sequence. The consensus sequence logo of the motif is shown below the alignment. (B) Position of the MADA motif at the N-terminus of the ZAR1 resistosome structure (Wang et al., 2019a), indicated by the red box. (C) Cell death phenotypes induced by NRC4 chimeras in which the N-terminus of NRC4 was replaced with that of MADA, MADAL, or non-MADA NLRs. $NRC4^{WT}$ and the autoactive $NRC4^{DV}$ were used as negative and positive controls. Photographs were taken 5 days after agroinfiltration. This figure has been adapted from Adachi et al. (Adachi et al., 2019a) and appears here with permission.

funnel-shaped pore (Wang et al., 2019a). This funnel structure is required for cell death, and electrophysiological studies established that the ZAR1 resistosome acts as a calcium-permeable cation channel, placing sustained calcium influx upstream of the cell death response (Bi et al., 2021). The structure provided a mechanistic model for how effector recognition at the C-terminal LRR is converted into a membrane-associated signaling event at the N-terminus, becoming the reference paradigm against which subsequent CC-NLR assemblies were compared.

1.5.2 Sr35: the second pentameric resistosome

The mechanism first observed for ZAR1 has since been validated in other CC-NLRs. The wheat receptor Sr35 differs from ZAR1 in that it recognizes its cognate effector, AvrSr35, directly through its LRR domain rather than by guarding a host component.

Despite this difference, Sr35 also assembles into a pentameric resistosome with the same overall wheel-shaped architecture and comparable calcium channel activity (Förderer et al., 2022; Zhao et al., 2022). Two receptors from distantly related plant lineages, using different recognition strategies, thus converge on the same activated architecture. This was taken as evidence for a shared activation pathway, temporarily establishing the pentamer as the canonical active state for CC-NLRs (Förderer et al., 2022).

1.5.3 Activation in the NRC network

In the NRC network the sensor–helper communication and activation mechanism differs substantially from the singleton model. NRC-S trigger the activation and subsequent oligomerization of their downstream helper NRCs without forming a stable complex with them (Contreras et al., 2022; Ahn et al., 2023). The absence of a detectable sensor–helper complex, combined with the observation that activated helpers oligomerize while their upstream NRC-S remain monomeric, is most consistent with a transient "activation-and-release" interaction rather than a persistent heterooligomeric signaling complex (Contreras et al., 2022; Ahn et al., 2023). Further work mapped this activity to the nucleotide-binding domain (NBD) of the sensor NLR, which is sufficient as a minimal unit to trigger helper activation and oligomerization (Contreras et al., 2023c). More recently, we identified the interfaces mediating this transient interaction (Toghani et al., 2026a). Using AlphaFold 3, we could predict a high-confidence model of the sensor NLR Rx in complex with its helper NRC2. We then validated the predicted contacts through loss- and gain-of-function mutagenesis (Toghani et al., 2026a). These sensor–helper interfaces are conserved across asterid NRC networks despite more than 120 million years of divergence. Notably, structure-guided mutagenesis of a lettuce sensor NLR was sufficient to expand the range of its helper compatibility (Toghani et al., 2026a). Together, these findings support the activation-and-release model and show that sensor–helper compatibility is determined by a small number of key contact interfaces.

The resting and activated states of the helper NLRs themselves also diverge from the known singleton model. NRC helpers such as NRC2 exist natively as homodimers rather than monomers (Selvaraj et al., 2024), meaning that upstream sensors must interact and convert a resting homodimer into a higher-order assembly rather than simply activating an autoinhibited monomer. Furthermore, activated NRC2, NRC3, and NRC4 assemble into hexameric resistosomes, deviating from the pentameric structures observed for singletons (Liu et al., 2023a; Madhuprakash et al., 2024; Seager et al., 2026). Remarkably, recently we have also shown that NRC7 from several

Solanaceae species assembles into an 11-mer resistosome structure upon activation (Toghani et al., 2026b), suggesting structural and oligomeric diversity even within the same closely related protein family. However, so far, no empirical structure of resting-state or activated NRC-S has been reported. Together, these findings highlight fundamental differences in the resting architecture and activation dynamics of the NRC network relative to singleton CC-NLRs, pointing to potential structural diversity within the NLR family.

1.5.4 CC_{G10}-NLRs and the octameric resistosome

The CC_{G10}-NLRs form a phylogenetically distinct clade within the CC-type NLRs, resolved as a separate lineage in NB-ARC phylogenies of angiosperm receptors and classified as one of the four main N-terminal classes in current plant NLR nomenclature (Shao et al., 2016; Kourelis et al., 2021). This class has remained comparatively understudied despite containing several classic resistance genes, including *Arabidopsis* RPS2 and RPS5 (Bent et al., 1994; Mindrinos et al., 1994; Qi et al., 2012; Pottinger and Innes, 2020), pepper Pvr4 (Kim et al., 2015; Lee et al., 2023), and soybean Rsv3 (Tran et al., 2018). Functionally, characterized CC_{G10}-NLRs behave like singletons. A genome-wide screen of pepper NLRs identified a group of ancient and autonomous NLRs (ANLs) carrying autoactive coiled-coil domains whose cell death activity requires neither NRCs nor the NRG1/ADR1 helper families, and which localize directly to the plasma membrane like ZAR1 (Lee et al., 2021). Conserved across seed plants, the CC_{G10} class represents an ancient family that is distinct from both canonical singleton CC-NLRs and the networked helpers discussed above.

This class is further distinguished by sequence divergence in its CC domain. Canonical CC-NLRs carry a conserved EDVID motif in the $\alpha 3$ helix that packs against the NB-ARC domain to support autoinhibition and oligomerization. The sequence variation of this motif tracks structural and functional divergence closely enough that its presence has been proposed as a predictor of canonical resistosome assembly (Sulkowski et al., 2025). CC_{G10}-NLRs lack this EDVID motif, nor do they carry an N-terminal MADA motif. In the case of pepper CC_{G10}-NLRs, cell death activity depends on the structural integrity of the $\alpha 1$ helix itself (Lee et al., 2021; Lee et al., 2023). These features indicated that the N-terminal architecture of CC_{G10}-NLRs diverges significantly from the canonical CC-NLR models, hinting at potential differences in structural configurations in this class.

This hypothesis has recently been confirmed experimentally. Cryo-EM analysis of an activated wheat CC_{G10}-NLR named Wheat AutoImmunity 3 ($WAI3^{GOF}$), phylogenetically close to *Arabidopsis thaliana* RPS2, demonstrated that it assembles into an octameric resistosome; RPS2 was likewise shown to form an octamer, indicating that this stoichiometry

is possibly conserved across this clade (Guo et al., 2026). The WAI3 resistosome triggers sustained cytosolic calcium influx via a distinct channel pore shaped by the divergent configuration of its CC domain, an arrangement that may be shared by other plant NLRs lacking the EDVID motif (Guo et al., 2026; Sulkowski et al., 2025). Within the CC-type NLRs alone, known resistosome architectures therefore span pentamers, hexamers, and octamers, and 11-mers.

1.5.5 TIR-NLR and CC_R-NLR signaling

Plant TIR domains function as NAD^+-cleaving enzymes whose catalytic activity is required for cell death (Wan et al., 2019; Horsefield et al., 2019). This activity is shared with animal SARM1 and a diverse family of prokaryotic TIR proteins involved in antiphage defense, indicating that TIR-based enzymatic signaling is an ancient module recruited independently across the tree of life (Essuman et al., 2022; Shi et al., 2022; Ofir et al., 2021).

Catalysis depends on proximity: cryo-EM structures of the activated ROQ1 and RPP1 resistosomes showed that effector recognition drives assembly into tetramers in which four TIR domains are brought together in an asymmetric arrangement, reconstituting composite NADase active sites so that the activated receptor functions as a holoenzyme (Ma et al., 2020; Martin et al., 2020). Subsequent structural work has extended this model to canonical TIR-NLR activation more broadly (Maruta et al., 2026; Huang et al., 2025). Meanwhile, induced proximity via substrate-driven condensation has emerged as an alternative route to the same activated state for TIR-only proteins, of which ***Arabidopsis thaliana*** and other land plants encode a substantial repertoire alongside the full-length TIR-NLRs (Song et al., 2024; Duxbury et al., 2020; Bayless et al., 2025).

What this holoenzyme produces has been resolved only recently, revealing not a single molecule but a family of nucleotide-derived second messengers. Beyond the variant cyclic ADP-riboses first associated with plant TIR activity, two classes of products have been identified as the biologically active signals: the phosphoribosyl adenosines pRib-AMP and pRib-ADP (Huang et al., 2022), and the ADP-ribosylated nucleotides ADPr-ATP and di-ADPR (Jia et al., 2022). Plant TIR domains have also been reported to act as 2′,3′-cAMP/cGMP synthetases (Yu et al., 2022). The product spectrum differs systematically between plant and prokaryotic TIR domains, with prokaryotes generating distinct cyclic ADPR isomers and, in some bacteria, histidine-ADPR (Bayless et al., 2023; Sabonis et al., 2025; Jia et al., 2023). The signal generated by a TIR-NLR is therefore diffusible and chemically specific, rather than being tied to the membrane insertion of the receptor itself.

These molecules are perceived by the EDS1 signaling hub, which operates as two heterodimeric modules, EDS1–PAD4 and EDS1–SAG101, that act downstream with the helper CC_R-NLRs ADR1 and NRG1, respectively (Pruitt et al., 2021; Lapin et al., 2019; Gantner et al., 2019). The second messengers bind an allosteric pocket at the heterodimer interface to promote association with the cognate helper: pRib-AMP/ADP acts on EDS1–PAD4–ADR1, while ADPr-ATP/di-ADPR acts on EDS1–SAG101–NRG1 (Huang et al., 2022; Jia et al., 2022). This mechanism provides a biochemical basis for the effector-dependent recruitment of NRG1 to EDS1–SAG101 observed *in planta* (Sun et al., 2021).

Specificity between the two branches is determined in part by cavity-lining residues in PAD4 and SAG101 (Dongus et al., 2022). The two branches are also functionally specialized: NRG1 is required primarily for the execution of cell death (Castel et al., 2019; Wu et al., 2019), whereas ADR1 contributes mainly to transcriptional defense reprogramming and the potentiation of pattern-triggered immunity (Saile et al., 2020; Pruitt et al., 2021). The EDS1 family and RNL helpers have coevolved as a functional module and are frequently lost together in lineages lacking TIR-NLRs, notably in monocots, pointing to a tightly constrained signaling unit (Lapin et al., 2020; Baggs et al., 2020; Tarr and Alexander, 2009). Recent work nonetheless indicates that the EDS1–PAD4–ADR1 node itself is conserved and functional in monocots as well as eudicots (Maruta et al., 2025). As in the NRC network, this pathway carries its own endogenous modulators, including the N-terminally truncated variant NRG1C, which antagonizes immunity mediated by full-length NRG1 helpers in *Arabidopsis thaliana* (Wu et al., 2022).

Despite the divergence in upstream signaling chemistry, the endpoint of both CC-NLR and TIR-NLR pathways converges. Helper CC_R-NLRs oligomerize upon activation and associate with the plasma membrane in a manner reminiscent of CC-NLR resistosomes (Feehan et al., 2023; Jacob et al., 2021; Wang et al., 2023). Helper RNLs are activated directly by both plant and bacterial TIR-derived small molecules to assemble into resistosome-like complexes (Yu et al., 2024; Ge et al., 2026). Calcium influx appears to be the shared output: RNL and CC-NLR resistosomes alike act as calcium-permeable channels, and calcium mobilized from both organellar and extracellular pools converges on the execution of cell death (Bi et al., 2021; Bi and Zhou, 2021; Ibrahim et al., 2026; Shen et al., 2024). However, it was recently shown NRG1-type helper NLRs localize to organellar membrane and channel internal calcium deposits from such as chloroplast to induce the calcium influx (Ibrahim et al., 2026), contrasting the known plasma membrane localization of CC-NLR resistosomes. Nevertheless, two receptor classes with entirely different recognition strategies and signaling mechanisms therefore converge on the same execution step, reinforcing the principle that oligomerization

into a defined higher-order assembly is the universal mechanism of NLR activation (Hu and Chai, 2023; Wu, 2025).

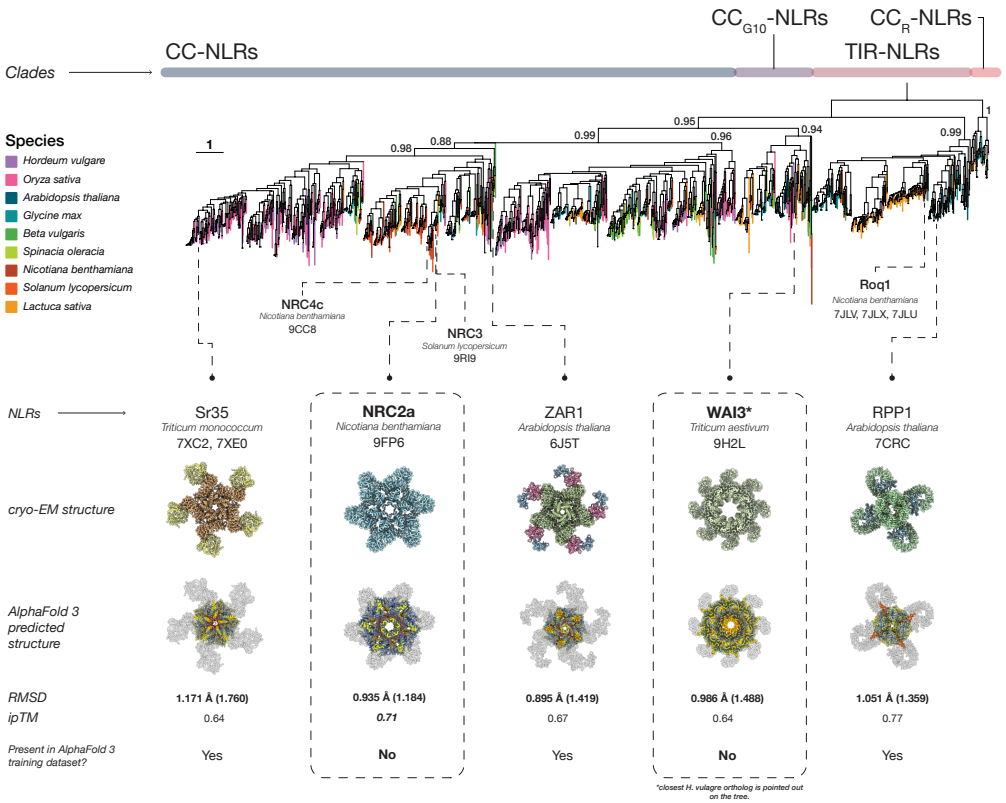


Figure 1.4: Structural diversity of plant NLR resistosomes across the four N-terminal classes. Maximum likelihood phylogenetic tree of NLRs compiled from nine plant species, with branches colored by species and the CC-NLR, CC$_{G10}$-NLR, CC$_R$-NLR, and TIR-NLR clades indicated above. Receptors with experimental structure of the activated resistosome complex are marked on the tree together with their Protein Data Bank accessions. Five of these are shown below in top view: the pentameric TmSr35 (Förderer et al., 2022; Zhao et al., 2022) and AtZAR1 (Wang et al., 2019a) resistosomes, the hexameric NbNRC2a resistosome (Madhuprakash et al., 2024), the octameric wheat WAI3 resistosome (Guo et al., 2026), and the tetrameric AtRPP1 resistosome (Ma et al., 2020). For each, the corresponding AlphaFold 3 prediction is shown beneath the experimental structure, with the root mean square deviation between the two given for the full complex and, in parentheses, for all atoms, together with the interface predicted template modeling (ipTM) score of the prediction. The final row indicates whether the experimental structure was available at the time the AlphaFold 3 training set was assembled. The phylogenetic tree is adapted from one published in Contreras et al. (Contreras et al., 2023a) and appears here with permission; the remaining panels were generated for this thesis.

1.6 Oligomerization and cell death across the domains of life

1.6.1 Nucleotide-dependent oligomerization as a shared signaling principle

Plant NLRs represent one branch of the STAND clade of P-loop NTPases, a group defined by shared domain architecture and nucleotide-binding chemistry that also encompasses animal NOD-like receptors, the apoptotic regulators APAF-1 and CED-4, and, more distantly, the AAA+ ATPases (Leipe et al., 2004; Duxbury et al., 2021; Seraphim and Houry, 2020; Khan et al., 2022). What these proteins share is not sequence similarity across their peripheral domains, but a common signaling logic: a central nucleotide-binding module acts as a conformational switch, with activation transduced by converting an autoinhibited monomer, or an autoinhibited higher-order assembly, into a defined oligomer that serves as a platform for downstream effector recruitment or activation (Hu and Chai, 2023; Saur et al., 2021).

The apoptosome illustrates this principle in its simplest form, where APAF-1 assembles upon cytochrome *c* binding into a wheel-shaped heptamer that recruits and activates initiator caspases (Yuan et al., 2013; Pang et al., 2015). The inflammasomes of animal innate immunity, originally described as molecular platforms for activating inflammatory caspases (Martinon et al., 2002), follow an analogous scheme toward different outputs, whether through templated self-propagation of NLRC4 into a disc (Hu et al., 2015), the cage-to-disc transition of NLRP3 (Andreeva et al., 2021; Xiao et al., 2023), or the proteasome-dependent release of an active NLRP1 fragment (Barry et al., 2023; Sundaram et al., 2024; Fu et al., 2024).

This principle extends well beyond animals and plants. Fungal NLR proteins have nucleotide-binding scaffolds fused to amyloid signaling motifs, that act as oligomerization templates leading to the conversion of a downstream effector domain into a self-propagating amyloid state, as seen in the HET-s/HET-S incompatibility system and the NWD2 receptor (Uehling et al., 2017; Daskalov et al., 2016; Seuring et al., 2012; Greenwald et al., 2010; Gaspar and Pawlowska, 2022). Similar motifs occur across bacterial genomes (Dyrka et al., 2020). Bacterial NLR proteins have since been shown to function in antiphage defense, representing part of a broader paradigm in which components of eukaryotic cell-autonomous immunity have prokaryotic or archaeal antecedents acquired through horizontal gene transfer and domain shuffling (Kibby et al., 2023; Gao et al., 2022; Aravind et al., 2024;

Wein and Sorek, 2022; Bernheim et al., 2024; Culbertson and Levin, 2023; Leão et al., 2024; Johnson and Kranzusch, 2022).

The NLR family itself is correspondingly ancient within eukaryotes: plant NLRs have been traced to chlorophyte green algae and retain recognizable structural and functional hallmarks across that evolutionary distance (Ortiz and Dodds, 2018; Busche and Laflamme, 2024; Feng et al., 2024), while animal NLRs are already present and diversified in sponges (Yuen et al., 2014). What is conserved across these systems is proximity-driven activation mediated by an oligomerizing scaffold rather than a fixed oligomeric geometry. NLR assemblies range from the heptameric APAF-1 apoptosome to the octameric CED-4 apoptosome (Qi et al., 2010), and from discs and filaments in animal inflammasomes to the tetramers, pentamers, hexamers, octamers, and 11-mers identified in plants. That stoichiometry remains variable even where the overall activation mechanism is conserved is a central theme addressed directly in Chapter 5.

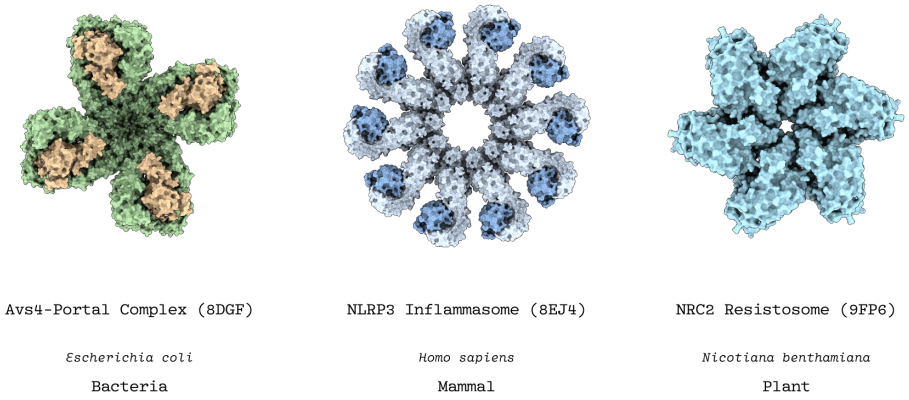


Figure 1.5: Nucleotide-binding immune receptors assemble into higher-order complexes across the domains of life.
Experimentally determined structures of three activated receptor complexes, shown in top view. Left, the bacterial antiviral STAND protein Avs4 of *Escherichia coli* (green) bound to the portal protein of phage PhiV-1 (orange), which it recognizes as a conserved viral signature (PDB 8DGF) (Gao et al., 2022). Center, the disc-shaped active human NLRP3 inflammasome in complex with NEK7 and the adaptor ASC, which recruits caspase-1 (PDB 8EJ4) (Xiao et al., 2023). Right, the hexameric resistosome formed by the helper NLR NRC2 of *Nicotiana benthamiana* (PDB 9FP6) (Madhuprakash et al., 2024). Although the three receptors differ in their N-terminal domains, recognition specificities, and downstream outputs, each converts pathogen or phage detection into a defined ring-shaped oligomer. Structures are not shown to a common scale.

1.6.2 Executing cell death: convergent routes to membrane permeabilization

This evolutionary comparison is informative about how these pathways end as well as how they begin. In animals, both pyroptosis and necroptosis terminate in the permeabilization of the plasma membrane by a dedicated executor protein distinct from the upstream receptor. Inflammasome-activated caspases cleave gasdermins, releasing an N-terminal fragment that oligomerizes within the membrane to form a large β-barrel pore (Ding et al., 2016; Ruan et al., 2018; Broz et al., 2020). Similarly, during necroptosis, RIPK3 phosphorylates the pseudokinase MLKL, which then translocates to the plasma membrane to permeabilize it (Murphy et al., 2013; Petrie et al., 2019; Samson et al., 2020). Both executor classes are ancient: gasdermins occur in fungi and bacteria, where bacterial gasdermin pores function in antiphage defense (Chen and Broz, 2024; Johnson et al., 2024; De Schutter et al., 2021). In animals, sensing and execution are therefore separated across multiple regulatory steps, and the terminal rupture of the membrane is itself a regulated event mediated by NINJ1 (Degen et al., 2023; Newton et al., 2024). These pathways exhibit significant crosstalk, with pore-forming executors recurring widely as the terminal step in regulated cell death (Ai et al., 2024; Snyder and Oberst, 2021; Mesa-Galloso et al., 2021).

In the canonical NLR pathways, singleton and helper NLR resistosome funnel formed by the N-terminal α1 helices inserts directly into the membrane and acts as a calcium-permeable channel. Consequently oligomerization and membrane permeabilization are mediated by a single NLR receptor rather than distributed across an extended protease cascade (Bi et al., 2021; Förderer et al., 2022; Bi and Zhou, 2021).

Even so, these plant and animal systems converge on a shared proximal trigger, as sustained calcium influx precedes cell death in both kingdoms (Bi et al., 2021; Chen et al., 2026). However, more recently some NLRs have been reported to require non-NLR components for immune signaling. Plants encode a family of MLKL-like proteins that function in TIR-NLR immunity and mediate cytoplasmic calcium influx (Mahdi et al., 2020; Shen et al., 2024), and a wheat receptor pair was recently shown to execute cell death through an auxiliary MLKL helper (Bennett et al., 2026). Dedicated executors and self-executing receptors can therefore both operate in plants across distinct signaling branches (Maekawa et al., 2023; Pitsili et al., 2019).

1.7 The evolution of NLR immune receptors

1.7.1 The classical view: co-evolution with pathogen effectors

NLRs are among the most evolutionarily dynamic families of plant proteins and are under constant pressure to evolve (Adachi et al., 2023b; Prigozhin and Krasileva, 2021; Tamborski and Krasileva, 2020). Selection can drive the recognition of new pathogen effectors or host targets, or promote evasion from effector-mediated suppression. The signature of this process is most visible in the LRR, where exposed residues on the concave side show evidence of diversifying selection, both between species and within populations (Prigozhin and Krasileva, 2021; Prigozhin et al., 2025; Liu et al., 2005).

At the repertoire level, NLR gene numbers range from a handful to well over a thousand per genome. They are generally understood to follow a birth-and-death model, in which gene duplication continually generates new receptors, some of which are retained for their recognition capability while others are lost or pseudogenized (Michelmore and Meyers, 1998; Seong et al., 2020; Wingen et al., 2025). As a result, presence/absence variation is common, in a way that a receptor present in one accession of a species may be entirely absent from another (Van De Weyer et al., 2019; Barragan and Weigel, 2021). The forces behind this turnover are not limited to selection for new recognition specificities. Mutation, recombination, and the rates of duplication and deletion all contribute, and population genomics offers a way to weigh their relative contributions across a repertoire rather than one receptor at a time (Wingen et al., 2025). Pathogens add a further layer of pressure by directly targeting immune receptors: the Solanaceae helper NLRs NRC2 and NRC3 are suppressed by the cyst nematode effector SS15 and the oomycete effector AVRcap1b (Derevnina et al., 2021). In some species NRC3 has accumulated substitutions in the effector-binding loop that allow it to evade suppression (Contreras et al., 2023b).

1.7.2 Emergence of networks from singleton ancestors

The evolution of CC-NLRs in plants reflects a transition from an ancestral singleton to complex networks with specialized partners. This shift, driven by the need for rapid adaptation to evolving pathogen effectors, resulted in the formation of dedicated sensor and helper NLRs, each fulfilling distinct roles during the immune response (Wu et al., 2018; Wu et al., 2017). This division of labor is thought to offer several advantages. It alleviates evolutionary constraints on individual NLRs, enabling faster diversification and functional specialization. Sensor NLRs can also more readily incorporate additional domains,

such as IDs or the SD, to expand their detection capabilities without being constrained by execution requirements (Cesari et al., 2014; Seong et al., 2020; Białas et al., 2021). This modularity, seen in both paired and networked configurations, is likely a key driver of the adaptability of the plant NLR family.

1.7.3 Genomic organization and the mechanics of NLR diversification

The distribution of NLR genes in plant genomes provides further insight into their evolutionary history. In plants such as *Arabidopsis*, about half of the NLRs are found in clusters formed through tandem gene duplication or unequal crossing-over events (Lee and Chae, 2020; Van De Weyer et al., 2019). Others are scattered throughout the genome and often lack introns, which suggests expansion via retroduplication (Kim et al., 2017). These duplication processes allow NLRs to develop new functions and domain architectures, leading to diverse gene clusters containing either phylogenetically related genes or a mix of different NLR types (Adachi et al., 2019b).

Sensor–helper NLR pairs, such as the rice pairs Pik-1/Pik-2 and Pia (RGA5/RGA4), are typically found in close genetic proximity, often in head-to-head arrangements (Cesari et al., 2014; Białas et al., 2021; Stein et al., 2018). This close linkage likely facilitates transcriptional co-regulation and ensures that functionally dependent sensor and helper pairs are co-inherited. This is especially important given that some helpers are intrinsically active and could trigger autoimmunity in the absence of their corresponding sensor (De La Concepcion et al., 2018; Ashikawa et al., 2008).

In contrast, the NRC network in asterids shows a different pattern, where NRC-S and NRC-H are mostly genetically dispersed despite their common evolutionary origin (Wu et al., 2017). Within this network, some helper clades can nonetheless remain linked with their cognate NRC-S, as seen for the NRC6 and Hero clades (Lüdke et al., 2023). The ancestral helper NRC0 is likewise found in a conserved cluster with the NRC-S that depend on it (Sakai et al., 2024). Overall, NRC-S and NRC-H belong to distinct but related phylogenetic clades. The network dates back roughly 120 million years and probably evolved from an initial sensor/helper gene pair derived from an ancestral singleton NLR (Goh et al., 2024; Sakai et al., 2024). In the TIR-NLR/CC_R-NLR network, helper NLRs like NRG1 and ADR1 and their upstream TIR-NLRs likewise form distinct phylogenetic groups and may have evolved from similar functional pairings (Liu et al., 2023b).

1.7.4 Asymmetric evolution of sensors and helpers

A prediction that follows from this division of labor is that sensors and helpers should experience different selective regimes. Sensors track rapidly evolving effectors and are expected to diversify, whereas helpers are constrained by the need to remain compatible with many upstream partners and are expected to be more conserved. The broad generalization in the literature is indeed that sensors evolve rapidly while helpers remain conserved (Goh et al., 2024). Within the NRC network, NRC0 stands out as the most ancient and conserved helper and is often genetically linked to its dependent sensors, whereas more derived helpers have undergone lineage-specific expansions and occupy varied genomic positions (Goh et al., 2024; Sakai et al., 2024; Lee et al., 2021). A comparable pattern is seen for ZAR1, which has remained atypically conserved across angiosperm evolution while the ZRK guardees on which it depends for recognition have greatly diversified (Adachi et al., 2023b). This generalization, however, is unlikely to be the whole story. Helpers face competing selective demands: they must remain compatible with multiple upstream sensors while also evading effectors that target them directly, two pressures that pull in opposite directions (Derevnina et al., 2021; Contreras et al., 2023b).

1.7.5 Evolutionary pressures beyond R–AVR co-evolution

The selective pressures described so far are most often framed in terms of classic R–AVR co-evolution, and the recognition surfaces of sensor NLRs clearly bear this signature. A substantial part of NLR sequence evolution, however, appears to concern the relationships among NLRs themselves rather than their interactions with pathogens. This includes the requirement that a receptor signal correctly and exclusively to its intended partners, that paralogous nodes remain insulated from one another. Insulation of this kind would preserve functional redundancy, minimize undesired cross-activation between pathways, and provide a route to evade pathogen effectors that suppress individual helpers (Selvaraj et al., 2024; Derevnina et al., 2021; Contreras et al., 2023b). Moreover, the network activity should be kept in check by modulators such as NRCX (Adachi et al., 2023a).

1.8 Genome-scale resources for studying NLR evolution

1.8.1 The expansion of sequenced plant genomes

Over the past two decades, the number of publicly available plant genome assemblies has grown from a handful of model species to many hundreds spanning the land plant phylogeny, driven largely by falling sequencing costs and the maturation of long-read technologies (Marks et al., 2021). This rapid accumulation of genomes has opened the door to large-scale comparative analyses of plant gene families, allowing tracing gene gain and loss, copy-number variation, and lineage-specific expansions across broad evolutionary distances. Sampling has also deepened within species, so that many crops are now represented by pangenomes assembled from dozens of accessions rather than by a single reference genome (Zhou et al., 2022; Li et al., 2023; Van Workum et al., 2024). For a gene family as variable as the NLRs, this shift matters, since a single reference genome captures only a fraction of the receptors segregating in a species.

However, the value of such comparisons depends critically on the consistency of the underlying gene annotations. Unlike genome assembly, gene annotation has not been standardized: genomes are annotated with different gene predictors, different sources and depths of transcript or protein evidence, and different parameter choices. This can lead to orthologous sequences to be modeled as a gene in one species but missed entirely in another (Salzberg, 2019). This annotation heterogeneity can systematically bias comparative results. Mixing annotation methods across a clade has been reported to inflate the apparent number of lineage-specific genes by up to 15-fold (Weisman et al., 2022; Van Oss and Carvunis, 2019), whereas errors in gene number inferred from draft assemblies are pervasive (Denton et al., 2014). The same problem confounds presence/absence analyses, where the apparent absence of an ortholog in a heterogeneous dataset may reflect poor or inconsistent annotation rather than true gene loss.

1.8.2 Why NLRs are particularly affected

NLRs are highly diverse and rapidly evolving, which makes them a compelling subject for evolutionary study but also a particularly challenging gene family to annotate consistently (Barragan and Weigel, 2021; Michelmore and Meyers, 1998; Shao et al., 2016; Prigozhin et al., 2025). Their modular, repetitive domain structure, their tendency to occur in dense tandem clusters, and their frequent location in repeat-rich or transposon-rich regions all frustrate standard gene-prediction pipelines. Because of this, often specialized motif-aware tools are required to recover

the full repertoire (Steuernagel et al., 2015; Steuernagel et al., 2020; Li et al., 2016; Kourelis et al., 2021; Jupe et al., 2012; Kim et al., 2017).

The consequences are not evenly distributed across the family. Clustered NLRs are prone to being merged into chimeric gene models or split across several partial ones; pseudogenized copies may be annotated as functional or omitted altogether, and receptors carrying integrated domains are especially liable to be truncated at the point of fusion. Because these failure modes is largely dependent on the pipeline, the same locus can be represented quite differently in two genomes annotated by different groups, and comparative counts of NLR repertoire size can reflect annotation practice as much as biology. Curated resources such as NCBI RefSeq provide high-quality, internally consistent annotations, but only for a limited number of reference genomes (Pruitt et al., 2007), which leaves the great majority of published assemblies outside any harmonized standard.

1.8.3 Deep-learning gene annotation

Gene annotation has long relied on *ab initio* gene finders built on hidden Markov models, an approach pioneered in eukaryotes in part on compact fungal genomes (Ter-Hovhannisyan et al., 2008; Min et al., 2017). These tools remain widely used, but they depend on genome-specific training or on transcript and protein evidence, and their parameters and outputs still vary from assembly to assembly. A more recent generation of deep-learning annotators aims to close this gap by predicting gene models directly from genomic DNA sequence. Helixer, for example, uses a deep neural network to assign each base to a structural category and then resolves these base-wise predictions into gene models with a hidden Markov model (Stiehler et al., 2021). Because a single model is trained once per lineage on curated NCBI RefSeq annotations and then applied uniformly across genomes, such tools can produce internally consistent, harmonized annotations at scale (Stiehler et al., 2021; Holst et al., 2023). The paradigm is developing quickly, with recent tools coupling deep learning to differentiable HMMs and to large DNA foundation models (Gabriel et al., 2024; Liu et al., 2025; Vader et al., 2026).

However, no pipeline is perfect. While tools like Helixer are trained on high-quality datasets such as NCBI RefSeq, applying a single model uniformly across genomes harmonizes gene models rather than necessarily improving them. Deep-learning predictors are themselves imperfect and carry their own systematic biases, so a harmonized annotation is best understood as internally consistent rather than error-free (Sugihara et al., 2026). With that qualification, harmonized annotation removes one of the main confounders in comparative analysis of a family as annotation-sensitive as the NLRs.

1.9 Structural biology and computational modeling of NLRs

1.9.1 What experimental structures have revealed

Over the past few years, a succession of cryo-EM structures has transformed our understanding of NLR-mediated immunity. These resolve several activated receptor complexes, including the ZAR1 and Sr35 pentamers, the tetrameric TIR-NLR resistosomes ROQ1 and RPP1, and most recently, the hexameric NRC2 and NRC4 assemblies alongside the resting-state NRC2 homodimer (Wang et al., 2019a; Förderer et al., 2022; Martin et al., 2020; Ma et al., 2020; Liu et al., 2023a; Madhuprakash et al., 2024; Selvaraj et al., 2024). Collectively, these structures established that NLR activation proceeds through oligomerization, while demonstrating that oligomeric state and signaling output differ systematically across classes.

Despite these advances, each structure required substantial effort in protein purification, sample preparation, and data collection, and as a result coverage remains restricted to a small number of model receptors relative to the size of the family. Moreover, within these structures, the N-terminal α1 helix that forms the CC-type resistosome funnel is structurally dynamic and frequently unresolved or only partially resolved in the resulting maps (Förderer et al., 2022; Zhao et al., 2022; Liu et al., 2023a; Madhuprakash et al., 2024). Consequently, the region responsible for membrane insertion and cell death execution is missing from experimental data and has remained elusive.

1.9.2 Protein structure prediction and its extension to complexes

Computational structure prediction offers a complementary route. AlphaFold 2 made accurate monomer prediction available (Jumper et al., 2021; Senior et al., 2020), and its multimer extension allowed researchers to model protein complexes (Evans et al., 2021). AlphaFold 3, in addition to faster prediction speed and overall improvements, added the capacity to model proteins together with ligands, ions, and small molecules (Abramson et al., 2024). This feature was proven to be highly relevant for NLR biology, as CC-type NLR resistosomes could be modeled with lipids binding to the N-terminal α1 helix, simulating the plasma membrane.

These tools also report a set of internal confidence metrics that indicate how far a given model should be trusted. The most widely used are the per-residue predicted Local

Distance Difference Test (pLDDT), the pairwise Predicted Aligned Error (PAE), and the model-wide predicted template modeling (pTM) and ipTM scores (Jumper et al., 2021; Evans et al., 2021; Abramson et al., 2024). Each captures a different aspect of a prediction and they are most informative when interpreted together, since a complex can score modestly overall while still containing local regions that are modeled with high confidence and are biologically meaningful (Posbeyikian et al., 2026). These metrics also carry known biases. The ipTM score, for instance, is sensitive not only to the size of the interface but also to the total length of the proteins, and it decreases when one partner is disordered even if the interface itself is modeled correctly. Several complementary interface scores have been developed to address this, including ipSAE, pDockQ, and the Local Interaction Score (LIS), which allow predicted complexes to be assessed at scale without manual inspection of PAE plots (Posbeyikian et al., 2026).

Prediction is also a stochastic process. These tools use random seeds to initialize the masking or subsampling of the multiple sequence alignment (MSA), and in the diffusion-based models such as AlphaFold 3 the seed additionally sets the initial noise pattern from which the structure is generated. Because proteins exist as ensembles of related conformations rather than as single static structures, a given run may converge on a local solution. Sampling several seeds and asking whether the same model is recovered each time is therefore a practical way to distinguish robust structural features from stochastic variation (Posbeyikian et al., 2026). The depth of the MSA matters for the same reason: AlphaFold 2 depends heavily on deep evolutionary profiles and tends to struggle with fast-evolving or lineage-specific proteins, whereas AlphaFold 3 relies more on learned structural priors and is comparatively less sensitive to shallow alignments (Posbeyikian et al., 2026). This is directly relevant for NLRs, which are among the most rapidly evolving gene families in plants and frequently return shallow alignments.

1.9.3 Applications to plant immunity and their limits

Structure prediction has already been applied extensively to NLR biology, including structure-informed comparative analyses of receptors and effectors, classification of proteins whose relationships are not resolvable from sequence alone, and modeling-guided engineering of recognition specificity (Ibrahim et al., 2023; Seong and Krasileva, 2021; Barrio-Hernandez et al., 2023; Contreras et al., 2023b).

Because every predicted structure is effectively a hypothesis, benchmarking is the most direct way to decide how far to trust one. Where an experimental structure exists, it should be modeled alongside the targets of interest and used as a reference. Negative controls matter just as much, since confidence scores on their own often cannot separate a genuine complex from a plausible-looking artifact (Posbeyikian et al., 2026).

Low-confidence or unexpected predictions can be informative rather than simply being failures. Interpreted against what is already known about a protein family, they can point to architectures or functions that depart from the canonical model (Posbeyikian et al., 2026). NRC helpers, for example, generally model with high confidence as hexamers. Through a systematic inspection of outliers in a large-scale prediction we found that the NRC7 clade deviates from this arrangement, which was subsequently confirmed as an 11-mer by electron microscopy (Toghani et al., 2026b; Posbeyikian et al., 2026). The same logic applies to the sensor–helper distinction, since sensors typically fail to return confident oligomeric predictions while helpers generally model with higher confidence. This prediction confidence can itself carry functional information to classify NLRs (Toghani et al., 2025a; Pai et al., 2026; Posbeyikian et al., 2026).

However, predictions are therefore best treated as hypotheses to be tested rather than as direct substitutes for experimental structures. Predictions are made in the absence of the cellular context in which these receptors operate, lacking membranes, partners, or post-translational modifications unless these are supplied explicitly. They also depend on the depth and quality of the multiple sequence alignment available for a given receptor, which may vary across the NLR family.

1.10 Aims and thesis outline

The evolution of plant NLRs is most often interpreted through co-evolution with pathogen effectors. This framework accounts well for how recognition specificities are gained and lost, but it explains considerably less about the remainder of the receptor. The central question of this thesis is therefore what other forces shape NLR sequence and structure. I focus on the following topics: 1. how the evolutionary patterns determine or affect sensor-helper relationships?, 2. how paralogous receptors remain functionally insulated from one another?, and 3. to explore the structural and functional diversity of NLRs using structural modeling. To address these questions, I combined comparative genomics, phylogenomics, and computational structural modeling. I used the superasterids NRC network as the principal system, given its massive but lineage-specific expansion, containing both sensor and helper NLRs, and being well characterized experimentally.

In Chapter 2 I establish the resources on which the remainder of the thesis depends on. Because annotation heterogeneity can be mistaken for biological signal in a gene family as variable as the NLRs, I compiled three NLRome datasets at nested taxonomic resolutions and re-annotated the underlying genomes with a single deep-learning pipeline. This

allows repertoire size, class composition, and presence or absence of individual clades to be compared across genomes in a harmonized manner.

In Chapter 3 I examine two features of the NRC helper clade that does not directly relate to effector recognition. I first analyze the diversity of the N-terminal MADA motif across the clade, identify a sequence signature shared between the modulator NRCX and a previously undescribed group of NRC4 sequences, and test the functional consequences of that signature in cell death assays. I then turn to the interfaces which NRC2 self-associates in its resting state, and determine whether these have diverged between paralogous helpers in the manner expected if selection acts to keep their signaling pathways insulated from one another.

In Chapter 4 I extend my analysis from individual helpers to entire networks. I compare the architecture and genomic organization of the NRC network between the Asterales and the Solanales, establish experimentally sensor-helper relationships in lettuce, and test whether these relationships shaped the evolution of the network.

In Chapter 5 I address the same problem from a structural perspective. I benchmark AlphaFold 3 against the CC-NLR resistosomes that have been determined experimentally, and then apply the same approach to receptors whose structures remain unsolved, in order to establish how far oligomeric state varies across the CC-type NLRs and whether it can be inferred computationally.

In Chapter 6 I test the extent to which sensors and helpers can be distinguished by their predicted oligomeric structures, evaluate this criterion against the sequence-based ones currently used for the same purpose, and apply it to NLR pairs and networks for which functional data are limited.

Finally, in Chapter 7 I discuss what I learned about the forces acting on immune receptors within a network, and about the extent to which structural prediction can assisst experimental characterization and classification of NLR proteins.

Chapter 2: Curation of harmonized plant NLRome datasets on multiple taxonomic resolutions

The work described in this chapter was carried out in collaboration with Yu Sugihara (The Sainsbury Laboratory). For the superasterid dataset, I collected and compiled the genome assemblies from NCBI Datasets (`https://www.ncbi.nlm.nih.gov/datasets/`), and ran the Helixer and NLRtracker pipelines together with Yu Sugihara. For the Solanaceae dataset, Yu Sugihara and I collected the genomes together, and ran the Helixer, NLRtracker, and BUSCO pipelines. All downstream analyses, data curation, and interpretations presented in this thesis were performed by me.

2.1 Introduction

Over the past two decades, the number of publicly available plant genome assemblies has grown from a handful of model species to many hundreds spanning the land plant phylogeny, driven largely by falling sequencing costs and the maturation of long-read technologies (Marks et al., 2021). This rapid accumulation of genomes has opened the door to large-scale comparative analyses of plant gene families, allowing researchers to trace gene gain and loss, copy-number variation, and lineage-specific expansions across broad evolutionary distances. However, the value of such comparisons depends critically on the consistency of the underlying gene annotations. Unlike genome assembly, gene annotation has not been standardized: genomes are annotated with different gene predictors, different sources and depths of transcript or protein evidence, and different parameter choices, so that orthologous sequences may be modeled as a gene in one species but missed entirely in another (Salzberg, 2019). This annotation heterogeneity can systematically bias comparative results. It has been reported that mixing annotation methods across a clade can inflate the apparent number of lineage-specific genes by up to 15-fold (Weisman et al., 2022; Van Oss and Carvunis, 2019), and errors in gene number inferred from draft assemblies are pervasive, with a majority of genes assigned the wrong copy number in some datasets (Denton et al., 2014). The same problem confounds presence/absence analyses, where the apparent absence of an ortholog or paralog in a heterogeneous dataset may reflect poor or inconsistent annotation rather than true gene loss. Distinguishing biological signal from annotation artifact therefore requires harmonized, high-quality datasets, making the production of consistently annotated genome collections a prerequisite for robust comparative genomics.

In plant immunity, the nucleotide-binding leucine-rich repeat receptor (NLR) family of proteins has been extensively studied due to its crucial role in pathogen recognition and the activation of immune responses (Jones and Dangl, 2006; Jones et al., 2016; Dodds et al., 2024). NLRs are highly diverse in plants and rapidly evolving, which makes them a compelling subject for evolutionary study but also a particularly challenging gene family to annotate consistently (Barragan and Weigel, 2021; Michelmore and Meyers, 1998; Shao et al., 2016; Prigozhin et al., 2025). NLR annotation has historically been difficult: their modular, repetitive domain structure, their tendency to occur in dense tandem clusters, and their frequent location in repeat-rich or transposon-rich regions all frustrate standard gene-prediction pipelines, so that specialized motif-aware tools are typically required to recover the full repertoire (Steuernagel et al., 2015; Steuernagel et al., 2020; Li et al., 2016; Kourelis et al., 2021; Jupe et al., 2012; Kim et al., 2017). As a result, NLR gene models are especially sensitive to how a genome was annotated, and comparative studies of this family depend on high-quality and comparable annotations. Yet among publicly available genomes, many lack author-provided annotations altogether, and where annotations do exist the pipelines and parameter choices used to generate them differ from study to study, which can markedly affect how gene families such as NLRs are resolved (Toghani and Kamoun, 2024; Toghani et al., 2026c; Weisman et al., 2022). Curated resources such as NCBI RefSeq provide high-quality, internally consistent annotations, but only for a limited number of reference genomes (Pruitt et al., 2007).

Gene annotation has long relied on *ab initio* gene finders built on hidden Markov models. This approach was pioneered in eukaryotes on compact fungal genomes: GeneMark-ES introduced unsupervised self-training of a hidden Markov model (HMM) directly on genomic sequence (Ter-Hovhannisyan et al., 2008), and pipelines such as FunGAP later combined several such predictors with evidence-based model selection to annotate fungal assemblies (Min et al., 2017). These tools remain widely used, but they depend on genome-specific training or on transcript and protein evidence, and their parameters and outputs still vary from assembly to assembly. A more recent generation of deep-learning annotators aims to close this gap by predicting gene models directly from genomic DNA sequence. Helixer, for example, uses a deep neural network to assign each base to a structural category (intergenic, UTR, coding, or intron) and then resolves these base-wise predictions into gene models with an HMM (Stiehler et al., 2021). Because a single model is trained once per lineage and then applied uniformly across genomes, such tools can produce internally consistent, harmonized annotations at scale, sidestepping much of the pipeline-to-pipeline heterogeneity discussed above. Helixer's models are trained on curated NCBI RefSeq annotations spanning hundreds of high-quality genomes within a clade, so that the network learns a reference-grade

representation of gene structure that generalizes to newly sequenced species of the same lineage (Stiehler et al., 2021; Holst et al., 2023). Dedicated pre-trained models have since been released for fungal and invertebrate genomes alongside those for plants and vertebrates (Holst et al., 2023). This paradigm is developing quickly, with recent tools coupling deep learning to HMMs (Gabriel et al., 2024), to large DNA foundation models (Liu et al., 2025), and to lineage-tailored architectures that push accuracy further, including for fungal genomes (Vader et al., 2026). However, applying a single model uniformly across genomes harmonizes gene models rather than necessarily improving them. The aim is to make annotations mutually comparable, not to claim that any individual gene model is more accurate than an author-provided one. Deep-learning predictors are themselves imperfect and carry their own systematic biases, so a harmonized annotation is best understood as internally consistent rather than error-free (Sugihara et al., 2026).

To study NLR diversity and molecular evolution in plants across different evolutionary depths, I compiled three large-scale datasets from public genomes at three distinct taxonomic resolutions. All three were assembled following the same workflow (Figure 2.1), which begins from high-quality public genome assemblies and branches on a single question: whether harmonized, high-quality annotations already exist for the genomes in question. Where they do, as for RefSeq, the existing proteomes can be used directly; where they do not, gene models are generated de novo with Helixer so that the whole collection rests on a common annotation footing. Both branches then converge on NLR annotation with NLRtracker, followed by curation of the resulting NLR set.

The first dataset comprises 180 RefSeq-annotated green plant (Viridiplantae) genomes and takes the first of these branches. Because these assemblies were accompanied by high-quality reference annotations, I ran NLRtracker (Kourelis et al., 2021) directly on their proteomes to annotate protein domains and extract the NLR repertoire (Toghani and Kamoun, 2024). To examine the phylogeny and evolution of the NRC network more closely, I compiled a second dataset of 230 superasterid genomes. Since many of these lacked pre-existing annotations, I annotated the entire set uniformly using the deep-learning-based gene predictor Helixer (Stiehler et al., 2021) to generate harmonized gene models, subsequently extracting the NLRs with NLRtracker (Toghani et al., 2025b). Finally, to resolve the intra- and inter-specific evolution of Solanaceae NLRs, I assembled a third dataset of 346 Solanaceae genomes. As with the superasterid dataset, I generated de novo annotations using Helixer and extracted the corresponding NLRs (Toghani et al., 2026c). In this chapter, I describe each of these NLR datasets in detail, establishing the methodological foundation for the work presented in the subsequent chapters.

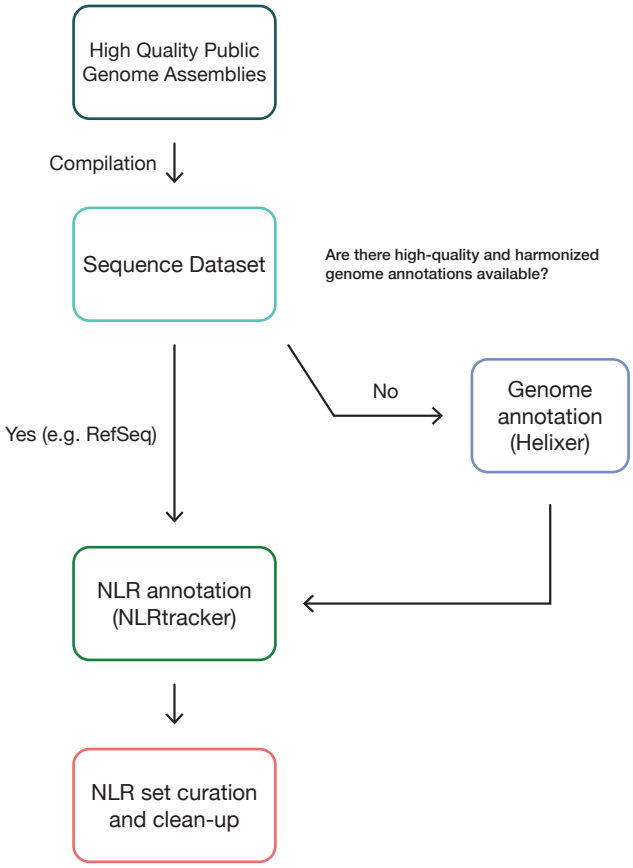


Figure 2.1: Workflow used to compile the harmonized NLR datasets described in this chapter. Starting from high-quality public genome assemblies, I compiled sequence datasets and evaluated whether harmonized, high-quality genome annotations were already available. Where suitable annotations existed (such as RefSeq reference genome annotations (Pruitt et al., 2007)), I used them directly. Where they were lacking, I generated gene models using the *ab initio* gene predictor Helixer (Stiehler et al., 2021; Holst et al., 2023) to ensure all genomes in the dataset shared a uniform annotation standard. Both workflows were followed by NLR annotation with NLRtracker (Kourelis et al., 2021), and subsequently by manual curation and filtering of the resulting NLR dataset.

2.2 Functional annotation of 180 RefSeq reference plant proteomes reveals a dataset of 113,684 NLR proteins

The first and taxonomically broadest dataset was assembled from the NCBI RefSeq database (Pruitt et al., 2007), which provides curated, internally consistent annotations for a limited set of reference genomes. I selected 180 reference plant (Viridiplantae; up to September 2024) proteomes spanning a wide phylogenetic range, from algae and gymnosperms through early-diverging flowering plants to monocots and dicots, to survey NLR diversity across the green plant lineage on a uniformly annotated foundation. Because these proteomes were already annotated to RefSeq's reference standard, no

de novo gene prediction was required, and I applied NLRtracker (Kourelis et al., 2021) directly to each proteome. NLRtracker integrates InterProScan (Blum et al., 2025) to identify the constituent domains and motifs of NLR proteins and classifies sequences according to the NLR definition established for RefPlantNLR (Kourelis et al., 2021). Where NCBI-provided annotation BUSCO scores were available (156 of the 180 genomes), completeness was uniformly high (median 98.4%; Figure 2.2A), in line with the reference-grade quality expected of RefSeq annotations.

Using this pipeline, 169 of the 180 species contained at least one predicted NLR. All eleven species without a single NLR were chlorophyte green algae. Across the 169 proteomes, NLRtracker cataloged 113,684 NLRs, distributed highly unevenly (Figure 2.2). Per-genome counts ranged from none to 4,155, with a median of 468 and a mean of 632, and NLRs accounted for a median of 1.0% of each proteome (up to 7.7%). Counts varied by two orders of magnitude, from 33 in *Cucurbita maxima* to 4,155 in *Quercus robur*, which is consistent with the lineage-specific expansion and contraction thought to characterize this gene family. Proteome size was only weakly associated with NLR count (Pearson $r = 0.42$; Figure 2.2B), suggesting that repertoire size is not simply a function of overall gene number.

Classifying NLRs by their N-terminal signaling domain recovered a structural breakdown broadly in line with expectations (Figure 2.2D). CC-NLRs were the most abundant class (53,708; 47.2%), followed by TIR-NLRs (27,076; 23.8%) and a smaller set of CC_R-NLRs / RPW8-type helper NLRs (1,360; 1.2%). A further 16.6% could not be unambiguously resolved among the CC/CC_R/CC_{G10} types, and 10.8% remained undetermined. Unlike TIR-only proteins, TIR-NLRs were absent from monocots, recapitulating the well-documented loss of the TIR-NLR class in this lineage (Tarr and Alexander, 2009; Shao et al., 2016; Liu et al., 2023b) and suggesting that the dataset recovers established evolutionary patterns.

Beyond NLR extraction, NLRtracker returned functional annotations for the complete proteome of each species, providing a comparative framework for the domain- and motif-level analyses developed in the following chapters. For each proteome, the pipeline produces identifier lists, GFF3 annotations, and FASTA sequences for both NLRs and NLR-associated genes, together with full-length and deduplicated NB-ARC domain sequences, iTOL-ready domain annotation files, a per-sequence domain metadata table, and the raw InterProScan output. The complete dataset, along with a RefSeq species metadata table and a per-genome table of proteome and NLR statistics, is openly available on Zenodo (Toghani and Kamoun, 2024).

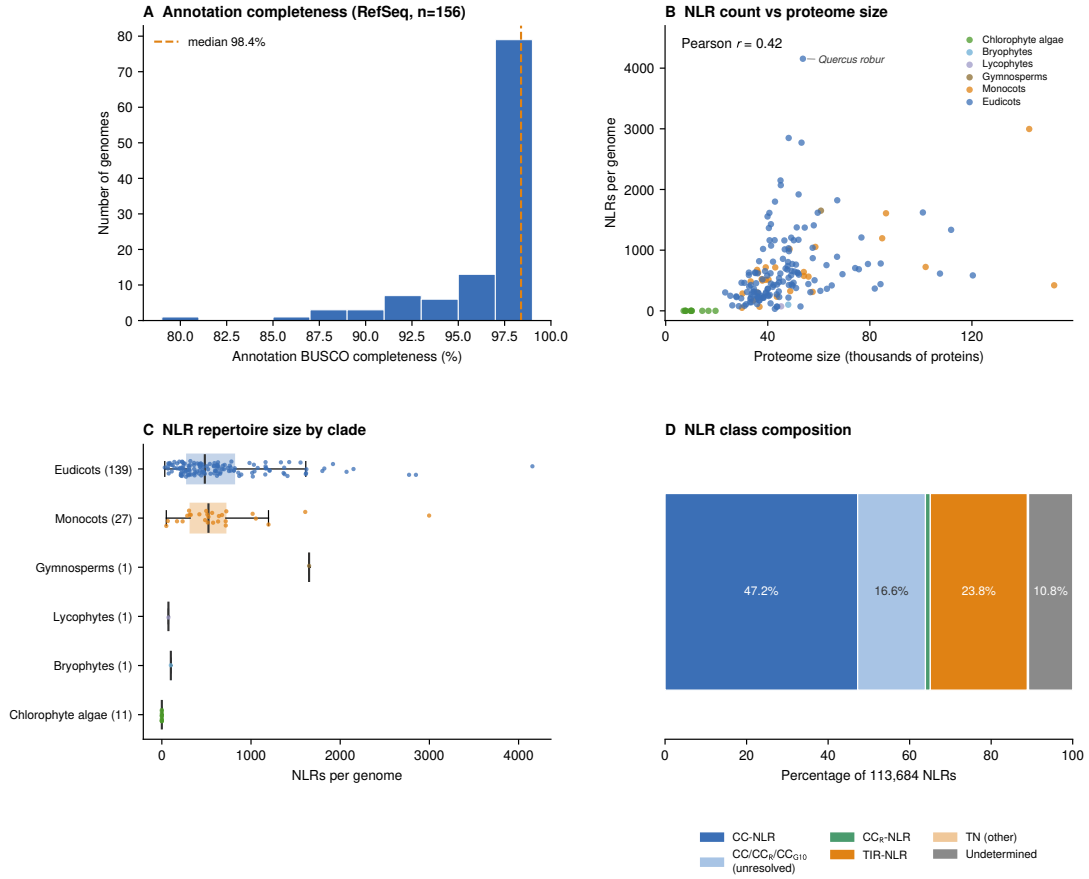


Figure 2.2: Overview of the RefSeq reference-plant NLR dataset.
(A) Distribution of NCBI annotation BUSCO completeness for the 156 of 180 genomes with available scores; the dashed line indicates the median (98.4%). **(B)** NLR count per genome against proteome size, colored by major clade; NLRs are absent from the chlorophyte algae and expand in land plants, with *Quercus robur* carrying the largest repertoire. **(C)** NLR repertoire size by clade, ordered from land plants to algae; the number of genomes per clade is given in parentheses. **(D)** Composition of the 113,684 NLRs by N-terminal signaling domain class, as assigned by NLRtracker.

2.3 Deep-learning-based annotation of 230 superasterid genomes reveals a harmonized dataset of 91,366 NLRs

The second dataset was assembled to support phylogenetic and evolutionary analysis of the NRC network, and therefore narrows the taxonomic scope to the superasterids while increasing sampling density. I collected 230 high-quality, publicly available superasterid genomes, representing 227 distinct taxa across 59 families and 14 orders. Potato (*Solanum tuberosum*) is represented by four independent assemblies, while all other taxa are represented by a single genome. In contrast to the RefSeq collection,

many of these genomes lacked author-provided annotations, and those that were available had been generated with different pipelines and parameter settings. To eliminate this as a source of variation, I set aside the existing annotations and produced harmonized *de novo* annotations for the entire set using the deep-learning annotator Helixer v0.3.2 (Stiehler et al., 2021; Holst et al., 2023). This was run with the `--lineage land_plant` option so that a single land-plant model was applied uniformly across all genomes. This workflow yielded 10,124,265 predicted protein sequences, with a median proteome size of 38,192 (interquartile range 31,628–48,612). Coding sequences and proteomes were extracted from the Helixer GFF output with GffRead v0.12.7 (Pertea and Pertea, 2020), and NLRs and NLR-associated sequences were identified with NLRtracker (Kourelis et al., 2021) using InterProScan for domain assignment (Blum et al., 2025).

Annotation completeness was assessed with BUSCO v5.5.0 against the `viridiplantae_odb10` lineage dataset (Manni et al., 2021). The harmonized annotations appeared to be of consistently high quality (Figure 2.3A): BUSCO completeness averaged 95.1% (median 96.7%), with 89.1% of genomes exceeding 90% completeness and 67.0% exceeding 95%. Missing and fragmented BUSCOs were correspondingly low (median 0.7% and 2.4%, respectively). Only five genomes fell below 80% completeness: ***Striga hermonthica*** (65.9%), ***Jacaranda mimosifolia*** (70.4%), ***Lactuca virosa*** (72.9%), ***Drosera capensis*** (78.6%), and ***Artemisia tridentata*** subsp. ***tridentata*** (79.3%). In most of these the deficit was driven by fragmented rather than missing BUSCOs, as in ***J. mimosifolia*** (23.1% fragmented versus 6.5% missing) and ***S. hermonthica*** (20.2% versus 13.9%), which points to fragmented underlying assemblies rather than a failure of the annotation model; ***L. virosa*** was the exception, with 20.0% of BUSCOs missing outright. It is also worth noting that two of these species, the obligate parasite ***S. hermonthica*** and the carnivorous ***D. capensis***, have specialized lifestyles associated with genuine gene loss (Li et al., 2025), so reduced BUSCO completeness in such taxa need not reflect annotation quality alone. These five genomes span nearly the full range of NLR content observed across the dataset, from 86 in ***D. capensis*** to 1,246 in ***A. tridentata*** (Figure 2.3B), and BUSCO completeness showed no correlation with NLR count across the collection as a whole (Pearson $r = 0.14$). With the exceptions discussed below, this suggests that the variation in NLR repertoire size described later is unlikely to be driven primarily by differences in annotation quality.

Five genomes exceeded 100,000 predicted proteins, and these fall into two apparent classes. In ***Craterostigma plantagineum*** (199,483 proteins), 76.9% of BUSCOs were duplicated against only 22.4% single-copy, which most likely reflects genuine retention of homoeologous gene copies in a polyploid genome; the same signature, in milder form, probably explains the elevated duplication seen in ***Fraxinus americana*** (98.8%

duplicated) and *Nicotiana tabacum* (94.4%). The remaining cases show the opposite signature: *Diospyros kaki* (181,147), *Dorcoceras hygrometricum* (149,066), and *Artemisia tridentata* subsp. *tridentata* (132,727) all retain single-copy BUSCO profiles (1.4–1.6% duplicated), so haplotype redundancy is unlikely to account for their inflated gene counts. This is clearest for *D. kaki*: although the species is an autohexaploid, the assembly used here is a haploid representation, and its single-copy BUSCO profile indicates that retained homoeologs are unlikely to be the source of its 181,147 predicted proteins. Consistent with this, the proteomes obtained for the other two species substantially exceed the published evidence-based annotations for the same genomes: 43,377 genes were reported for *A. tridentata* (Melton et al., 2022) and 49,374 for *D. hygrometricum* (Xiao et al., 2015), roughly a third of the counts obtained here. All three are large or difficult to annotate—the *A. tridentata* assembly is 4.2 Gb and 77.99% repetitive, among the most repetitive plant genomes sequenced to date (Melton et al., 2022), while the *D. hygrometricum* assembly is an early short-read draft—and both show elevated fragmented BUSCOs (16.5% and 11.3%, respectively). The most likely explanation is therefore over-prediction of gene models in repeat-rich and fragmented sequence, where transposon-derived open reading frames and split loci can be readily called as genes. This failure mode would also be largely invisible to BUSCO, which assays whether a conserved core gene set is present but is insensitive to the addition of spurious models, so high BUSCO completeness should not by itself be read as evidence against over-annotation. NLR counts in these genomes should therefore be interpreted with corresponding caution, particularly for *A. tridentata*, whose 1,246 NLRs rank among the highest in the dataset.

Applying the NLRtracker definition, this pipeline recovered 91,366 NLRs across the 230 genomes. Per-genome counts spanned more than two orders of magnitude, from 12 to 2,049, with a median of 338 and a mean of 397, and NLRs accounted for a median of 0.80% of each proteome, up to 3.25%. As in the RefSeq dataset, proteome size explained only part of this variation (Pearson $r = 0.42$). The smallest repertoires belonged to parasitic and carnivorous lineages, namely *Cuscuta campestris* (12 NLRs), *Orobanche coerulescens* (19), and *Utricularia gibba* (32), consistent with the reduced immune receptor repertoires reported for plants with specialized lifestyles and correspondingly altered pathogen exposure (Liu et al., 2021; Baggs et al., 2020). The largest repertoires were found in the Lamiaceae, led by *Perilla frutescens* (2,049), *Pogostemon cablin* (1,479), and *Thymus mandschuricus* (1,466). At the order level, median NLR content varied several-fold, from 82 in Santalales and 167 in Caryophyllales to 401 in Solanales, 582 in Cornales, and 702 in Dipsacales, with the two most densely sampled orders, Lamiales ($n = 58$) and Solanales ($n = 40$), showing broad within-order dispersion (Figure 2.3C).

Classifying NLRs by their N-terminal signaling domain gave a breakdown broadly similar

to the RefSeq dataset (Figure 2.3D), with CC-NLRs the dominant class (47,276; 51.7%), followed by TIR-NLRs (12,606; 13.8%) and CC_R-NLRs / RPW8-type helper NLRs (884; 1.0%). A further 21.1% could not be unambiguously resolved among the CC/CC_R/CC_{G10} types, and 12.0% remained undetermined. The lower TIR-NLR fraction relative to the RefSeq green plant dataset (13.8% versus 23.8%) likely reflects the asterid-focused sampling of this collection.

Genome annotations are organized by phylogenetic order following the APG IV classification (The Angiosperm Phylogeny Group, 2016), with each order held in its own subdirectory containing the genome assembly, GFF annotation, CDS and protein FASTA files, and the corresponding NLRtracker output; compiled proteome and CDS files carrying source-assembly tags are also provided. As for the RefSeq dataset, the NLRtracker output comprises identifier lists, GFF3 annotations, and FASTA sequences for NLRs and NLR-associated genes, full-length and deduplicated NB-ARC domain sequences, iTOL-ready domain annotation files, a per-sequence domain metadata table, and the raw InterProScan output. The complete dataset, together with a species metadata table and a per-genome table of proteome, NLR, and BUSCO statistics, is openly available on Zenodo and Dryad (Toghani et al., 2025b).

2.4 Harmonized genome annotation of 346 Solanaceae genomes reveals 197,834 NLR immune receptors

The third and most finely resolved dataset narrows the scope to a single family. The Solanaceae encompass several agriculturally important crops, including potato, tomato, eggplant, pepper, and tobacco, alongside a rich diversity of wild relatives that serve as reservoirs of genetic variation. I assembled a set of 346 publicly available Solanaceae genomes spanning 85 species and 9 genera. Because many species are represented by multiple independent assemblies, the dataset supports comparison at both the interspecific and the intraspecific level, the resolution on which the analyses developed in later chapters relate to. This release builds directly on an earlier Solanaceae NLRome resource, which cataloged 66,665 NLRs from 124 published Solanaceae proteomes (Sugihara et al., 2023). That earlier dataset relied on heterogeneous, author-supplied annotations and so was subject to the pipeline-to-pipeline inconsistency described above. Here, the sampling was expanded to 346 genomes, and the author-supplied annotations were replaced throughout with harmonized *de novo* annotations generated using the same Helixer-based workflow applied to the superasterid dataset (Stiehler et al., 2021; Holst et al., 2023; Pertea and Pertea, 2020). NLRs and NLR-associated sequences were identified with NLRtracker (Kourelis et al., 2021) using

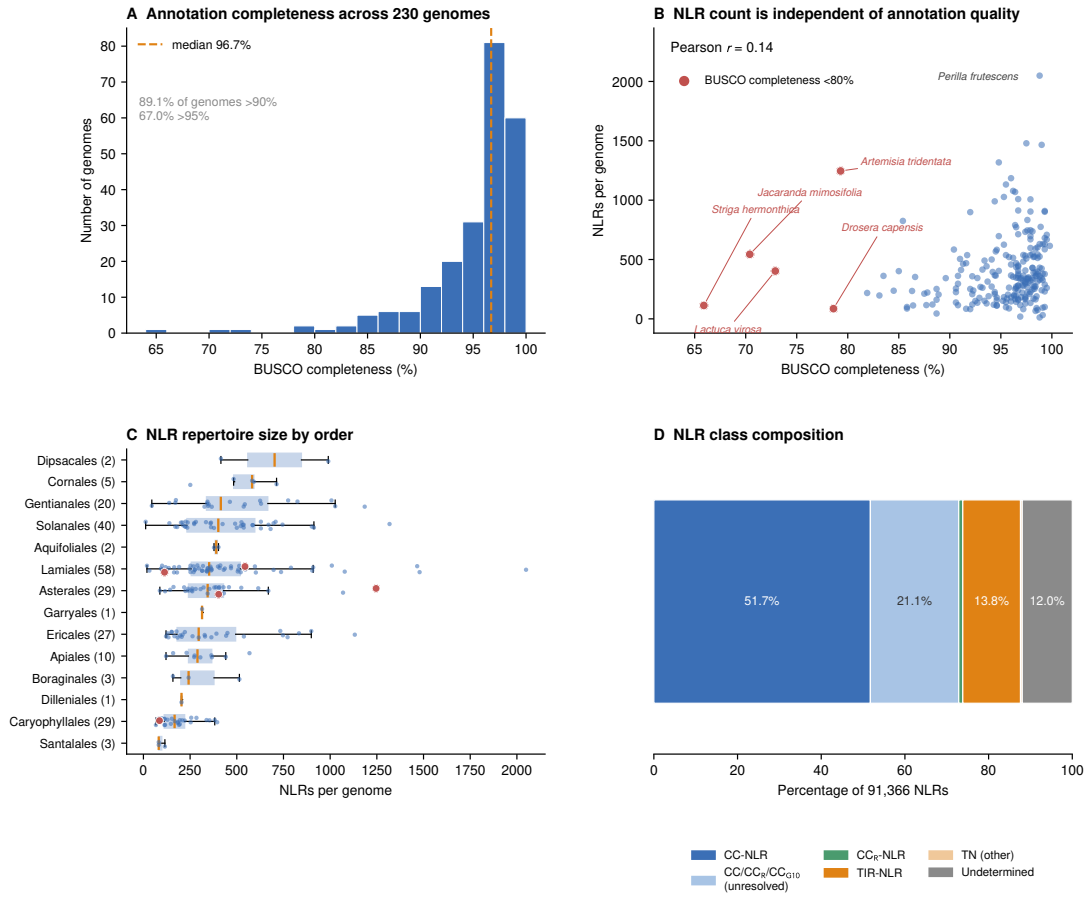


Figure 2.3: Overview of the harmonized superasterid NLR dataset.
(A) Distribution of BUSCO completeness across the 230 Helixer-annotated genomes, scored against `viridiplantae_odb10`; the dashed line indicates the median (96.7%). **(B)** NLR count per genome plotted against BUSCO completeness. The five genomes with completeness below 80% are highlighted in red and labeled. The absence of correlation (Pearson $r = 0.14$) indicates that repertoire size is not driven by annotation quality. **(C)** NLR repertoire size by taxonomic order, ordered by median; points show individual genomes and the number of genomes per order is given in parentheses. **(D)** Composition of the 91,366 NLRs by N-terminal signaling domain class, as assigned by NLRtracker.

InterProScan for domain assignment (Blum et al., 2025), and annotation completeness was assessed with BUSCO v6.0.0 against the family-appropriate `solanales_odb12` lineage dataset (Manni et al., 2021).

Across the 346 genomes the pipeline predicted 15,078,455 protein sequences, with a median proteome size of 35,970 (interquartile range 32,587–42,519). The harmonized annotations appeared to be of high and consistent quality (Figure 2.4A), with BUSCO completeness averaging 96.6% (median 97.8%), 97.4% of genomes exceeding 90% completeness, and 78.6% exceeding 95%. Both fragmented and missing BUSCOs were low, at a median of 1.0% and 1.1%, respectively. Elevated BUSCO duplication was common, with 49 genomes exceeding 50% duplicated BUSCOs. Because duplicated

single-copy orthologs are a recognized signature of polyploidy (Manni et al., 2021), and the Solanaceae include the autotetraploid cultivated potato as well as several other polyploid lineages (Fox et al., 2020; Sun et al., 2025), this pattern most likely reflects the retention of homoeologous gene copies rather than an annotation artifact, though the two cannot be fully disentangled from BUSCO scores alone. Reassuringly, BUSCO completeness showed no correlation with NLR count across the dataset (Pearson $r = 0.09$), as in the other two collections, which suggests that variation in NLR repertoire size is unlikely to be driven primarily by differences in annotation quality.

Applying the NLRtracker definition, this pipeline recovered 197,834 NLRs. Per-genome counts ranged from 139 to 2,675, with a median of 600 and a mean of 572, and NLRs accounted for a median of 1.44% of each proteome. Within this taxonomically narrow sample, proteome size was more strongly associated with NLR count than in the broader datasets (Pearson $r = 0.58$; Figure 2.4B), which would be consistent with polyploid genomes tending to carry both more genes overall and more NLRs. At the genus level, the pepper genus ***Capsicum*** ($n = 30$) had the highest median repertoire (630 NLRs), closely followed by ***Solanum*** ($n = 294$; median 610), while ***Nicotiana*** ($n = 11$) carried markedly fewer (median 281) (Figure 2.4C).

The value of this dataset lies in its dense within-species sampling. Nineteen species are represented by more than one genome, including 110 assemblies of ***Solanum tuberosum*** (potato), 59 of ***S. lycopersicum*** (tomato), and 24 each of ***Capsicum annuum*** (pepper) and ***S. candolleanum***. This sampling offers a first view of the extent of intraspecific variation in NLR content (Figure 2.4D). In cultivated tomato, NLR counts were relatively stable across all 59 assemblies (215–244, representing a spread of roughly 13%), which may point to limited presence/absence variation at the species level and to good reproducibility of the harmonized pipeline, although some of this stability could also reflect the comparatively uniform, high-quality assemblies available for tomato. In potato, by contrast, counts ranged far more widely, from 549 to 2,675 across 110 assemblies. This dispersion appears to track the varying ploidy and haplotype content of the potato assemblies, with the highest counts tending to come from genomes with elevated BUSCO duplication, and is therefore probably better interpreted in that light than as straightforward differences in gene copy number. These patterns are suggestive rather than definitive and disentangling true intraspecific variation from assembly- and ploidy-related effects will require the more controlled comparisons developed in the following chapters.

Classifying NLRs by their N-terminal signaling domain gave a breakdown consistent with the other datasets, with CC-NLRs the dominant class (111,968; 56.6%), followed by TIR-NLRs (27,724; 14.0%) and CC_R-NLRs / RPW8-type helper NLRs (1,117; 0.6%). A further 16.3% could not be unambiguously resolved among the CC/CC_R/CC_{G10} types, and 12.3%

remained undetermined.

The dataset provides the underlying genome assemblies, the complete Helixer-derived annotation outputs (proteome, CDS, and GFF files), and the full NLRtracker output archives, following the same file organization as the superasterid resource. It is openly available on Zenodo (Toghani et al., 2026c) and is intended to serve as a harmonized foundation for comparative genomics, NLR evolution, and immune-receptor engineering in the nightshade family.

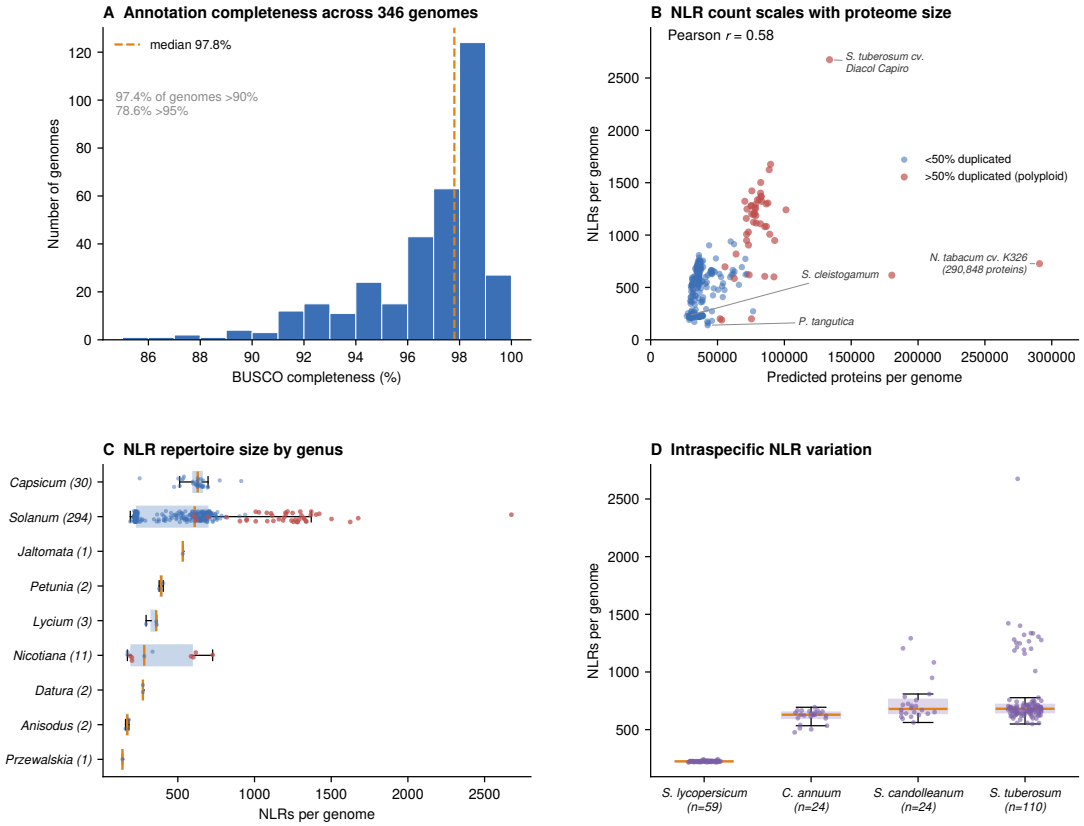


Figure 2.4: Overview of the harmonized Solanaceae NLR dataset.
(A) Distribution of BUSCO completeness across the 346 Helixer-annotated genomes, scored against `solanales_odb12`; the dashed line indicates the median (97.8%). **(B)** NLR count per genome against predicted proteome size, colored by BUSCO duplication level; genomes with more than 50% duplicated BUSCOs (red) are polyploid and tend to carry both larger proteomes and more NLRs. Assemblies at the extremes are labeled: *S. tuberosum* cv. Diacol Capiro (most NLRs), *N. tabacum* cv. K326 (largest proteome), *S. cleistogamum* (smallest proteome), and *Przewalskia tangutica* (fewest NLRs). **(C)** NLR repertoire size by genus, ordered by median; points show individual genomes and the number of genomes per genus is given in parentheses. **(D)** Intraspecific NLR variation for the four most densely sampled species.

2.5 Conclusions and Discussion

In this chapter I curated three NLR datasets spanning nested taxonomic resolutions, together comprising 756 plant genomes and more than 400,000 cataloged NLRs. The three collections were designed to complement one another: 180 RefSeq reference proteomes provide a broad, uniformly annotated view of NLR diversity across the green plant lineage; 230 superasterid genomes narrow the focus to the phylogenetic distribution of the NRC network while increasing sampling density; and 346 Solanaceae genomes add the depth needed to compare NLR repertoires within, as well as between, species. Except for the RefSeq set, which was already annotated to a consistent standard, all genomes were re-annotated de novo with a single Helixer model and processed through an identical NLRtracker pipeline, so that differences between genomes reflect biology rather than the pipeline-to-pipeline heterogeneity that motivated this work.

Across all three datasets, the harmonized annotations were of high completeness, and NLR content was uncorrelated with annotation quality, suggesting that the repertoire variation described here is largely biological in origin. It is important, however, to be clear about what harmonization does and does not achieve. Re-annotating every genome with a single Helixer model makes gene models comparable across the collection, but it does not make them perfect. Like any annotation pipeline, Helixer has its own systematic limitations, and it can over-predict genes in repeat-rich or fragmented assemblies, miss atypical or highly divergent NLRs, and inflate counts in polyploid genomes (Sugihara et al., 2026). The datasets presented here therefore trade per-genome optimality for cross-genome consistency; their value lies in comparability rather than in any claim to be the definitive annotation of a single genome. With these limitations in mind, the collections nonetheless recovered well-established patterns, including the absence of NLRs in green algae and the loss of the TIR-NLR class in monocots, providing confidence that they capture genuine evolutionary signal. Together, they establish a consistent foundation for the comparative and evolutionary analyses performed in the following chapters.

Chapter 3: Functional and molecular evolution of the NLR proteins beyond R-AVR co-evolution

Part of the work described in this chapter has been published in Selvaraj et al. (Selvaraj et al., 2024) and appears here with permission. Cryo-EM structure determination was carried out by Muniyandi Selvaraj. Biochemical and cell death assays for NbNRC2 homodimer section were performed by Mauricio Contreras and Muniyandi Selvaraj. My contribution comprised the phylogenomics and the structural analyses described below.

3.1 Introduction

Nucleotide-binding and leucine-rich repeat (NLR) immune receptors are essential components of innate immunity across the tree of life, detecting intracellular pathogen components and initiating immune responses to counteract infection (Jones et al., 2016; Contreras et al., 2023a). Plant NLRs are activated by pathogen-secreted virulence proteins, termed effectors, which pathogens deliver into host cells to modulate host physiology (Jones and Dangl, 2006). They belong to the signal transduction ATPases with numerous domains (STAND) superfamily (Leipe et al., 2004) and typically exhibit a tripartite domain architecture consisting of an N-terminal signaling domain, a central nucleotide-binding domain, and C-terminal superstructure-forming repeats (Takken et al., 2006). The central domain, termed NB-ARC (nucleotide-binding adaptor shared by APAF-1, plant R proteins, and CED-4) in plant NLRs, is a signature of this protein family and acts as a molecular switch, mediating the conformational changes required for activation (Takken et al., 2006). A hallmark of NLR activation in eukaryotes and prokaryotes is oligomerization into higher-order immune complexes, termed resistosomes in plants or inflammasomes in mammalian and bacterial cells. These higher order immune complexes initiate immune signaling through diverse mechanisms and often culminate in programmed cell death (Wang et al., 2019b; Bi et al., 2021; Förderer et al., 2022).

Plant NLRs fall into different phylogenetic classes carrying distinct N-terminal domains, of which the coiled-coil (CC) type is the most widespread (Shao et al., 2019). A canonical plant CC-NLR is *Arabidopsis thaliana* AtZAR1, which is held in a resting monomeric state in complex with its partner pseudokinase RKS1. Perception of its cognate effector AvrAC primes individual AtZAR1 monomers, which leads to ADP to ATP exchange and undergo conformational changes and subsequently assemble into pentameric resistosomes

(Wang et al., 2019b; Bi et al., 2021). Following resistosome assembly, CC-NLR oligomers insert into the plasma membrane and act as calcium-permeable channels, initiating immune signaling and cell death, as shown for AtZAR1 and the wheat CC-NLR TmSr35 (Bi et al., 2021; Förderer et al., 2022). NLRs can operate in different configurations: as functional singletons, as sensor–helper pairs, or within more elaborate immune receptor networks (Adachi et al., 2019b; Wu et al., 2017). Functional singletons such as AtZAR1 and TmSr35 combine both effector perception and immune signaling in a single genetic unit (Adachi et al., 2019b; Saintenac et al., 2013). Many other NLRs, however, have subfunctionalized over evolutionary time into sensors dedicated to pathogen perception and helpers that execute immune signaling, as exemplified by the rice NLR pairs Pik-1/Pik-2 and RGA5/RGA4 (Césari et al., 2014; Maqbool et al., 2015). In the NRC (NLR required for cell death) network of solanaceous plants, multiple sensor CC-NLRs signal redundantly through an array of downstream paralogous helper CC-NLRs to confer disease resistance to diverse pathogens and pests, including oomycetes, fungi, nematodes, insects, viruses, and bacteria (Wu et al., 2017; Derevnina et al., 2021). Effector perception by NRC-dependent sensors (NRC-S) triggers conformational changes in their central nucleotide-binding domain that signal to downstream helper NRCs, activating them, and driving their oligomerization into a helper resistosome (Contreras et al., 2022; Contreras et al., 2023c; Ahn et al., 2023; Toghani et al., 2026a). Assembly into a resistosome is a shared endpoint between AtZAR1 and the NRCs, although the NRC assemblies described to date are hexameric or undecameric rather than pentameric (Liu et al., 2023a; Madhuprakash et al., 2024; Seager et al., 2026; Toghani et al., 2026b).

The N-terminal region is central to this activation mechanism. The MADA motif, located in the N-terminal α1 helix of about 20% of all angiosperm CC-NLRs, is required for plasma membrane integration of resistosome and the induction of cell death (Adachi et al., 2019a). This motif, along with related sequences such as the MAEPL motif found in non-flowering plants, is functionally conserved across evolutionarily distant NLRs capable of triggering cell death (Adachi et al., 2019a; Adachi et al., 2023b). MADA-type NLRs include the ancient singleton ZAR1 as well as helper NLRs from distinct phylogenetic clades such as Pik-2 and NRC4, pointing to a shared origin for these CC-NLRs and indicating that the MADA-type α1 helix has been a feature of the family since the emergence of land plants (Adachi et al., 2019a). As sensor NLRs specialized in pathogen detection, however, they appear to have lost or degenerated their MADA motif, rendering them unable to initiate immune signaling on their own. Consistent with this, NRC helpers carry a MADA motif whereas NRC-S do not, and the N-termini of MADA- or MAEPL-containing NLRs can functionally replace the N-terminus of the helper NRC4, while the N-termini of sensor CC-NLRs cannot (Adachi et al., 2019a;

Adachi et al., 2023b). These observations fit an evolutionary "use-it-or-lose-it" model, in which the conserved N-terminal motifs of sensor NLRs degenerated once these receptors came to rely on helpers for cell death execution.

NLRs are among the most evolutionarily dynamic families of plant proteins and are under constant pressure to evolve, whether to gain recognition of new pathogen effectors or effector targets, or to evade effector-mediated suppression (Seong et al., 2020; Prigozhin and Krasileva, 2021). This pressure is most often framed in terms of R–AVR co-evolution, which is usually observed in NLR domains involved in effector recognition. Yet the transition from multifunctional singleton ancestors to networks of specialized partners suggests that additional selective pressures act on NLRs (Adachi et al., 2019b; Wu et al., 2017). The division of labor between NRC-S and NRC-H relaxes evolutionary constraints on individual receptors, allowing faster diversification and functional refinement. It helps unlock greater evolutionary flexibility. Sensor NLRs can readily acquire additional domains, such as integrated domains (IDs) or the Solanaceous Domain (SD), to expand their detection capabilities (Jubic et al., 2019; Kourelis et al., 2023). A further layer of complexity comes from NLRs that modulate the activity of other NLRs in ways that do not fit the standard sensor–helper pairing. NRCX, a full-length CC-NLR closely related to the helpers NRC2 and NRC3, is one such case: its N-terminal MADA sequence is non-functional, and silencing it in *N. benthamiana* leads to autoimmunity that depends on NRC2 and NRC3, but not NRC4 (Adachi et al., 2023a). NRCX therefore appears to have evolved to regulate the activity of its closest paralogs, likely preventing inappropriate activation of their pathways (Adachi et al., 2023a). Modulators of this kind imply that complex network configurations evolved in part to maintain network balance and prevent misregulation, and that NLR evolution is also shaped by relationships among NLRs themselves rather than just by their interactions with pathogens.

Two aspects of these intra-network relationships remain poorly resolved. First, although the MADA motif is functionally important and broadly interchangeable, it is not known how diverse this region is within NRC helper clade, nor whether the degeneration seen in NRCX represents an isolated event or a recurrent evolutionary outcome. Second, the extent to which the model of NLR activation derived from singletons such as AtZAR1 and TmSr35 applies to helper CC-NLRs is unknown. Paralogous helpers within the NRC network signal redundantly yet remain functionally insulated from one another, and how this insulation is achieved at the level of protein–protein interactions has not been addressed.

In this chapter, I examined both questions using the NRC helper clade of the Solanaceae family as a model system. The NRC helpers are massively expanded in this family, form

a well-supported clade, and have been established experimentally, making them well suited to comparative analysis. I first analyzed the diversity of the N-terminal MADA motif across the clade. I then turned to the resting state of NRC2, which we recently showed to accumulate *in planta* as a homodimer rather than a monomer (Selvaraj et al., 2024). I used phylogenomics across the NRC clade to ask whether the interfaces mediating this self-association have diverged between paralogous helpers. Together, these analyses address how NRC helpers and more broadly NLR proteins have evolved not only in response to pathogens, but also in relation to one another and to their compatibility or incompatibility within a shared network.

3.2 The N-terminal MADA motif has diversified in NRC helpers

While the function and interchangeability of the MADA motif are well established (Adachi et al., 2019a), the sequence variation observed among MADA-containing NLRs, including non-cell-death-inducing variants, points to potential functional divergence within the family (Adachi et al., 2019a; Adachi et al., 2023a). I hypothesized that diversification within the MADA motif could lead to NLR neofunctionalization. To test this, I analyzed the MADA motif diversity across the NRC helper clade, which provides an ideal model system. Using NLRtracker, I constructed an NLRome dataset from 123 Solanaceae proteomes across the genera *Nicotiana*, *Physalis*, *Capsicum*, and *Solanum* (Kourelis et al., 2021; Sugihara et al., 2023). From phylogenetic analysis, I extracted 1,796 NRC sequences for downstream investigation. NRC helpers are massively expanded in the Solanaceae and form a well-supported clade within CC-type NLRs (Selvaraj et al., 2024; Lee et al., 2021; Goh et al., 2024) (Figure 3.1). Using reference NRC sequences (Kourelis et al., 2021; Adachi et al., 2023a; Oh et al., 2023), I resolved 15 major subclades within the helper clade. Although all NRC helpers possess a recognizable MADA motif, some, such as NRCX (Adachi et al., 2023a), do not induce cell death and instead act as negative modulators (Wu et al., 2017; Lee et al., 2021).

Because the MADA motif resides within the N-terminal α1 helix of CC-NLRs, I implemented a computational pipeline to predict secondary structure using NetSurfP-3.0 (Høie et al., 2022) and extract the α1 helix from each sequence. I then clustered these sequences using MCL (Van Dongen, 2008) and mapped the resulting clusters onto the NRC phylogenetic tree (Figure 3.2). Sequence logos from each cluster were used to compare clades. This unbiased clustering matched the phylogenetic relationships of the NRC helpers and revealed that the N-terminal α1 helix has diversified in both length and amino acid composition.

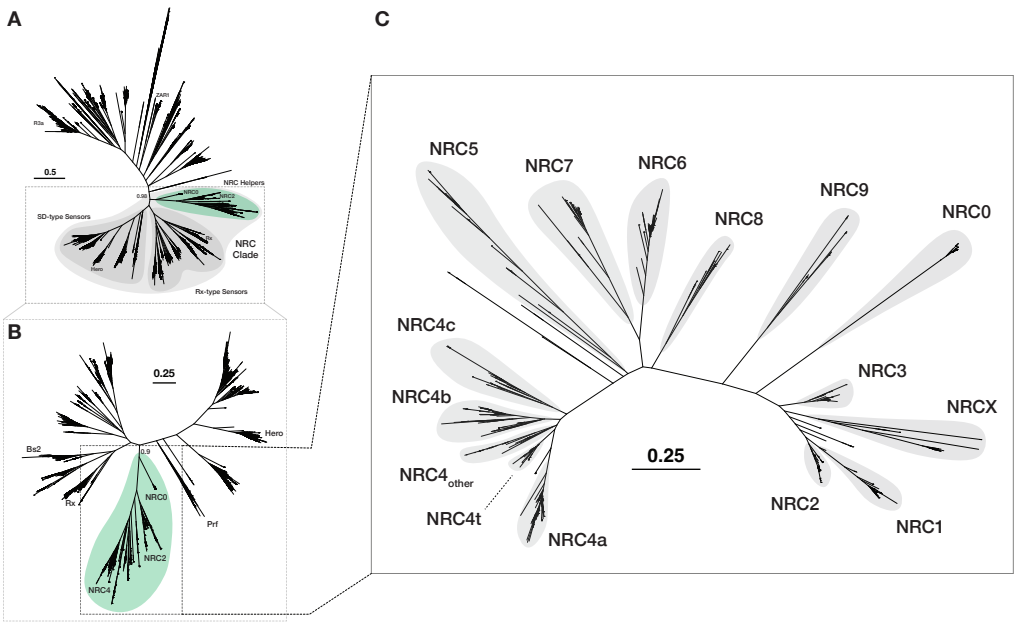


Figure 3.1: NRC sequences form a well-supported monophyletic clade within the CC-NLR and NRC superclade phylogenetic trees.

(A) The NRC superclade, comprising NRC helpers, Rx-type NRC-S, and SD-type NRC-S, forms a distinct and well-supported clade within a phylogenetic tree of 19,760 CC-NLRs from the Solanaceae. (B) 1,796 NRC sequences form a monophyletic clade within the NRC superclade, indicating a shared evolutionary history. (C) Based on reference NRC sequences, 15 distinct subclades are resolved within the NRC clade; subclade relationships are supported by bootstrap values (Selvaraj et al., 2024).

3.2.1 NRCX carries a signature threonine at the fourth position of its MADA motif

Building on previous observations that chimeras between the NRCX N-terminus and NRC4 fail to induce cell death (Adachi et al., 2023a), I compared N-terminal sequences of the NRCX clade with its sister clades, NRC2 and NRC3. NRCX carries a conserved threonine at the fourth position of its MADA motif (T4) that is absent in both NRC2 and NRC3 (Figure 3.3A). I hypothesized that this T4 residue represents a degenerative mutation that renders the MADA motif of NRCX incapable of triggering cell death.

3.2.2 A subset of NRC4 sequences share the NRCX N-terminal signature

Prompted by the identification of the NRCX T4 signature, I screened other NRC clades for the same feature. The T4 signature was also present in the NRC4/5 superclade (Figure 3.3B). Mapping these sequences onto the NRC phylogeny revealed that most NRC4 variants carrying the T4 residue belong to a distinct subclade sister to NRC4a (Figure 3.4).

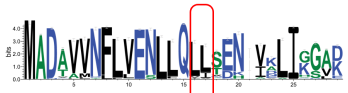
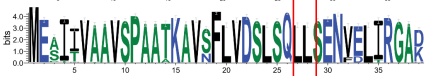
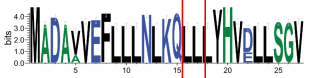
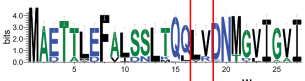
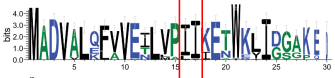
Cluster	No. Seq.	Clade	Consensus Sequence
C1	638	NRC4/5	
C2	358	NRC2/3	
C3	180	NRC6	
C4	140	NRC7	
C5	105	NRC1	
C6	102	NRC0	
C7	50	NRCX	
C8	36	NRC8	
C9	35	NRC1	
C10	25	NRC9	

Figure 3.2: The N-terminal MADA motif has diversified in NRC helpers.
Unbiased clustering of N-terminal sequences from NRC helpers matches their phylogenetic relationships. The N-terminal α1 helix has diversified in both length and sequence. The total number of sequences in each cluster is indicated. The red box marks the alignment of N-terminal sequences at their conserved leucine/isoleucine stretch. Y-axis shows the frequency of each amino acid in the given position within that cluster.

I designated this subclade as "NRC4t". The NRC4t clade mirrors NRCX in two aspects: it forms a distinct lineage closely related to a canonical helper clade (NRC4a), and it carries the characteristic T4 N-terminal signature.

3.2.3 NRC4t proteins induce cell death as autoactive mutants

Given their N-terminal sequence similarity to NRCX, I tested whether representative NRC4t proteins are competent to trigger cell death. I hypothesized that, like NRCX, NRC4t variants would fail to induce cell death as autoactive mutants due to the T4 substitution. To test this, I transiently expressed two potato NRC4t variants (StNRC4p and StNRC4h) alongside autoactive mutants carrying a D-to-V substitution in the MHD motif, and screened for cell death (constructs cloned and provided by Jiorgos Kourelis; Table 3.1).

Contrary to the initial hypothesis, both NRC4t DV mutants also induced strong cell death,

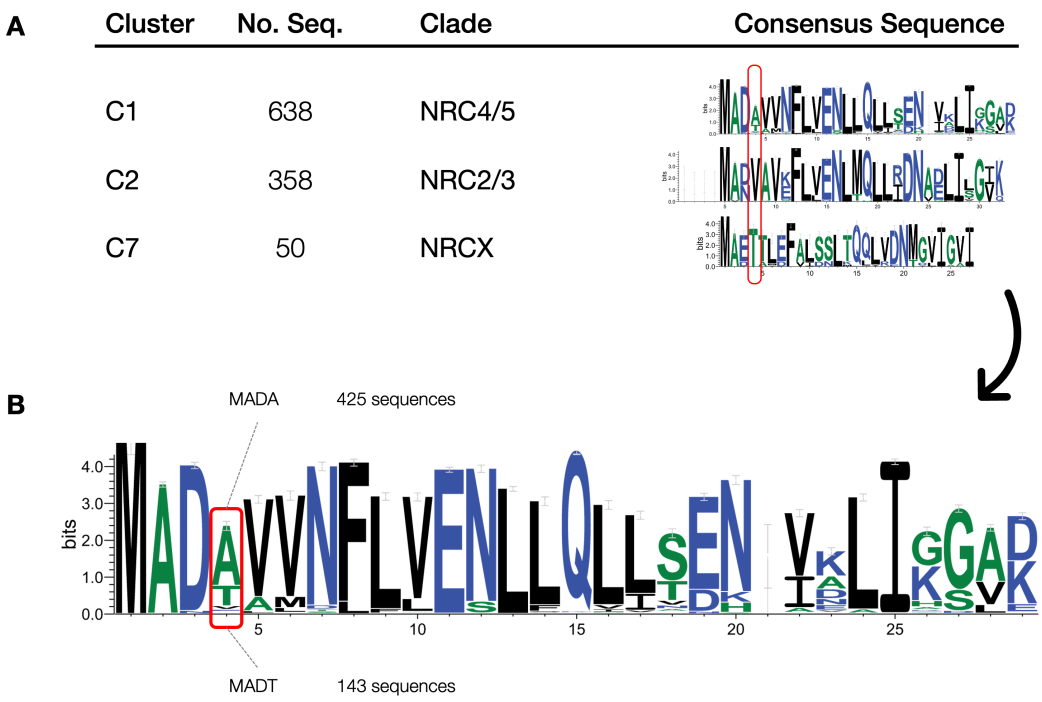


Figure 3.3: NRCX has a distinct threonine at the fourth position of its MADA motif.
(A) NRCX has a unique threonine at the fourth position (T4) of its MADA motif compared with NRC2 and NRC3. (B) The same T4 N-terminal signature is found in approximately 25% of NRC4/5 clade sequences. The red box indicates the fourth position of the MADA motif.

while their wild-type variants did not cause any visible cell death. To confirm this finding and exclude interference from endogenous helpers, I repeated the assay across two independent replicates in an *nrc2/3/4* knockout line of *N. benthamiana*, and obtained consistent results (Figure 3.5B). NbNRC4, a potato NRC4a homolog (StNRC4e), NbNRCX, and StNRCX, together with their corresponding DV mutants, served as controls: wild-type proteins were not expected to cause cell death, NRCX DV mutants served as negative controls (Adachi et al., 2023a), and NRC4 DV mutants served as positive controls. The controls behaved as expected: StNRCX did not cause cell death, whereas StNRC4e triggered cell death as an autoactive DV mutant (Figure 3.5A). These findings demonstrate that NRC4t receptors do not behave like NRCX in this assay; whether they operate as canonical helpers or fulfill specialized regulatory roles on other NRC helpers remains to be determined.

3.3 Resting-state NRC2 forms a homodimer

Recently, we showed that NbNRC2, unlike the canonical functional singleton NbZAR1, self-associates in co-immunoprecipitation assays (Figure 3.6A) (Selvaraj et al., 2024). Moreover, we could not detect hetero-association between NbNRC2 and its paralogs

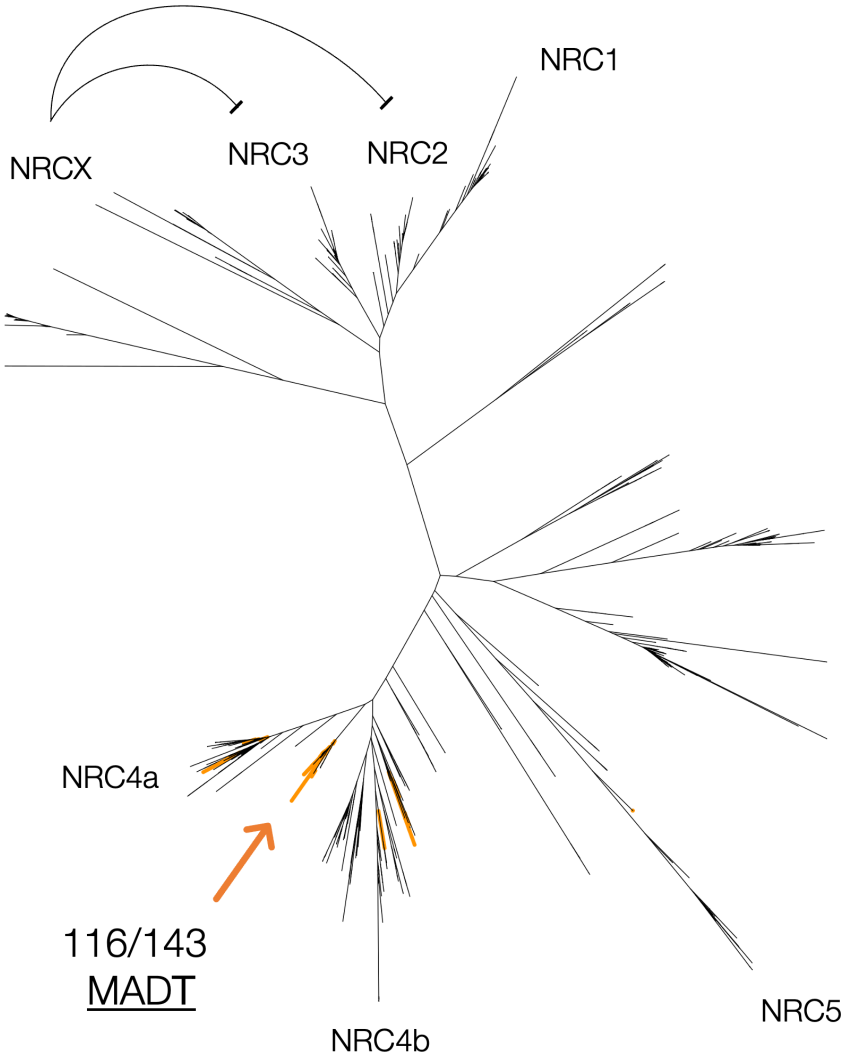


Figure 3.4: NRC4 sequences with the T4 N-terminal signature form a distinct phylogenetic clade in the NRC helper tree.
116 of 143 NRC4 sequences carrying the T4 signature (indicated by the orange line) map to a distinct subclade sister to NRC4a, mirroring the phylogenetic position of NRCX relative to NRC2 and NRC3.

NbNRC3 and NbNRC4, or NbZAR1. This suggests that NbNRC2 preferentially self-associates in *N. benthamiana* and does not form stable heterocomplexes with its paralogs. This conclusion was independently corroborated by live-cell confocal microscopy showing that resting-state NbNRC2 and NbNRC4 do not co-localize *in planta* (Selvaraj et al., 2024).

To investigate the NbNRC2 resting state we performed cryo-EM 2D class averaging (Figure 3.6B) and resolved the atomic model of the resting-state NbNRC2 homodimer. The cryo-EM model included the NBD, HD1-WHD, and LRR domains, while the N-terminal CC domain was absent from the density (Figure 3.6C) (Selvaraj et al., 2024). This struc-

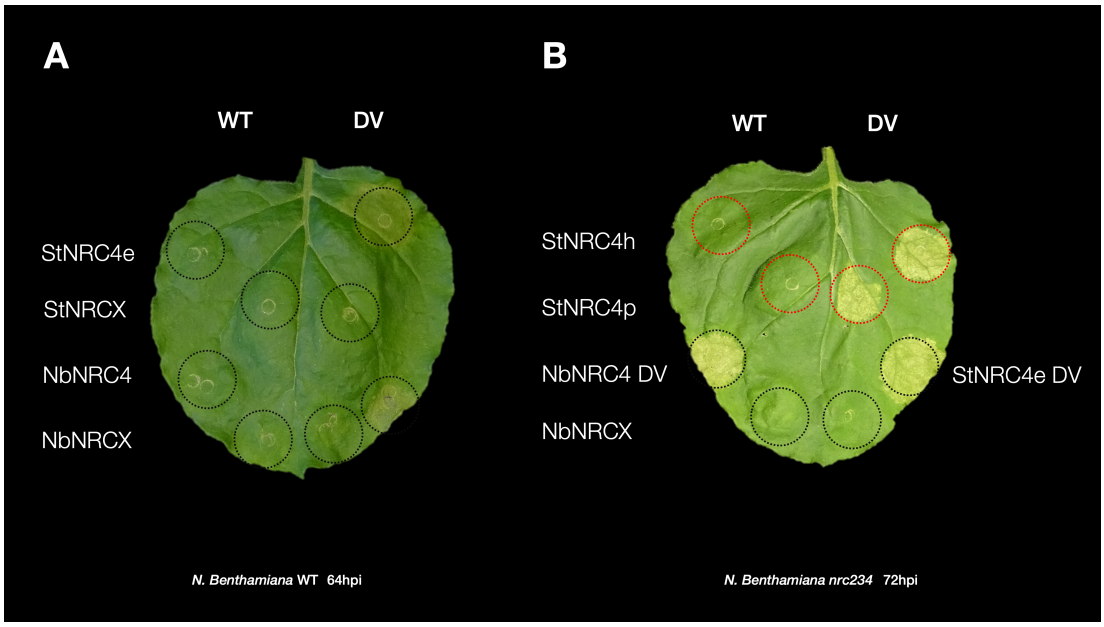


Figure 3.5: NRC4t sequences are capable of inducing cell death as autoactive mutants.
(A) StNRC4e, a homolog of NRC4a, induces cell death in *N. benthamiana* as an autoactive (DV) mutant. Like NbNRCX, StNRCX is unable to induce cell death as either wild type or an autoactive mutant in wild-type *N. benthamiana* (images taken at 64 hpi). (B) The NRC4t variants StNRC4h and StNRC4p (marked by red circles) induced cell death as autoactive DV mutants in an *nrc2/3/4* knockout line of *N. benthamiana* (72 hpi). Neither protein triggered cell death as wild type. Constructs used are listed in Table 3.1.

ture revealed the contact sites between the two protomers, comprising eight stretches that form three main interaction interfaces across the NBD, HD1-WHD, and LRR domains (Figure 3.6C,D) (Selvaraj et al., 2024). Because NbNRC2 does not form stable heterocomplexes with other NRC paralogs despite their shared evolutionary origin, we hypothesized that these homodimerization interfaces diversified across NRC helpers to insulate their association and therefore activation pathways.

3.4 Evolution of homodimerization in NRCs

To test whether the homodimerization interfaces have diversified across NRC helpers, I performed a phylogenomic analysis across the expanded NRC clade of the Solanaceae. Using the 123-proteome Solanaceae NLRome dataset (Sugihara et al., 2023), phylogenetic analysis yielded 993 unique NRC sequences. Together with 99 reference NRC sequences from *N. benthamiana*, *C. annuum*, *S. lycopersicum*, and *S. tuberosum*, these were used to construct a detailed phylogenetic tree of the NRC helper clade. This resolved the 15 subclades noted above, with clade sizes ranging from 31 to 129 sequences.

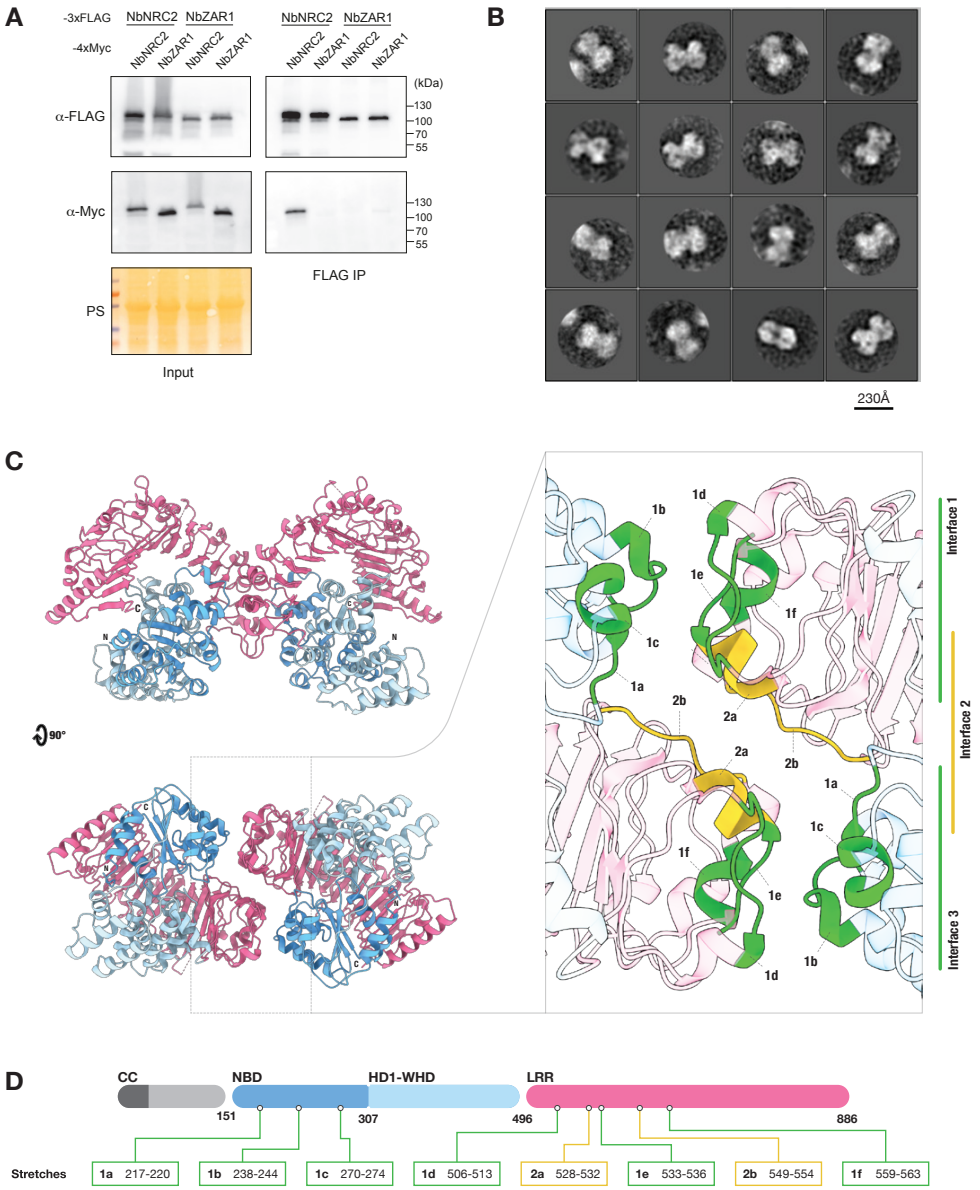


Figure 3.6: The helper NLR NRC2 forms a homodimer *in planta*.

(A) Co-immunoprecipitation assay testing the self-association of NRC2 in *N. benthamiana*. C-terminally 3×FLAG-tagged NbNRC2 or NbZAR1 was co-expressed with C-terminally 4×Myc-tagged NbNRC2 or NbZAR1; immunoprecipitation was performed with anti-FLAG beads (FLAG IP), and total extracts (Input) and IPs were immunoblotted with the antisera indicated on the left, with approximate molecular weights (kDa) on the right. NbNRC2 self-associates but does not associate with NbZAR1. Protein loading is shown by Ponceau stain (PS). (B) Representative cryo-EM 2D class averages of the resting-state NbNRC2 homodimer (scale bar, 230 Å). (C) Atomic model of the NRC2 homodimer in two orthogonal views (left), with the resolved NBD, HD1-WHD, and LRR domains; the N-terminal CC domain was absent from the cryo-EM density. The inset (right) details the three contact interfaces between the two protomers and the amino acid stretches (1a–1f, 2a, 2b) that form them. (D) Schematic of the NRC2 domain architecture and boundaries, with the residue ranges of the eight dimerization stretches indicated. This figure has been adapted from one published in Selvaraj et al. (Selvaraj et al., 2024) and appears here with permission.

3.4.1 The dimerization interfaces have diverged between NRC clades

When I compared the eight dimerization stretches identified in the NbNRC2 homodimer (Selvaraj et al., 2024) across all NRC sequences, they were noticeably less conserved than classic NLR motifs such as the P-loop and MHD (Figure 3.7A). This suggests that these interfaces diversified during NRC evolution. Within individual NRC clades, however, these regions were substantially more conserved (Figure 3.7B).

To quantify this, I calculated the relative amino acid ratio within each major clade (NRC2, NRC3, NRC4) compared with the others, using a 50-fold cutoff and requiring at least 80% of sequences within a clade to carry the identified variant (scripts provided by Yu Sugihara). With this analysis I identified clade-unique polymorphisms. For example, the NRC2 and NRC3 clades, which contain the representative proteins tested in the co-immunoprecipitation assays, showed 5 and 3 unique polymorphisms, respectively, across the 44 amino acids comprising the homodimer interface (Figure 3.7B).

3.4.2 Entropy analysis reveals clade-specific variation in NRC homodimerization stretches

To evaluate these patterns further, I used Shannon entropy analysis to quantify position-specific amino acid diversity. In this metric, a fully conserved position has an entropy score of 0, whereas a position with all 20 amino acids occurring at equal probability reaches a maximum entropy of 4.32 bits. I set a threshold of 1.5 bits to identify highly variable positions, following previous studies (Prigozhin and Krasileva, 2021).

All eight dimerization stretches scored above 1.5 bits when all 1,092 NRC sequences were analyzed together, but remained conserved within 11 of the 15 individual NRC clades (Figure 3.S1 and Figure 3.S2). In six clades (NRC0, NRCX, NRC2, NRC3, NRC4a, and NRC4t), all residues across the eight stretches scored below the 1.5-bit threshold, showing little to no sequence diversity. In five clades (NRC1, NRC6, NRC7, NRC4b, and NRC4other), the eight stretches remained below this threshold except for one to three variable residues per clade. In the remaining four clades (NRC8, NRC9, NRC5, and NRC4c), the stretches contained up to 12 variable residues with entropy values above 1.5 bits, reflecting higher sequence divergence. In contrast, the catalytic P-loop and MHD motifs remained below the 1.5-bit cutoff in 13 of the 15 clades, displaying minimal sequence variation (Figure 3.S1 and Figure 3.S2).

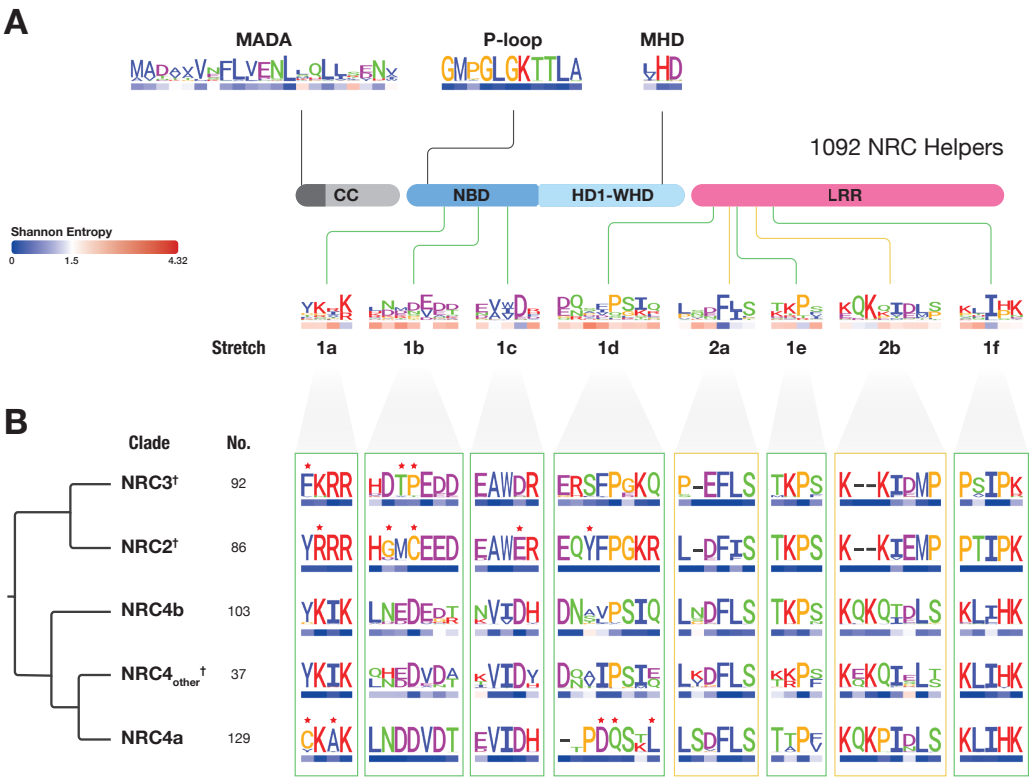


Figure 3.7: The NRC homodimerization interface has diversified throughout the evolution of the NRC clade.

(A) Schematic of NRC domain architecture (CC, NBD, HD1-WHD, and LRR), with the approximate positions of previously characterized NLR motifs (MADA, P-loop, MHD) and the eight amino acid stretches at the dimerization interface indicated. A database of 993 unique NRC sequences and 99 reference NRC sequences from 123 solanaceous genomes was used to generate consensus sequence patterns for each motif/region; Shannon entropy scores per position are indicated in bits, with a 1.5-bit threshold distinguishing highly variable (>1.5 bits) from non-variable (<1.5 bits) positions (Prigozhin and Krasileva, 2021). (B) All 1,092 NRC sequences were grouped by phylogeny and consensus sequence logos were generated within each clade for the eight dimerization stretches. Clades are ordered by phylogenetic relationship with the number of sequences per clade indicated. Red stars mark clade-unique amino acid polymorphisms; † marks clades containing the *N. benthamiana* NRCs used in the co-immunoprecipitation experiments. This figure has also been published in Selvaraj et al. (Selvaraj et al., 2024) and appears here with permission.

3.5 Conclusions and Discussion

In this chapter I examined two features of the NRC helper clade that are not easily explained by co-evolution with pathogen effectors: the diversification of the N-terminal MADA motif, and the divergence of the interfaces through which these receptors self-associate in their resting state. In both cases, the relevant selective pressures appear to arise from within the network itself, driven by the requirement that a receptor signals independently and correctly, rather than from direct recognition of pathogen molecules.

The MADA analysis began from the observation that while all NRC helpers carry the motif, not all of them execute cell death. Unbiased clustering of N-terminal α1 helix sequences recovered the phylogenetic structure of the clade and showed that the α1 helix has diversified in both length and sequence. This is notable for a region whose function is generally considered interchangeable between distantly related NLRs (Adachi et al., 2019a). Within this diversity, NRCX carries a threonine at the fourth position (T4) of the motif that is absent from its closest functional relatives, NRC2 and NRC3. Our working hypothesis was that this T4 residue represents a degenerative substitution that renders the MADA motif non-functional, consistent with the inability of NRCX to trigger cell death and with the loss of activity in NRCX–NRC4 N-terminal chimeras (Adachi et al., 2023a).

Screening for this signature across the wider clade revealed that it is not unique to NRCX. Around a quarter of NRC4/5 sequences also carry the T4 signature, and most of these form a distinct subclade, NRC4t, that is sister to NRC4a in much the same way that NRCX sits sister to NRC2 and NRC3. Given this parallel in both sequence and phylogenetic placement, I hypothesized that NRC4t might represent a second, independently derived group of modulators.

However, functional assays did not support this idea. Both potato NRC4t variants tested induced strong cell death as autoactive mutants, both in wild-type plants and in *nrc2/3/4* knockout line, behaving like canonical helper NLRs like NRC2 and NRC4. Moreover, I could detect the T4 pattern in other NRC4 and NRC5 proteins not belonging to NRC4t clade. Additionally, not all NRC4t clade sequences contained the T4 signature. The most direct interpretation is that the T4 substitution alone is not sufficient to abolish cell death activity, and that the functional degeneration of NRCX depends on additional features within its N-terminus or elsewhere in the protein. This is supported by lack of cell death with NRCX chimeras containing NRC4 α1 (Adachi et al., 2023a). Whether NRC4t proteins act as conventional helpers or have acquired distinct regulatory roles remains an open question. This will require identifying the upstream NRC-S that signal through them or helpers being regulated by them.

In the second half of the chapter, I addressed the homodimer resting state of NRC2. NRC2 accumulates *in planta* as a homodimer and does not form stable heterocomplexes with the paralogs NRC3 and NRC4, despite their shared evolutionary origin and overlapping roles in the network (Selvaraj et al., 2024). The phylogenomic data support that expectation. Across all 1,092 NRC sequences, the eight dimerization stretches were markedly less conserved than the P-loop and MHD motifs, suggesting that they diversified as the clade expanded. Within individual clades, however, these stretches were well conserved. Each of the major clades also carried unique polymorphisms at these positions. This pattern of between-clade divergence combined with within-clade conservation matches what is expected when selection acts to preserve self-association while preventing cross-association between paralogs, paralleling theoretical models for the emergence of orthogonal signaling pathways following gene duplication (Capra et al., 2012; McClune and Laub, 2020).

Together with the structural data, these findings suggest that the diversification of dimerization interfaces serves to insulate helper nodes from one another within the network. Such insulation maintains functional redundancy, minimizes undesired cross-activation between pathways, and provides a way to evade pathogen effectors that suppress individual helpers (Derevnina et al., 2021; Contreras et al., 2023b).

However, several caveats apply to these conclusions. The phylogenomic analyses draw exclusively on Solanaceae genomes, and the 15 clades differ substantially in size, meaning the entropy and polymorphism comparisons are not equally powered across all clades. Furthermore, co-immunoprecipitation captures stable complexes and would not necessarily detect weak or transient interactions between paralogs, which may still occur during activation and release. Most importantly, the link between interface divergence and activation insulation remains correlative. The sequence analyses show that these positions have diverged in the expected pattern, but do not directly prove that this divergence prevents heterodimerization. Although stretch swaps between NbNRC2a and NbNRC3 lead to loss of association between the NRC2 chimera and the WT variant (Selvaraj et al., 2024), testing will require swapping or mutating interface residues between paralogs to determine whether heterodimerization is acquired.

A further consideration links the two themes of this chapter. Tomato NRC2 has recently been shown to assemble beyond the homodimer, forming tetramers and higher-order oligomers at elevated concentrations in an inactive conformation. Dimerization and higher-order oligomerization appear to stabilize the resting state, sequestering the receptor from assembling into an active resistosome, while mutations at the dimeric or interdimeric interfaces enhance pathogen-induced cell death (Ma et al., 2024). However, we could not detect a tetramer for NbNRC2 resting state.

In the SlNRC2 tetramer, two homodimers associate through an interface distinct from the dimerization contacts analyzed above. This contact is formed by residues at the N-terminus of the CC domain of one protomer (including M14, R18, and D32) packing against the LRR domain of an adjacent protomer at positions including Y739 and T743 (Figure 3.8). Notably, the CC residues involved fall directly within or adjacent to the N-terminal α1 helix carrying the MADA motif. If this arrangement is functional *in vivo*, this region would serve two distinct roles: executing cell death upon activation and maintaining the resting receptor in an autoinhibited higher-order state prior to activation. Whether this tetramerization interface is general to the NRC clade or specific to NRC2 remains unresolved. Moreover, why NRCs dimerize or why NLRs can exist as monomers, dimers, and higher oligomers in their resting state remains unanswered here.

More broadly, these results illustrate that NLR evolution is shaped by internal network dynamics as well as by arms races with pathogens. The interfaces examined here reside within the N-terminal, NB-ARC, and LRR regions that mediate self-association and oligomerization, rather than the interfaces that bind effectors. The pressures acting on them stem from the need to maintain functional architecture within the network.

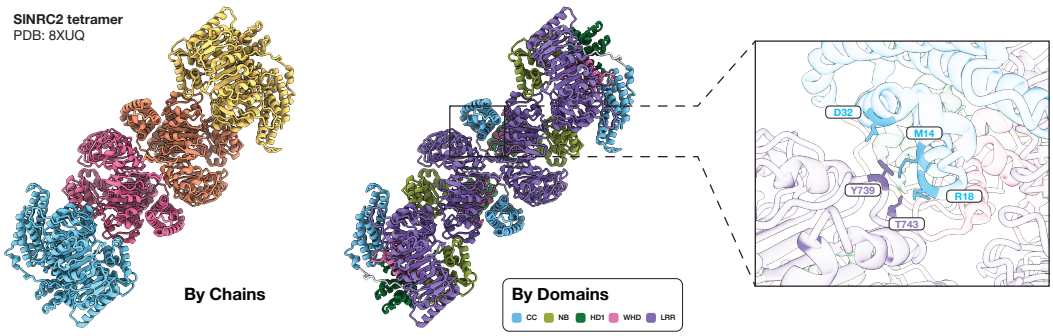


Figure 3.8: The SlNRC2 tetramer associates through an interface formed by the N-terminal α1 helix and the LRR domain.
Cryo-EM structure of the tomato NRC2 tetramer in its inactive conformation (PDB 8XUQ) (Ma et al., 2024), shown colored by protomer chain (left) and by domain (center; CC, NBD, HD1, WHD, and LRR as indicated). The tetramer comprises two homodimers related by the contact boxed in the central panel. The inset (right) details this interdimeric interface, in which residues at the N-terminus of the CC domain of one protomer (M14, R18, D32) pack against residues of the LRR domain of the adjacent protomer (Y739, T743). The CC residues involved lie within or immediately adjacent to the N-terminal α1 helix that carries the MADA motif.

3.6 Supplementary figures



Figure 3.S1: Sequence logos and Shannon entropy of the conserved motifs and dimerization stretches across all NRC clades.

Consensus sequence logos and per-position Shannon entropy for the MADA, P-loop, and MHD motifs and the eight dimerization stretches (1a–1f, 2a, 2b), shown for the NRC-helper consensus (1,092 sequences) and for each of the 15 NRC clades grouped by phylogeny. The number of sequences per clade is indicated, and the colored squares on the right summarize representation across the *Nicotiana*, *Physalis*, *Capsicum*, and *Solanum* genera. Per-position entropy is colored from conserved (blue) to variable (red), with a 1.5-bit reference. This figure has also been published in Selvaraj et al. (Selvaraj et al., 2024) and appears here with permission.

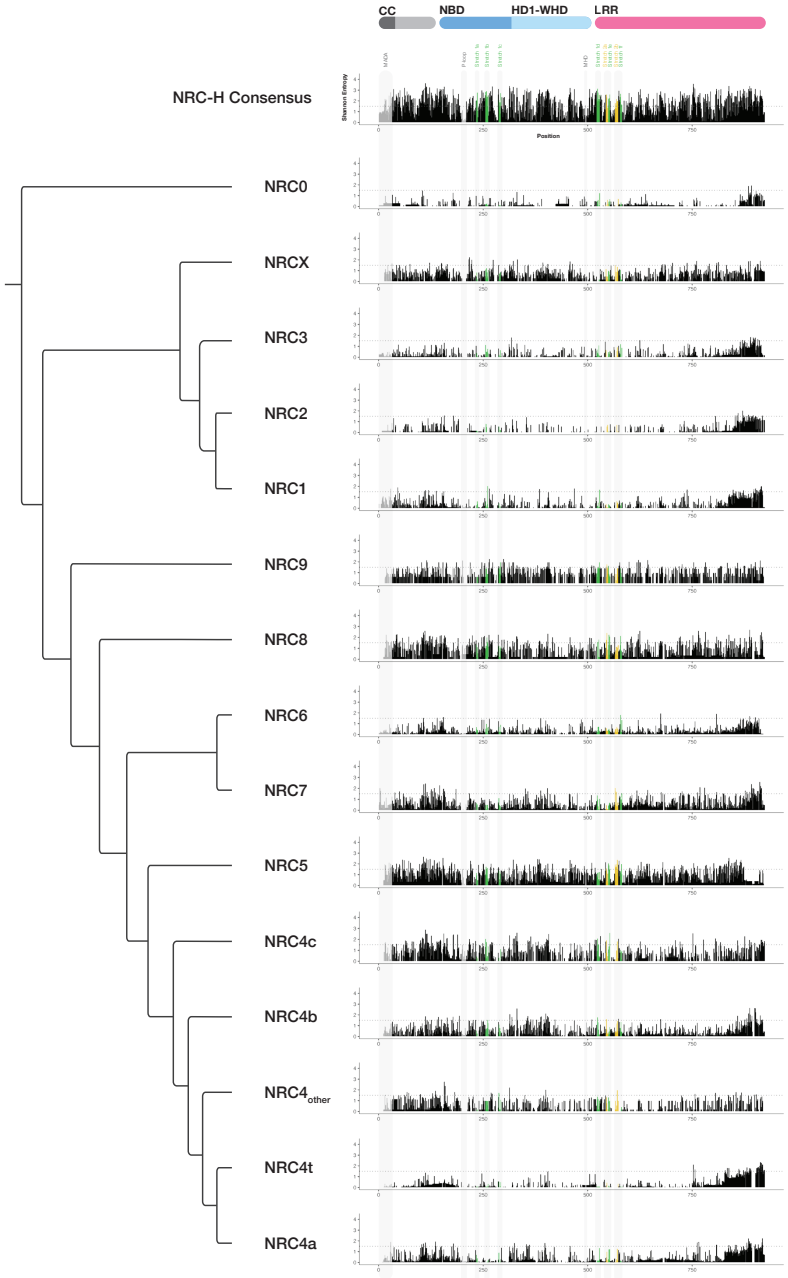


Figure 3.S2: Full-length Shannon entropy across NRC clades.
Per-position Shannon entropy calculated across the full-length alignment for the NRC-helper consensus (all 1,092 sequences) and within each of the 15 NRC clades, ordered by phylogeny. The domain architecture (CC, NBD, HD1-WHD, LRR) is shown at the top, with the positions of the MADA, P-loop, and MHD motifs and the eight dimerization stretches highlighted. The 1.5-bit threshold is indicated by the dotted line. This figure has also been published in Selvaraj et al. (Selvaraj et al., 2024) and appears here with permission.

3.7 Tables

Table 3.1: Constructs used in the NRC4t cell death assays.
Binary expression constructs used for agroinfiltration in *N. benthamiana* (Figure 3.5). Constructs 1–4 were assembled in the pICH47732 Golden Gate acceptor under a short 35S promoter with an omega translational enhancer and a C-terminal 3×FLAG tag. Constructs 5–12 were assembled in the pJK-001c level 2 acceptor under a double 35S promoter with a C-terminal 4×Myc tag and a 35S terminator. Superscripts denote the aspartate-to-valine substitution in the MHD motif that renders the receptor autoactive, and DV denotes the equivalent substitution where the residue position is not given. Plasmid parts are listed in the order acceptor backbone; promoter; epitope tag; terminator. Antibiotic resistances refer to selection of the *Agrobacterium* strain carrying each construct (Carb, carbenicillin; Kan, kanamycin; Rif, rifampicin; Gen, gentamicin).

No.	Construct	Antibiotic	*Agrobacterium* strain	Source	Plasmid parts
1	NbNRC4c	Carb/Rif/Gen	GV3101	Hiroaki Adachi	pICH47732; Short 35S+Ω; 3×FLAG; ocs
2	NbNRC4c^{DV}	Carb/Rif/Gen	GV3101	Hiroaki Adachi	pICH47732; Short 35S+Ω; 3×FLAG; ocs
3	NbNRCX	Carb/Rif/Gen	GV3101	Hiroaki Adachi	pICH47732; Short 35S+Ω; 3×FLAG; ocs
4	NbNRCX^{DV}	Carb/Rif/Gen	GV3101	Hiroaki Adachi	pICH47732; Short 35S+Ω; 3×FLAG; ocs
5	StNRC4e	Kan/Rif/Gen	GV3101	Jiorgos Kourelis	pJK-001c; 2×35S; 4×Myc; 35S
6	StNRC4e^{D479V}	Kan/Rif/Gen	GV3101	Jiorgos Kourelis	pJK-001c; 2×35S; 4×Myc; 35S
7	StNRC4h	Kan/Rif/Gen	GV3101	Jiorgos Kourelis	pJK-001c; 2×35S; 4×Myc; 35S
8	StNRC4h^{D478V}	Kan/Rif/Gen	GV3101	Jiorgos Kourelis	pJK-001c; 2×35S; 4×Myc; 35S
9	StNRC4p	Kan/Rif/Gen	GV3101	Jiorgos Kourelis	pJK-001c; 2×35S; 4×Myc; 35S
10	StNRC4p^{D478V}	Kan/Rif/Gen	GV3101	Jiorgos Kourelis	pJK-001c; 2×35S; 4×Myc; 35S
11	StNRCX	Kan/Rif/Gen	GV3101	Jiorgos Kourelis	pJK-001c; 2×35S; 4×Myc; 35S
12	StNRCX^{D478V}	Kan/Rif/Gen	GV3101	Jiorgos Kourelis	pJK-001c; 2×35S; 4×Myc; 35S

Chapter 4: Contrasting patterns of evolution in sensor and helper NLRs

The work described in this chapter has been published in Pai et al. (Pai et al., 2026) and appears here with permission. Cell death assays were performed by Hsuan Pai. I carried out the bioinformatic and evolutionary analyses and generated all the figures.

4.1 Introduction

Plants rely on a layered innate immune system to defend themselves against a broad range of pathogens (Jones and Dangl, 2006; Dodds et al., 2024). Intracellular NLR receptors form a central component of this system, recognizing pathogen-secreted effector proteins and initiating defense responses that often culminate in localized programmed cell death to restrict pathogen spread (Jones and Dangl, 2006). NLRs constitute one of the most diverse protein families in plants, with individual members evolving at strikingly different rates (Prigozhin and Krasileva, 2021; Michelmore and Meyers, 1998; Seong et al., 2020). What remains poorly resolved is how much of this variation in evolutionary tempo reflects the specific functional role a given receptor plays.

NLRs typically exhibit a tripartite architecture: an N-terminal signaling domain, most often a coiled-coil (CC) or Toll/interleukin-1 receptor (TIR) domain; a central NB-ARC domain that governs activation; and a C-terminal leucine-rich repeat (LRR) domain that mediates pathogen detection either directly or indirectly (Kourelis et al., 2021). Based on NB-ARC phylogeny, they fall into four main classes: CC-NLRs, CC_{G10}-NLRs, RPW8-type CC_R-NLRs, and TIR-NLRs (Kourelis et al., 2021). This architecture is not invariant; a subset of receptors carry additional integrated domains that act as decoys or baits for effector recognition (Kroj et al., 2016; Sarris et al., 2016). Approximately 20% of CC-NLRs also carry a conserved N-terminal MADA motif required for cell death execution (Adachi et al., 2019a). Upon activation, CC-NLRs assemble into higher-order complexes termed resistosomes, which insert into the plasma membrane to act as calcium-permeable channels via their N-terminal helices (Wang et al., 2019a; Bi et al., 2021). Recent computational advances now allow such assemblies to be modeled in simulated membrane environments with high confidence (Selvaraj et al., 2024; Toghani et al., 2025a).

Many NLRs do not act autonomously. Some function as genetically linked pairs, such as

the rice CC-NLR pairs Pik-1/Pik-2 and RGA5/RGA4, or the *Arabidopsis thaliana* TIR-NLR pair RPS4/RRS1 (Kroj et al., 2016; Sarris et al., 2016). Others are organized into wider networks where sensors depend on helpers in one-to-many or many-to-one configurations (Adachi et al., 2019b). The CC_R-NLRs NRG1 and ADR1 act downstream of TIR-NLR signaling in this manner (Castel et al., 2019; Saile et al., 2020).

The NRC network, first described in solanaceous species such as tomato and *Nicotiana benthamiana*, is the best-characterized CC-NLR network. Sensor NLRs (NRC-S) detect effectors but cannot signal independently, depending instead on helper NLRs (NRC-H) to transduce signals and execute cell death (Wu et al., 2017; Adachi et al., 2019b). NRC-S fall into two structural subtypes: Rx-type NRC-S with canonical domain architecture, and SD-type NRC-S carrying an additional N-terminal Solanaceous domain (Seong et al., 2020). Upon effector detection, sensors trigger the transition of their helpers from a resting homodimer into hexameric resistosomes (Contreras et al., 2022; Ahn et al., 2023; Selvaraj et al., 2024). The mechanism transmitting this signal between sensor and helper is consistent with a transient activation-and-release interaction rather than a stable, persistent complex (Contreras et al., 2022; Ahn et al., 2023; Contreras et al., 2024).

The network is thought to have emerged roughly 100 million years ago in the superasterid lineage and has expanded to varying degrees across asterid families, reaching its greatest complexity in the lamiids (Wu et al., 2017; Goh et al., 2024; Sakai et al., 2024). Within the network, NRC0 stands out as the most ancient and conserved helper and is frequently linked genetically to the NRC-S that depend on it. In contrast, derived helpers have undergone lineage-specific expansions and occupy varied genomic positions, with some clustered alongside NRC-S, such as NRC6, and others dispersed (Goh et al., 2024; Sakai et al., 2024; Lee et al., 2021).

The broad generalization in the field is that NRC-S evolve rapidly while helpers remain conserved (Goh et al., 2024). However, as shown in Chapter 3, paralogous helpers are themselves subject to diversifying selection that preserves homodimerization while preventing cross-activation between paralogs (Selvaraj et al., 2024). This variation in evolutionary tempo is not arbitrary. NLR repertoires range from a handful to over a thousand genes per genome and follow a birth-and-death model, where duplication continually generates new receptors, some of which are retained for recognition while others are lost or pseudogenized (Michelmore and Meyers, 1998; Seong et al., 2020). On the other hand, pathogens have evolved to target plant immune receptors. The Solanaceae helper NLRs NRC2 and NRC3 are suppressed by the cyst nematode effector SS15 and the oomycete effector AVRcap1b (Derevnina et al., 2021). In return, in some species NRC3 has accumulated mutations in the effector-binding loop that allows evading suppres-

sion (Sugihara et al., 2025). Redundancy within the network ensures that unsuppressed nodes maintain signaling when others are blocked (Derevnina et al., 2021). Under this model helpers face multiple evolutionary constraints. They need to retain compatibility with multiple upstream NRC-S while evading pathogen effectors that target them.

Most of our knowledge about the network derives from the Solanaceae family, leaving open the question of whether its architecture and evolutionary dynamics are general asterid features or specific to that family. The Asterales, a campanulid order that includes lettuce (*Lactuca sativa*), offer a useful comparison. Lettuce is a classic model for resistance genetics. Its NLR repertoire was organized by Meyers and colleagues into 42 phylogenetic Resistance Gene Candidate (RGC) groups, later consolidated into eight Major Resistance Clusters (MRCs) distributed across six chromosomes (Meyers et al., 1998; McHale et al., 2009; Christopoulou et al., 2015). Early work on the lettuce RGC2 locus distinguished two NLR classes: a diverse type I and a conserved type II, with contrasting copy numbers and evolutionary trajectories (Kuang et al., 2004). This distinction has since been recognized in other systems, including *Arabidopsis thaliana* and maize (Prigozhin and Krasileva, 2021). Despite this rich background in resistance genetics, the NRC network in the Asterales has received little attention.

In this chapter, I address three primary questions: how extensively the NRC network has expanded in the Asterales relative to the Solanales, and how its members are arranged in the genome; whether a functional sensor–helper network operates in lettuce; and whether helpers and distinct sensor subgroups exhibit different signatures of selection within the genus *Lactuca*. Building on the harmonized annotation framework developed in Chapter 2, I show that the network is comparatively less expanded in the Asterales. Nevertheless, lettuce maintains a functional, partially redundant network supported by two helpers. Furthermore, NRC-S dependent on the ancient helper NRC0, which reflects the ancestral state of the NRC network, evolve under markedly stronger constraint than those that have shifted onto more derived helpers.

4.2 The NRC network has expanded less in Asterales relative to Solanales

Because the NRC network is distributed across the asterids, I first compared its gene content between the Solanales (a lamiid order) and the Asterales (a campanulid order) (Figure 4.1A) (Wu et al., 2017). Applying NLRtracker (Kourelis et al., 2021) to 40 Solanales and 29 Asterales genomes that I annotated with Helixer (Stiehler et al., 2021; Holst et al., 2023), yielded 21,232 non-redundant NLR sequences. Phylogenetic analysis

of these sequences alongside RefPlantNLR references led to identification of the NRC network as a group of 7,020 sequences (Figure 4.S1) (Kourelis et al., 2021).

Total NLR content was broadly comparable between the two orders, averaging 300 sequences per genome in the Asterales and 338.6 in the Solanales. The proportion accounted for by the NRC network, however, differed: NRC-type receptors represented on average only 6.6% of the Asterales NLRome, compared with 50.3% in the Solanales (Figure 4.S2). While both orders carry comparable numbers of NLRs overall, the Asterales have expanded other NLR classes instead of the NRC network.

To resolve this pattern, I analyzed the 592 Asterales NRC network sequences, which comprise 88 helpers and 504 NRC-S (Figure 4.1B). The helpers split into two phylogroups: 27 sequences grouped with NRC0, while the remaining 61 formed a distinct clade alongside known Solanales helpers (such as NRC2 and NRC4), hereafter referred to as Ast-NRCs (Wu et al., 2017). All 504 Asterales NRC-S clustered in a single well-supported clade sister to Solanales SD-type NRC-S. None grouped with the Rx-type NRC-S prominent in the Solanales. Although these Asterales NRC-S are phylogenetically SD-type, they lack the N-terminal Solanaceous domain itself.

I observed the same patterns in lettuce. Its 17 NRC sequences (two helpers and 15 NRC-S) represent only 5% of its NLRome consisting of 359 NLRs. Instead, the repertoire is dominated by TIR-NLRs (206 sequences, over 57%). The CC-NLR, CC_R-NLR, and CC_{G10}-NLR classes each have 97, 9, and 42 sequences, respectively (Figure 4.S3). NLR expansion in lettuce has therefore been highly uneven, with the NRC network remaining among the least expanded.

4.3 Asterales NRC networks show lineage-specific expansion and variable genomic clustering

Given the differences between the two orders, I next examined the copy number and genomic organization of the network within the Asterales in more detail, focusing on *Lactuca*. I analyzed 40 helper and sensor sequences drawn from common lettuce and two wild relatives, *Lactuca saligna* and *Lactuca virosa*. Phylogenetic analysis resolved these into four well-supported groups: a single helper group containing sequences phylogenetically clustering either to NRC0 or to tomato NRC2 (termed Ast-NRC1 in each species), and three sensor groups, hereafter referred to as groups 1, 2, and 3 (Figure 4.2A). Mapping these onto the established lettuce NLR classification showed that they correspond to four previously defined RGC groups: the helpers to RGC7, and sensor groups 1, 2, and 3 to RGC26, RGC27, and RGC9, respectively (Figure 4.S3)

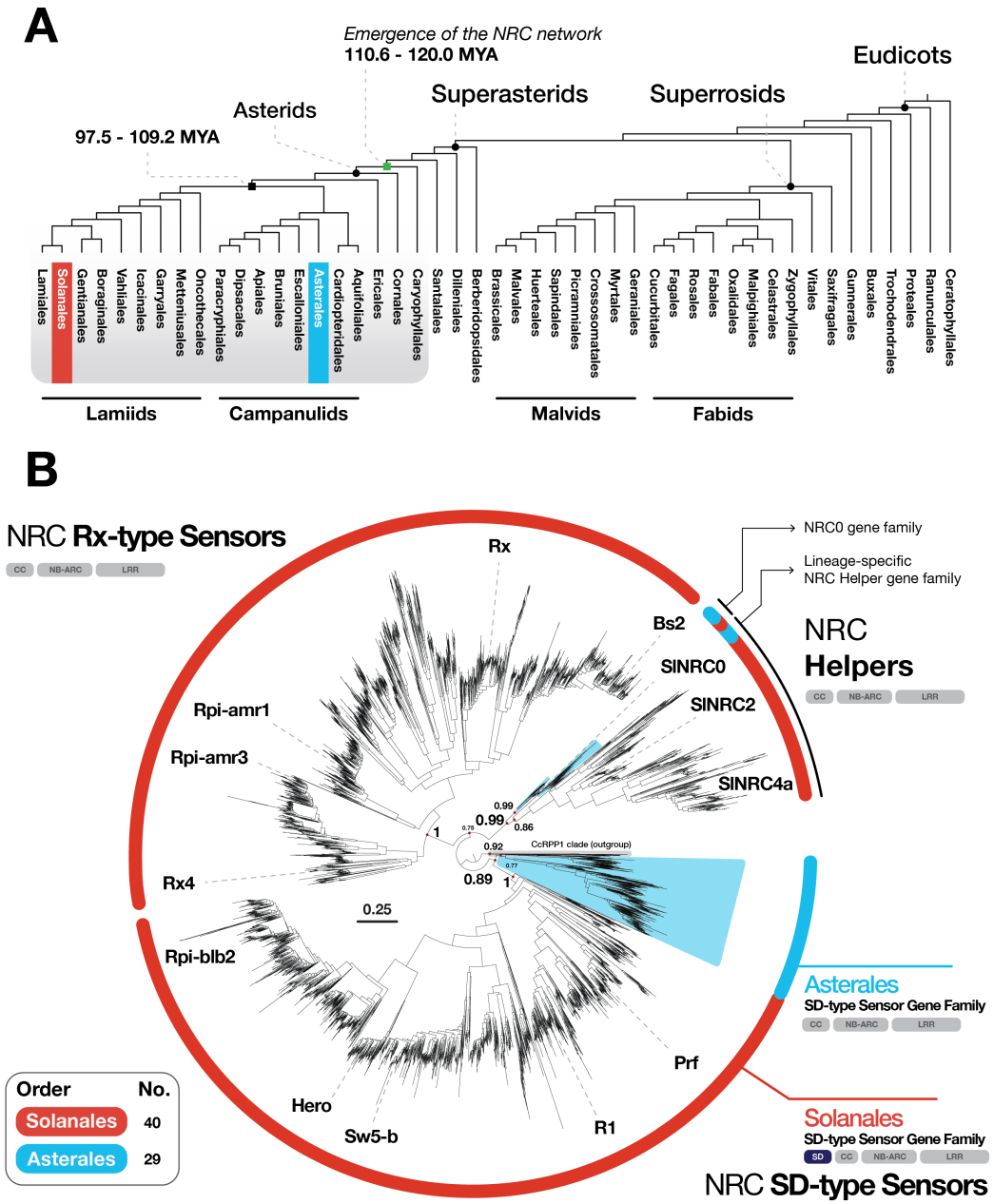


Figure 4.1: The NRC network has expanded less in Asterales than in Solanales.
(A) Taxonomic tree of eudicots indicating the timing of NRC network emergence within the asterids (Wu et al., 2017; Goh et al., 2024); relationships and divergence times were taken from TimeTree 5 (Kumar et al., 2022). MYA, million years ago. (B) Phylogenetic tree of NB-ARC domains from NRC network sequences derived from 40 Solanales and 29 Asterales genomes, built with FastTree 2 (Price et al., 2010) and rooted on CcRPP1. Asterales branches are highlighted and node values indicate bootstrap support. Helpers (NRC-H) and sensors (NRC-S) form distinct, well-supported groups; helpers divide into NRC0 and a lineage-specific group expanded in both orders, while NRC-S divide into SD- and Rx-type groups, the latter absent from the Asterales. SD, Solanaceous domain. This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

(Meyers et al., 1998; McHale et al., 2009).

Because NRC-H and NRC-S often occur in genetic clusters, as reported for NRC0 and for Solanaceae NRC6 (Sakai et al., 2024; Lee et al., 2021), I mapped these genes onto the lettuce chromosomes. Contrary to the pattern observed in other asterids, LsatNRC0 on chromosome 3 is isolated from any sensor sequence (Figure 4.2B) (Sakai et al., 2024). Ast-LsatNRC1, by contrast, resides within a physical cluster of less than 100 kb on chromosome 8 alongside NRC-S from groups 2 and 3, with all group 3 members localized to this locus. This cluster falls within the previously described MRC8c locus (Christopoulou et al., 2015; McHale et al., 2009). The remaining NRC-S, including all of group 1, are dispersed across chromosomes 4, 8, and 9 within MRC4, MRC8a, and MRC9a. LsatNRC0 is the only NRC gene outside any reported MRC.

The lack of chromosome-level assemblies for the wild lettuce species precluded a complete synteny analysis, although Ast-LvirNRC1 was located adjacent to group 2 and 3 NRC-S on a single contig (Figure 4.S4). Given these observations, I extended this analysis to five additional Asterales species, classified by their phylogenetic relationship to *Lactuca* helpers and NRC-S (Figure 4.S5), revealed no uniform genomic arrangement. In *Cichorium intybus*, CiNRC0 lies within 500 kb of a group 1 sensor, while both Ast-CiNRC1 copies cluster with group 2 and 3 NRC-S. In *Cynara cardunculus*, CcNRC0 is close to a group 1 sensor and Ast-CcNRC1 clusters with four group 1 NRC-S. In contrast, in *Helianthus annuus*, *Chrysanthemum lavandulifolium*, and *Codonopsis lanceolata*, helpers and NRC-S show no genomic clustering (Figures 4.S6 and 4.S7). Physical linkage within the network is therefore variable both within *Lactuca* and across the wider Asterales order.

4.4 LsatNRC0 and Ast-LsatNRC1 signal for distinct sensor groups

To test whether the phylogenetic grouping of these sequences into NRC-H and NRC-S reflects genuine functional dependency, we used *Agrobacterium*-mediated transient co-expression in an *nrc2/3/4* CRISPR knockout line of *Nicotiana benthamiana* (Wu et al., 2017). Because the cognate effectors of these lettuce NRC-S are unknown, we generated autoactive derivatives carrying a D-to-V substitution in the MHD motif (Adachi et al., 2019a). Expressed alone, none of the NRC-S produced a visible cell death response, whereas autoactive variants of both helpers triggered strong cell death in *Nicotiana benthamiana* leaves (Figure 4.S8).

NRC-S and NRC-H were then co-expressed pairwise. Of the 16 NRC-S tested, nine induced visible cell death when combined with one or both wild-type helpers, and these pheno-

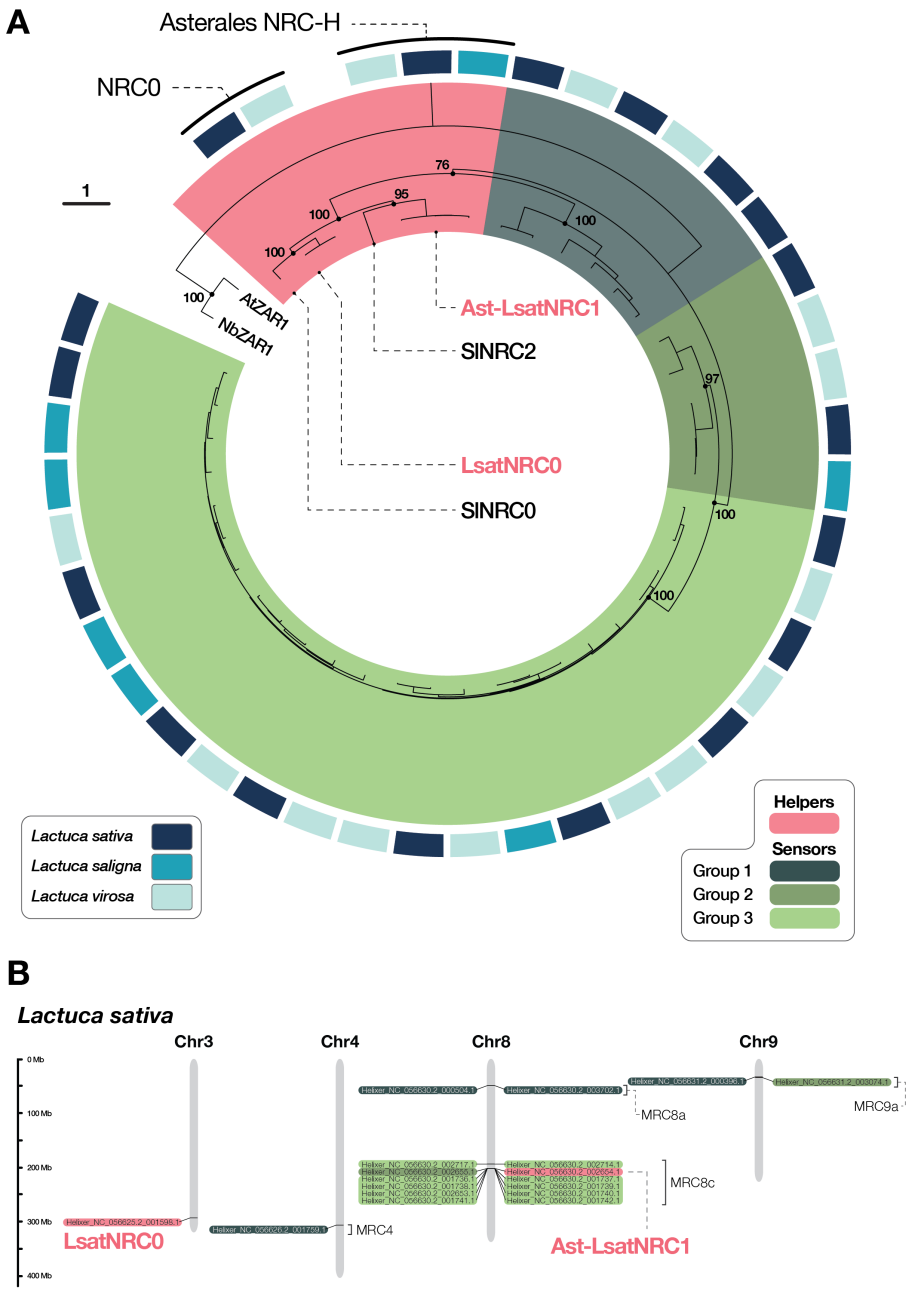


Figure 4.2: Unlike LsatNRC0, Ast-LsatNRC1 is genetically linked to NRC-S in lettuce.
(A) Phylogenetic tree of NB-ARC domains from 40 NRC sequences of three *Lactuca* species (*L. sativa*, *L. saligna*, *L. virosa*), constructed with IQ-TREE 2 (Minh et al., 2020). The *Lactuca* NRC family divides into one helper and three well-supported sensor groups; node values indicate bootstrap support. AtZAR1 and NbZAR1 were used as outgroups, and SlNRC0 and SlNRC2 as references. (B) Physical map of *L. sativa* NRC sequences. Ast-LsatNRC1 is genetically linked to group 3 NRC-S, whereas LsatNRC0 on chromosome 3 is not linked to any sensor. Previously described Major Resistance Clusters (MRCs) are indicated (McHale et al., 2009; Christopoulou et al., 2015). This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

types were reproducible across replicates (Figure 4.S8). These functional dependencies mapped cleanly onto the phylogeny: three NRC-S signaled exclusively through LsatNRC0 and all belonged to group 1; four signaled exclusively through Ast-LsatNRC1 and all belonged to group 3; and two NRC-S from group 2 could induce cell death with either helper (Figure 4.3A).

I also screened these sequences for the MADA motif required for helper-mediated cell death (Adachi et al., 2019a). Both helpers encoded clear MADA motifs, whereas none of the NRC-S did, matching the observation that only the helpers can execute cell death independently.

Functional dependencies in the wild lettuce species and across the wider Asterales remain to be validated experimentally. However, given their phylogenetic relationships and the fact that both NRC0 and group 1 NRC-S are absent from several of these genomes, functional specificity likely follows the same group-wise pattern. Together, these results define a functional NRC network in lettuce in which two helpers work downstream of nine NRC-S with partial redundancy (Figure 4.3B).

4.5 NRC-S dependent on Ast-LsatNRC1 are under stronger positive selection

To determine whether these functional differences are reflected in distinct selective regimes, I tested for positive selection across the *Lactuca* NRC groups using CodeML in PAML suite (Yang, 1997). These analyses compare a model permitting a subset of sites to evolve under positive selection ($\omega = d\text{N}/d\text{S} > 1$) against a null model with neutral and purifying selection only ($\omega \leq 1$). Where the selection model fits significantly better, individual sites under positive selection are identified by Bayes Empirical Bayes (BEB) analysis (Yang, 2005). For each group, I compared model M8 against its null counterpart M8a using a likelihood ratio test.

For the helpers, the test was not significant at the 1% threshold ($p = 0.037$), and BEB identified no positively selected sites. Sensor group 2 yielded a statistically significant likelihood ratio test ($p = 0.003$) but likewise lacked any BEB-supported sites. A significant test in the absence of identified sites is most conservatively interpreted as a false positive, as the M8a/M8 comparison is sensitive to alignment noise and gene conversion.

Sensor groups 1 and 3, by contrast, showed clear signatures of positive selection. In group 1, BEB identified five sites with posterior probabilities exceeding 0.95, three of which mapped to the CC domain on an AlphaFold 3 structural model (Figure 4.4B).

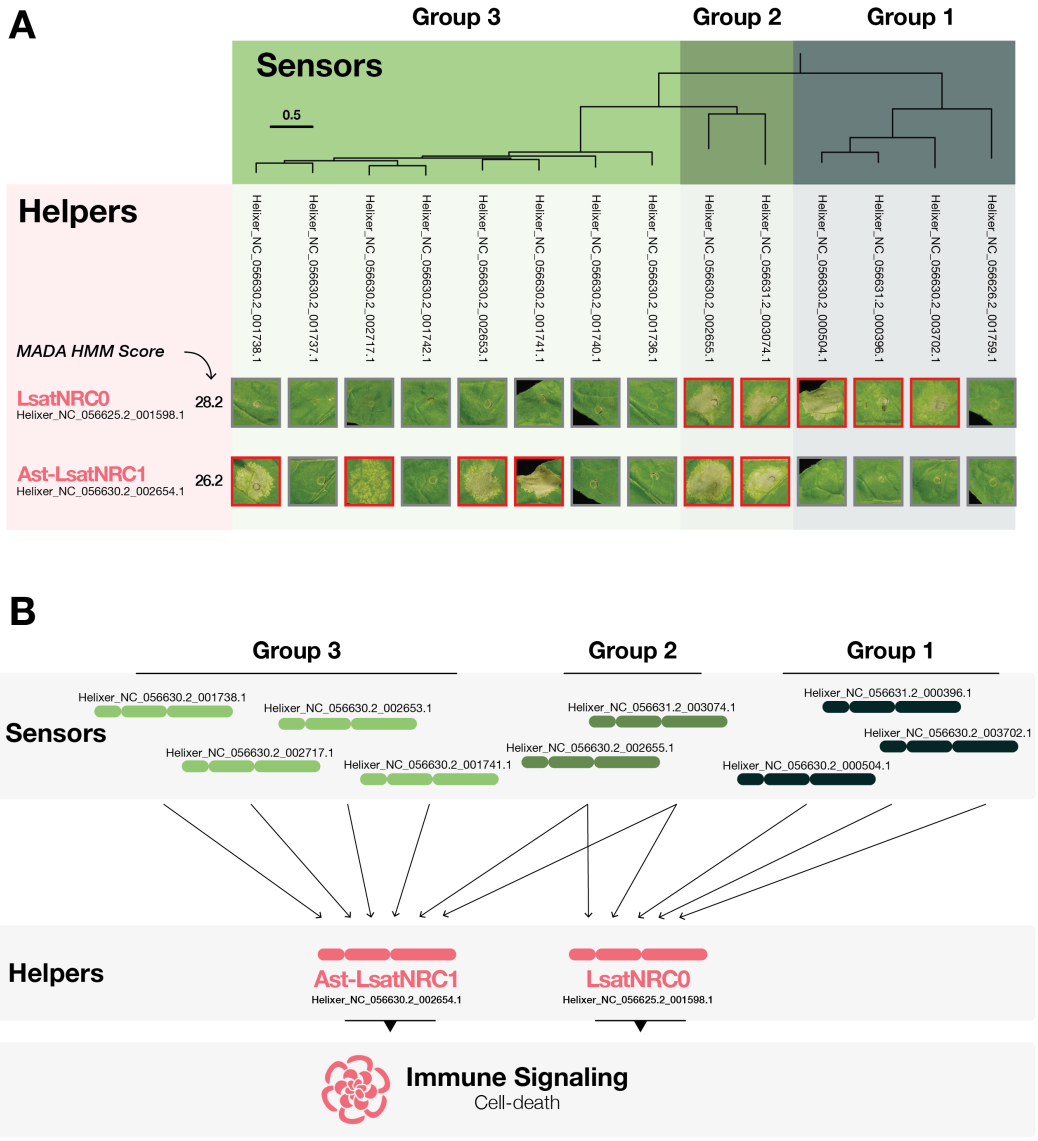


Figure 4.3: LsatNRC0 and Ast-LsatNRC1 mediate the cell death signaling of distinct NRC-S phylogenetic groups.

(A) Co-agroinfiltration cell death assays of NRC-H and NRC-S. NRC-S were expressed as autoactive (MHD D-to-V) mutants together with the two wild-type helpers; combinations producing cell death are outlined. Group 1 and group 3 NRC-S signal only through LsatNRC0 and Ast-LsatNRC1, respectively, while group 2 NRC-S signal through both. Protein accumulation was not assessed for combinations giving no visible cell death. (B) Model of the lettuce NRC network, in which nine validated NRC-S signal through the two helpers in a partially redundant manner. This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

Group 3 displayed the strongest signal, with 48 sites above 0.95 and 33 exceeding 0.99 posterior probability. The majority of these sites mapped to the concave face of the LRR domain, the surface typically responsible for direct effector recognition (Figure 4.4C) (Prigozhin and Krasileva, 2021).

To test whether average selective pressure varies across groups, I fitted a branch model assigning an independent ω to each phylogroup and compared it to a uniform one-ratio model. The four-ratio model fit the data significantly better ($2 \cdot \Delta \ln L = 305.66$, df $= 3$, $p = 6.5 \times 10^{-65}$), yielding $\omega = 0.12$ for the helpers, 0.35 for group 1, 0.39 for group 2, and 0.58 for group 3. The helpers showed the lowest average ω and group 3 the highest, matching the site-level analyses (Figure 4.S9). Although the four groups differ in sequence count and tree depth, their total tree dS lengths are comparable (1.51–2.87), as the higher sequence number in group 3 is offset by lower per-branch divergence. Branch-level heterogeneity under the free-ratio model is shown for reference in Figure 4.S10.

Finally, I calculated pairwise dN, dS, and ω values within each group using the Nei–Gojobori method, both across full-length sequences and separately for the CC, NB-ARC, and LRR domains (Figure 4.S11) (Nei and Gojobori, 1986). The helper group consistently had the lowest average ω. Sensor group 3 showed the highest overall ratios, while the CC domain of group 1 NRC-S yielded pairwise $\omega > 1$, indicating domain-specific positive selection rates (Figure 4.S12). Together, these analyses indicate that NRC-S signaling through Ast-LsatNRC1 evolve under stronger positive selection than those dependent on LsatNRC0, while the helpers themselves remain under strong purifying constraint.

4.6 Conclusions and discussion

Using the harmonized annotations described in Chapter 2, in this chapter I compared the NRC network between two asterid orders and used lettuce as a model to study the evolution of NRC network. Although the Asterales and Solanales carry comparable numbers of NLRs per genome, the NRC network accounts for only 6.6% of the Asterales repertoire compared with 50.3% in the Solanales. The Asterales have expanded other classes instead, most conspicuously the TIR-NLRs, which alone make up more than half of the lettuce NLRome. This fits the wider observation that the network proliferated heavily in lamiids relative to campanulids (Goh et al., 2024).

I observed the absence of Rx-type NRC-S from the Asterales. This group expanded extensively in the Solanales and contains many of the characterized resistance genes in solanaceous crops (Wu et al., 2017). The Asterales NRC-S cluster phylogenetically

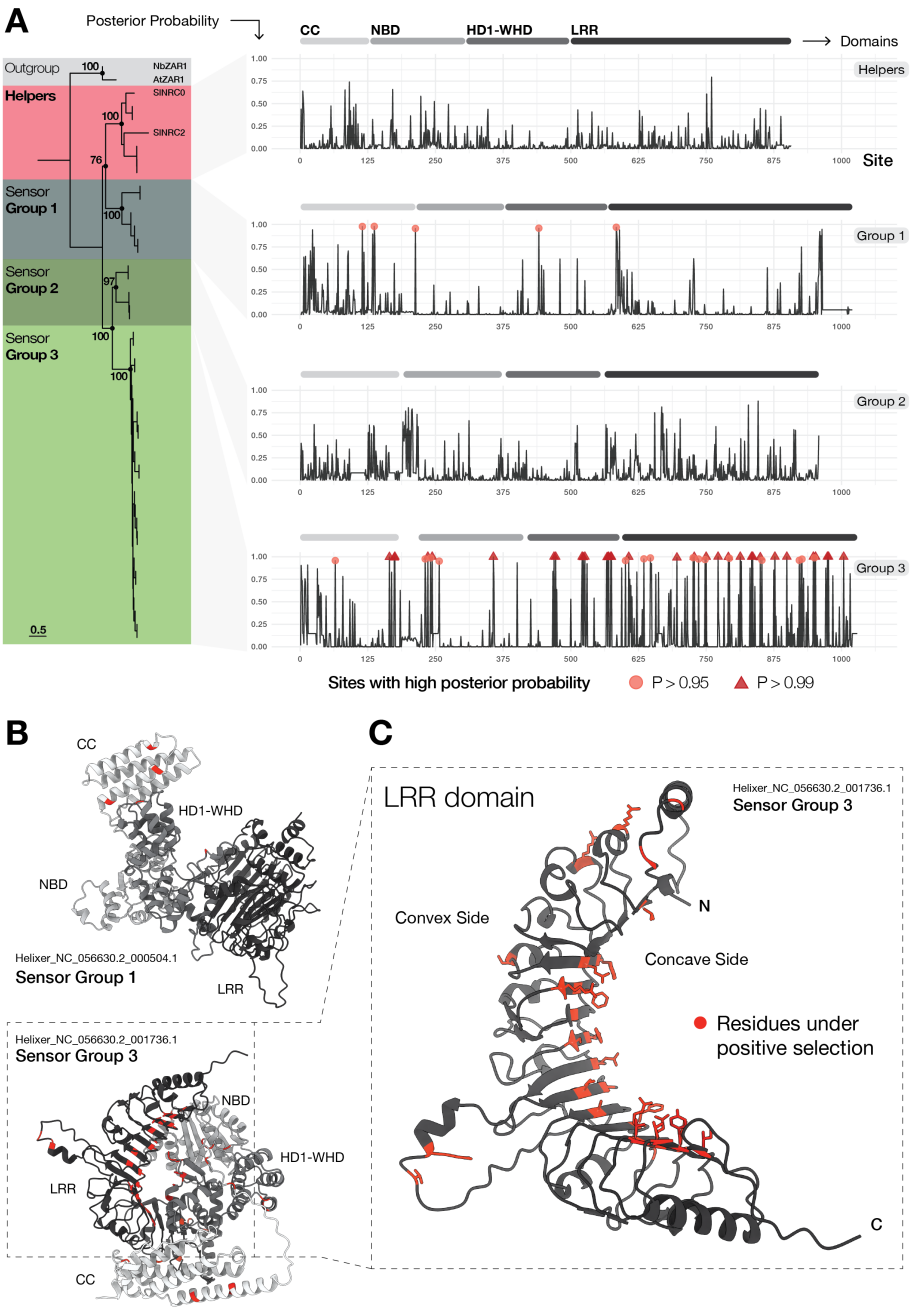


Figure 4.4: Ast-LsatNRC1-dependent NRC-S are under stronger positive selection than NRC0-dependent NRC-S and NRC helpers.

(A) Posterior probability that each site belongs to the positively selected class ($\omega > 1$), estimated by Bayes Empirical Bayes under model M8 in CodeML (Yang, 1997; Yang, 2005). Sites with $P > 0.95$ and $P > 0.99$ are marked by circles and triangles, respectively. No positively selected sites were detected in helpers or group 2 NRC-S, whereas groups 1 and 3 contain multiple selected sites, with most group 3 sites mapping to the LRR domain. (B) Positively selected sites (red) mapped onto AlphaFold 3 models of representative NRC-S from groups 1 and 3. (C) Close-up of the LRR domain of a representative group 3 sensor, showing positively selected sites concentrated on the concave face. CC, coiled-coil; NBD, nucleotide-binding domain; HD1, helical domain 1; WHD, winged-helix domain; LRR, leucine-rich repeat. This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

with Solanales SD-type NRC-S yet lack the Solanaceous domain itself. It has been proposed that this pattern reflects the secondary loss of SD-containing NRC-S in the Asterales (Seong et al., 2020). Consistent with this, the presence of SD-like NRC-S in more distantly related superasterids, such as sugar beet, argues that NRC-S with and without this integration were already present in their common ancestor, with the SD-carrying forms subsequently lost in ancestral Asterales. Resolving this will require denser phylogenetic sampling across the superasterids to determine when and how the SD integration originally arose.

Within lettuce, the network resolves into one helper and three sensor groups that map onto four previously described RGC groups (Meyers et al., 1998). Ast-LsatNRC1, together with all group 3 NRC-S and one group 2 sensor, lies within MRC8c, a cluster associated with quantitative trait loci for resistance to *Fusarium oxysporum* and *Verticillium dahliae* (Christopoulou et al., 2015). NRC-type receptors account for 12 of the 26 immune receptor genes at that locus. Although they have not yet been directly implicated in lettuce disease resistance.

The remaining NRC-S fall within MRC4, MRC8a, and MRC9a, of which MRC4 and MRC9a have been linked to downy mildew resistance (Christopoulou et al., 2015; McHale et al., 2009). LsatNRC0 sits outside any reported MRC and, unlike its homologs in several other asterids, is unlinked to any sensor (Sakai et al., 2024). Together with the variable genomic clustering observed across the other Asterales species surveyed here, this suggests that the ancient linkage of NRC0 to its NRC-S was disrupted by lineage-specific genomic rearrangements rather than lost systematically.

The functional assays extend earlier phylogenetic surveys of NRC network architecture (Goh et al., 2024) by directly testing which NRC-S signal through each helper. They reveal a graded shift in helper dependency across the sensor phylogeny: exclusive reliance on ancestral NRC0 in group 1, dual compatibility in group 2, and exclusive dependence on the derived Ast-LsatNRC1 in group 3. This makes lettuce a useful experimental system, as it allows helper dependency and evolutionary rate to be compared within a single, compact network rather than across distantly related species.

The selection analyses show that these functional categories track evolutionary rate: helpers are under the strongest purifying selection, group 1 NRC-S show limited diversification, and group 3 NRC-S display strong signatures of positive selection concentrated on the concave LRR surface implicated in effector binding.

The low sequence variability of the helpers matches observations from other helper systems, such as ADR1, which likewise belongs to the low-variability NLR class (Prigozhin and Krasileva, 2021; Saile et al., 2020). Group 1 NRC-S are unusual in having

their variability concentrated within the CC domain and the N-terminal portion of the LRR rather than the typical recognition interfaces, consistent with previous observations on NRC0-dependent NRC-S (Sakai et al., 2024); whether this pattern reflects relaxed selection or an active co-evolutionary process remains an open question. Group 2 NRC-S are comparatively invariant and represent the intermediate group capable of signaling through either helper. In contrast, group 3 NRC-S resemble highly polymorphic receptors such as *Arabidopsis thaliana* RPP1, displaying multiple positively selected sites clustered on the concave face of the LRR (Prigozhin and Krasileva, 2021).

A plausible model is that evolutionary constraint is shaped by signaling within the network (Figure 4.5). NRC-S that depend on the conserved NRC0 helper may be constrained both by the requirement to maintain helper compatibility and by the nature of their recognition targets, particularly if they monitor conserved effectors or guard essential host targets in a manner similar to the ZAR1–ZRK system. NRC-S that shifted onto a more recently derived helper appear relieved of these constraints, and their physical clustering indicates that recent tandem duplication accelerated their diversification. This pattern is visible in both orders, but is amplified in the Solanales, where a larger suite of derived, lineage-specific helpers supports a greatly expanded set of rapidly diversifying NRC-S. These interpretations should be treated with appropriate caution, as recent duplications and uneven group sizes introduce phylogenetic non-independence and sampling biases that site- and branch-model tests cannot completely eliminate.

The functional data also speak to the biochemical distinction between the two classes. None of the lettuce NRC-S triggered cell death when expressed alone, whereas both helpers did, and multiple lines of biochemical evidence indicate that NRC-dependent NRC-S likewise fail to form higher-order complexes either before or after activation (Contreras et al., 2022; Ahn et al., 2023; Contreras et al., 2024; Selvaraj et al., 2024). Taken together this is consistent with the view that these NRC-S have lost the capacity to assemble resistosomes altogether. Structural modeling supports the same conclusion and is developed in Chapter 6, where the lettuce network is placed alongside monocot NLR pairs to ask whether sensor and helper identity can be predicted structurally.

Why, then, should members of a single network evolve at such different rates? The most economical explanation is that evolutionary trajectory follows from functional position, and specifically from the dependencies each receptor must maintain (Figure 4.5). NRC0 and its dependent NRC-S are conserved across most asterid lineages (Goh et al., 2024; Sakai et al., 2024). Their low copy number and the conservation of their LRR domains, the region that normally mediates recognition, may indicate that they detect conserved effector molecules or guard essential host components in the manner of the ZAR1–ZRK system (Adachi et al., 2023b), either of which would limit how far the receptor can drift.

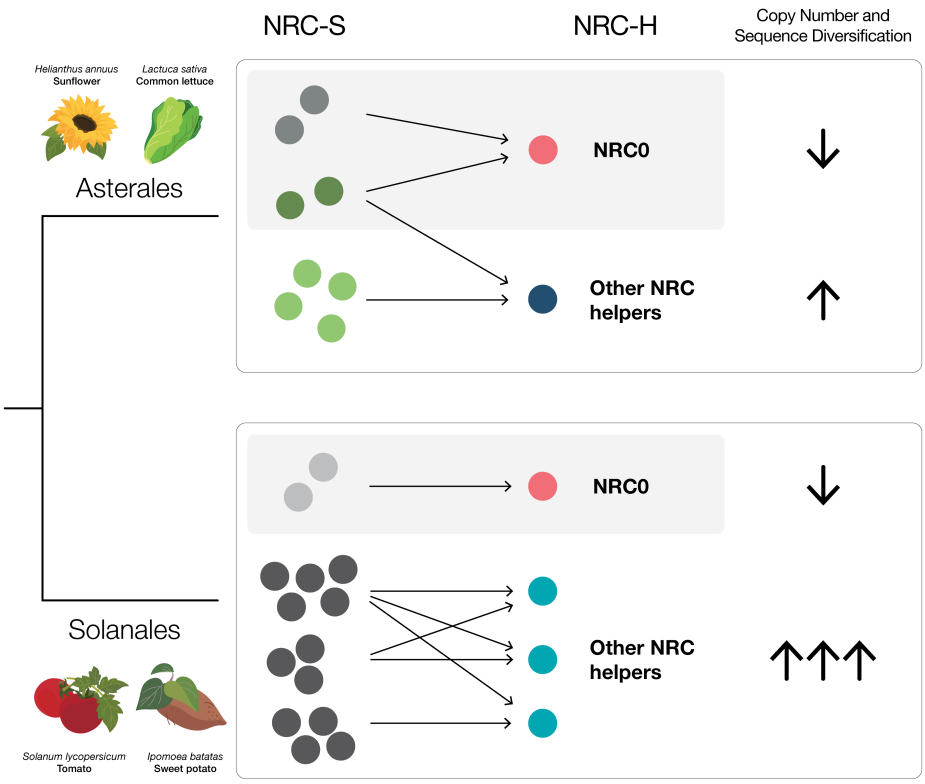


Figure 4.5: Evolutionary constraints and diversification patterns in NRC-H and NRC-S. In the proposed model, NRC-S dependent on NRC0 are present at lower copy number and are under weaker positive selection, likely reflecting the evolutionary constraints imposed by their dependence on a deeply conserved helper. NRC-S dependent on the other, more recently derived helpers have by contrast undergone greater expansion and diversification. The same asymmetry is seen in both orders but is far more pronounced in the Solanales, which carry several derived helpers and a correspondingly larger sensor repertoire. Representative species are shown for each order. NRC-H, NRC helper NLRs; NRC-S, NRC-dependent sensor NLRs. This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

A second, non-exclusive constraint is the requirement to remain compatible with NRC0 itself, since changes that compromise signaling to the helper would be selected against. NRC-S that have shifted onto a more recently derived helper appear correspondingly released from at least the latter constraint, and their physical clustering suggests that recent tandem duplication has accelerated their diversification, plausibly under pressure to recognize a wider range of effectors, to evade suppression, or to add redundancy to the pathway (Derevnina et al., 2021; Contreras et al., 2023b).

Together these results suggest that the differential tempo of evolution within the NRC network is not incidental but reflects the functional position each receptor occupies. The network confers resistance against an unusually broad range of pathogens, spanning bacteria, fungi, oomycetes, nematodes, and viruses, and individual members such as

NRC6 have taken on more specialized tissue- or pathogen-specific roles (Lee et al., 2021). A system in which conserved and rapidly evolving components coexist can retain reliable core signaling while continuing to sample new recognition specificities, and the comparatively simple lettuce network offers a model for how a shift in helper dependency might have licensed that diversification. It is also striking that closely related and functionally connected receptors can differ so markedly in evolutionary rate, echoing the contrast between diverse and conserved NLR types described at the lettuce RGC2 locus (Kuang et al., 2004). Establishing the molecular basis of helper specificity, and of sensor–helper communication more generally, is the natural next step, and would inform both resistance breeding and immune receptor engineering in asterid crops.

4.7 Supplementary figures

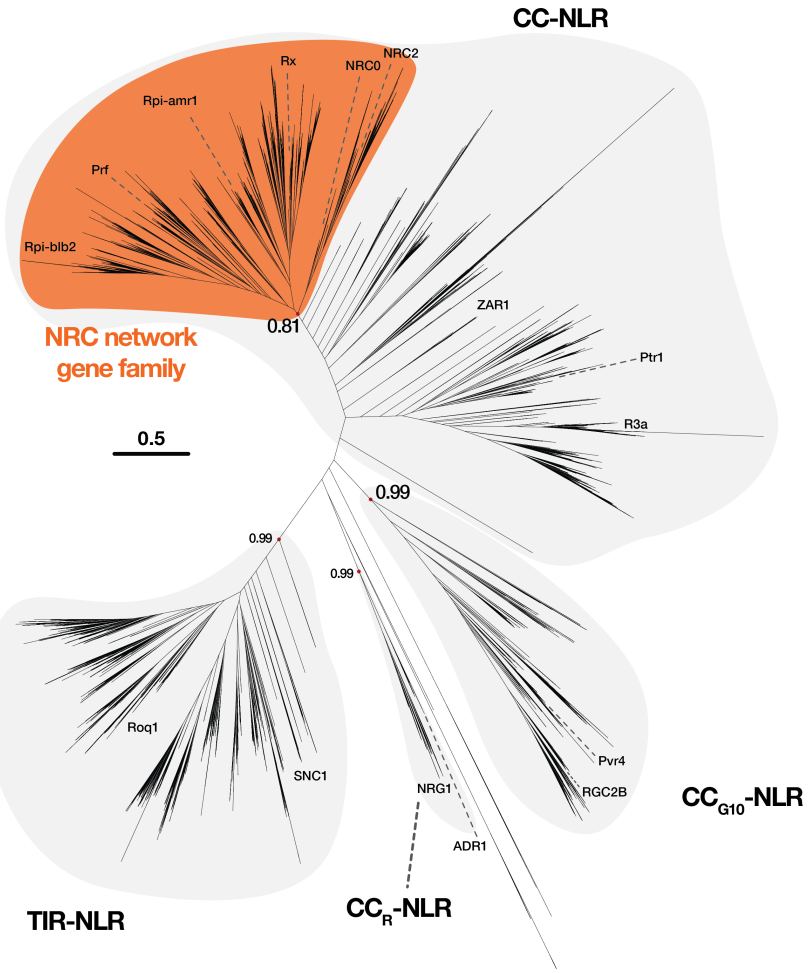


Figure 4.S1: Phylogenetic tree of NB-ARC domains from Solanales and Asterales NLRs with RefPlantNLR as the reference.
The tree from which the monophyletic NRC network was delimited, comprising all non-redundant NLR sequences extracted from the 40 Solanales and 29 Asterales genomes together with RefPlantNLR reference sequences (Kourelis et al., 2021). This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

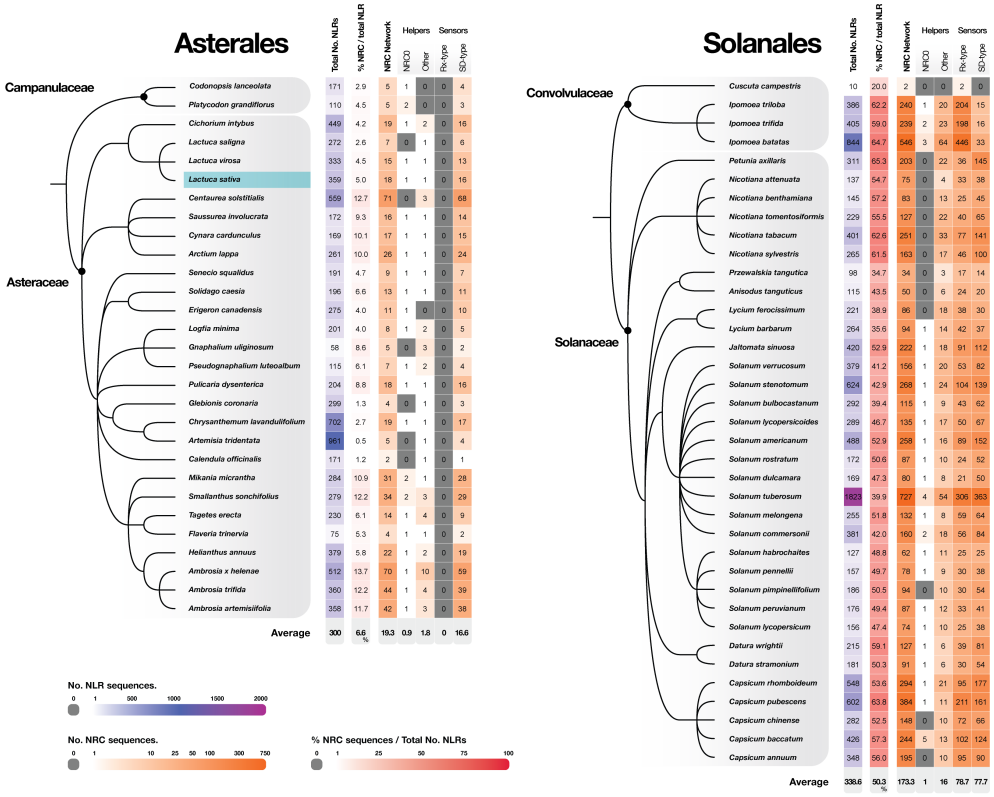


Figure 4.S2: Total NLRs, NRC network sequences, NRC0s, other NRC helpers, Rx-type NRC-S, SD-type NRC-S, and the percentage of NRC sequences per species.

Counts are shown for each of the studied Asterales and Solanales species alongside the corresponding phylogeny, with order-level averages given at the foot of each panel. This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

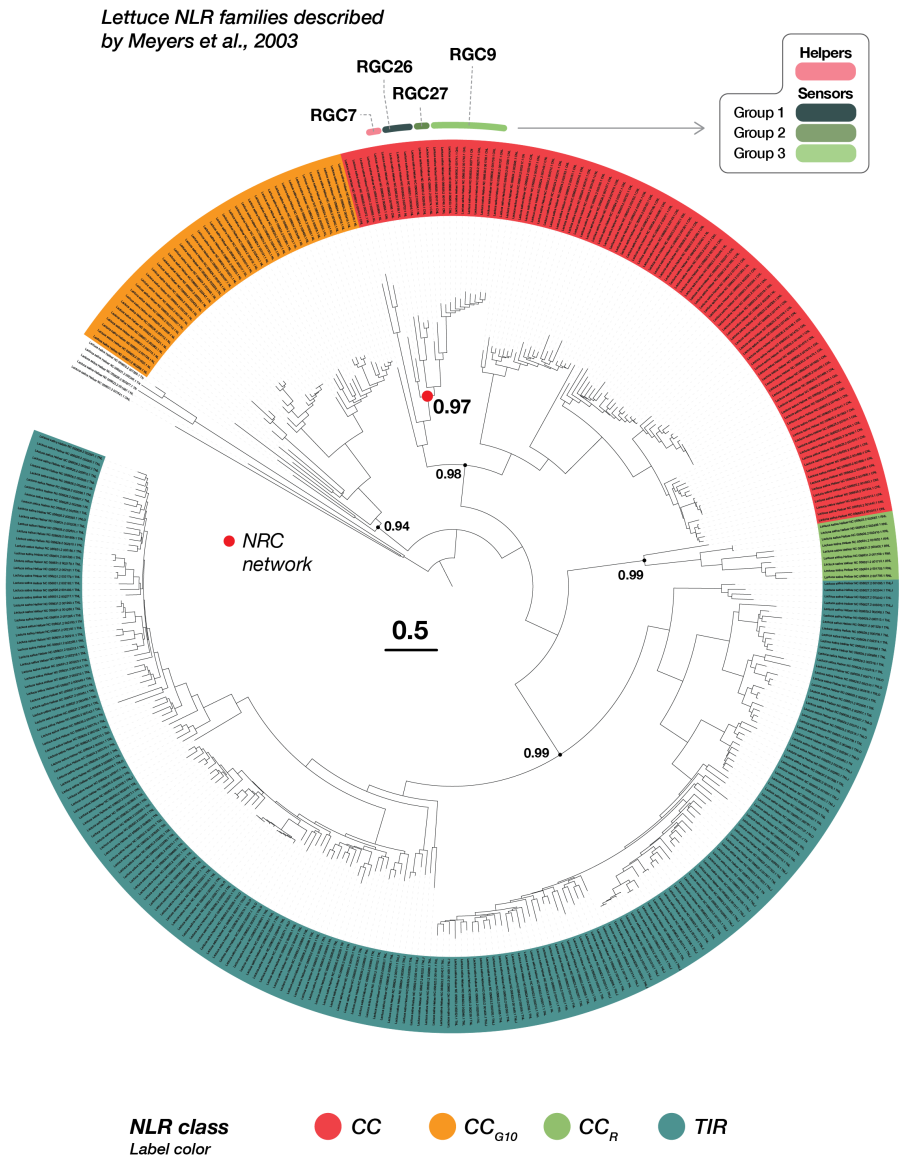


Figure 4.S3: Phylogenetic tree of the common lettuce (*Lactuca sativa*) NLRome featuring 359 sequences.

NLR classes and the previously described Resistance Gene Candidate (RGC) groups are indicated, showing the placement of the NRC network sequences within the wider lettuce NLRome (Meyers et al., 1998). This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

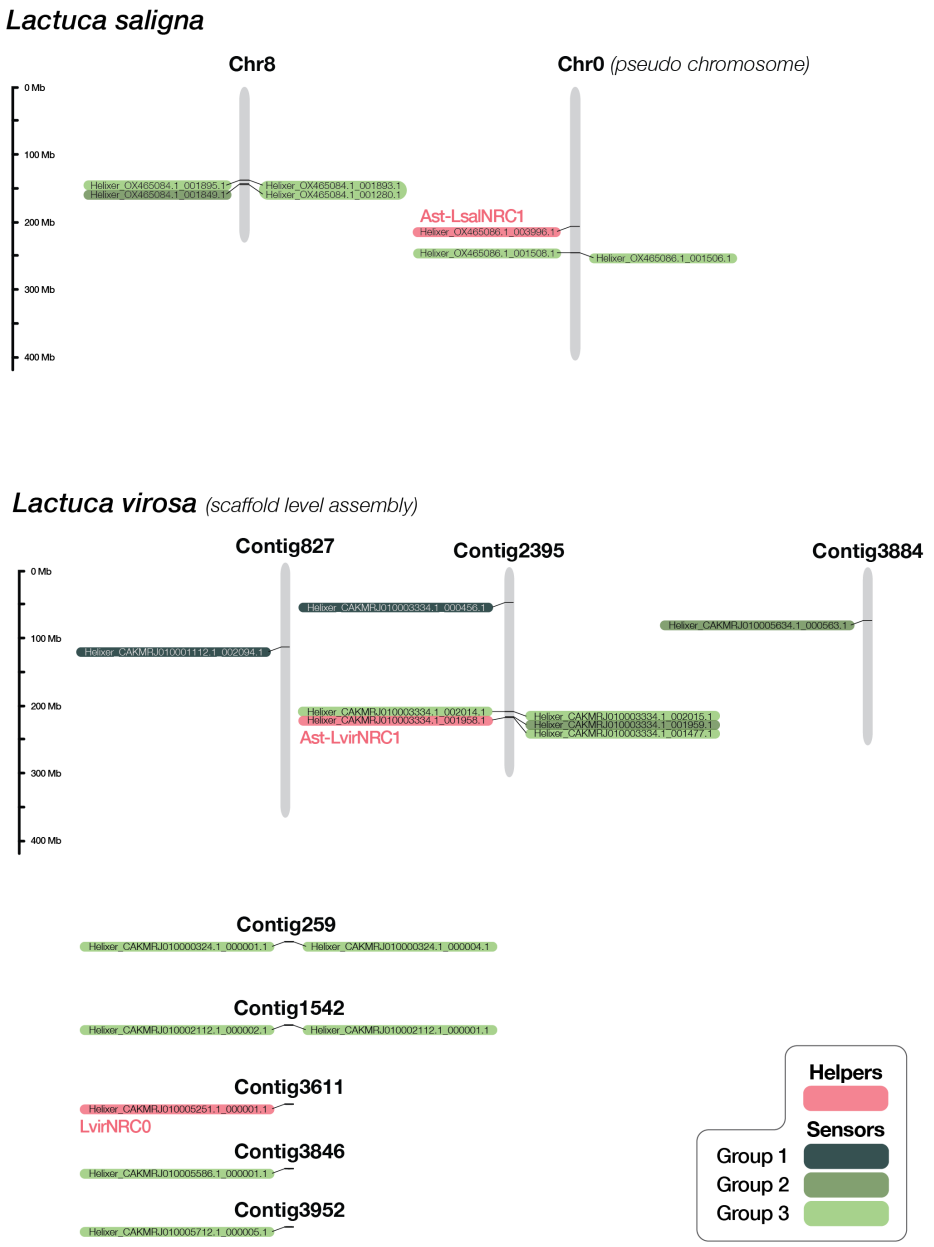


Figure 4.S4: Physical map of *Lactuca saligna* and *Lactuca virosa* NRC sequences. Because chromosome-level assemblies are unavailable for these species, NRC positions are shown on contigs; Ast-LvirNRC1 is located close to group 2 and group 3 NRC-S. This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

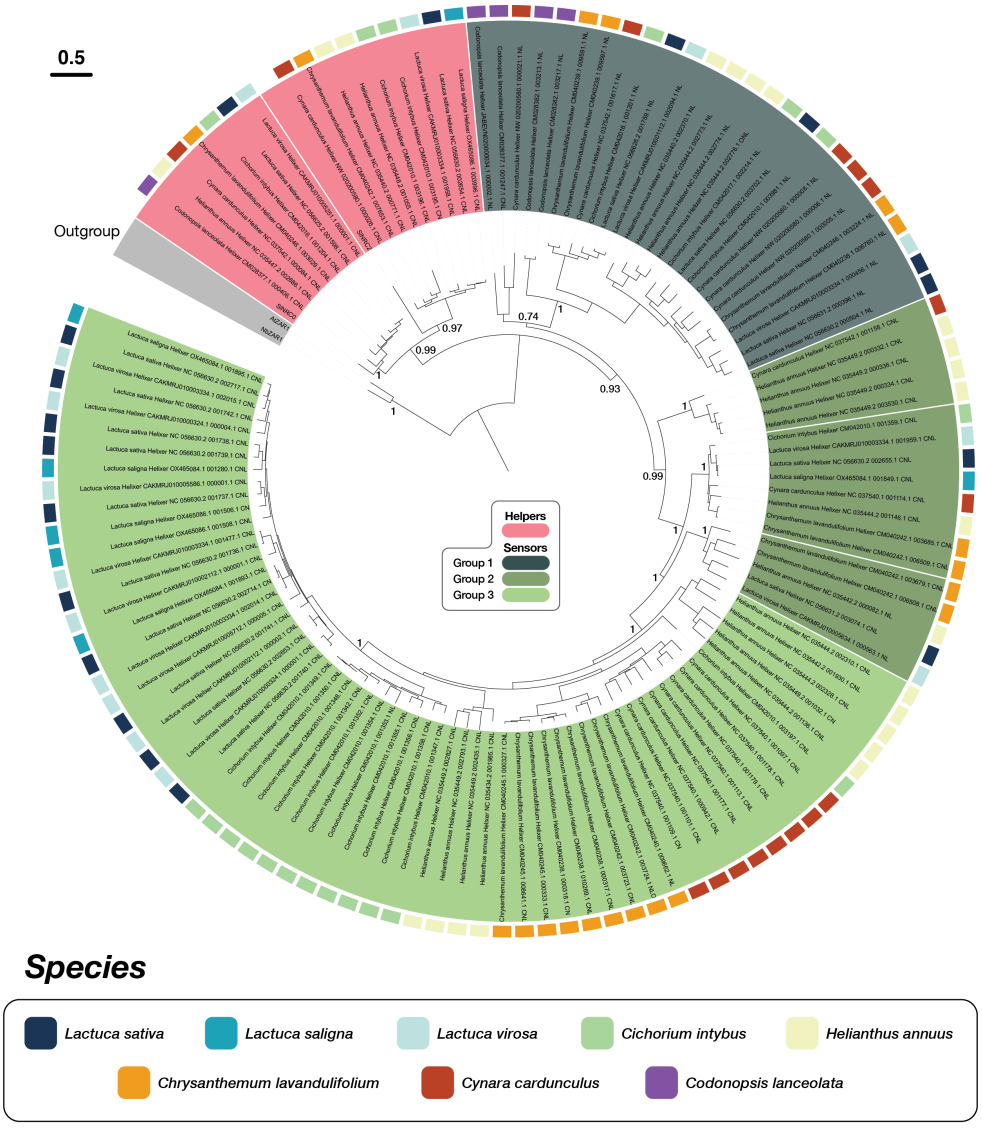


Figure 4.S5: Phylogenetic tree of NRC sequences from *Lactuca* species and five further Asterales.

Sequences from *Lactuca sativa*, *L. saligna*, and *L. virosa* are shown together with those of *Codonopsis lanceolata*, *Helianthus annuus*, *Cichorium intybus*, *Chrysanthemum lavandulifolium*, and *Cynara cardunculus*, and were used to assign NRC sequences in these species to the *Lactuca* helper and sensor groups. This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

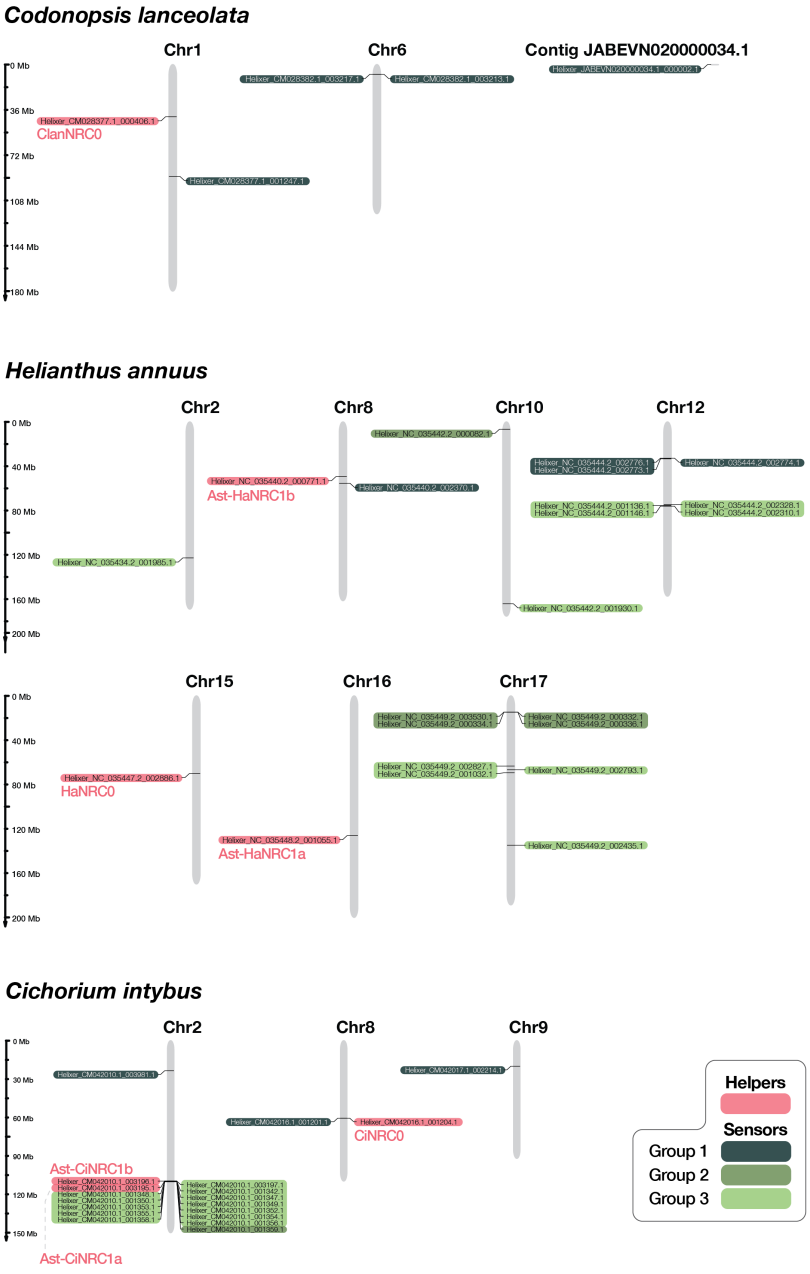


Figure 4.S6: Physical map of *Codonopsis lanceolata*, *Helianthus annuus*, and *Cichorium intybus* NRC sequences.

In *C. intybus*, CiNRC0 lies within 500 kb of a group 1 sensor while both Ast-CiNRC1 copies cluster with group 2 and 3 NRC-S; no helper is clustered with a sensor in the other two species. This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

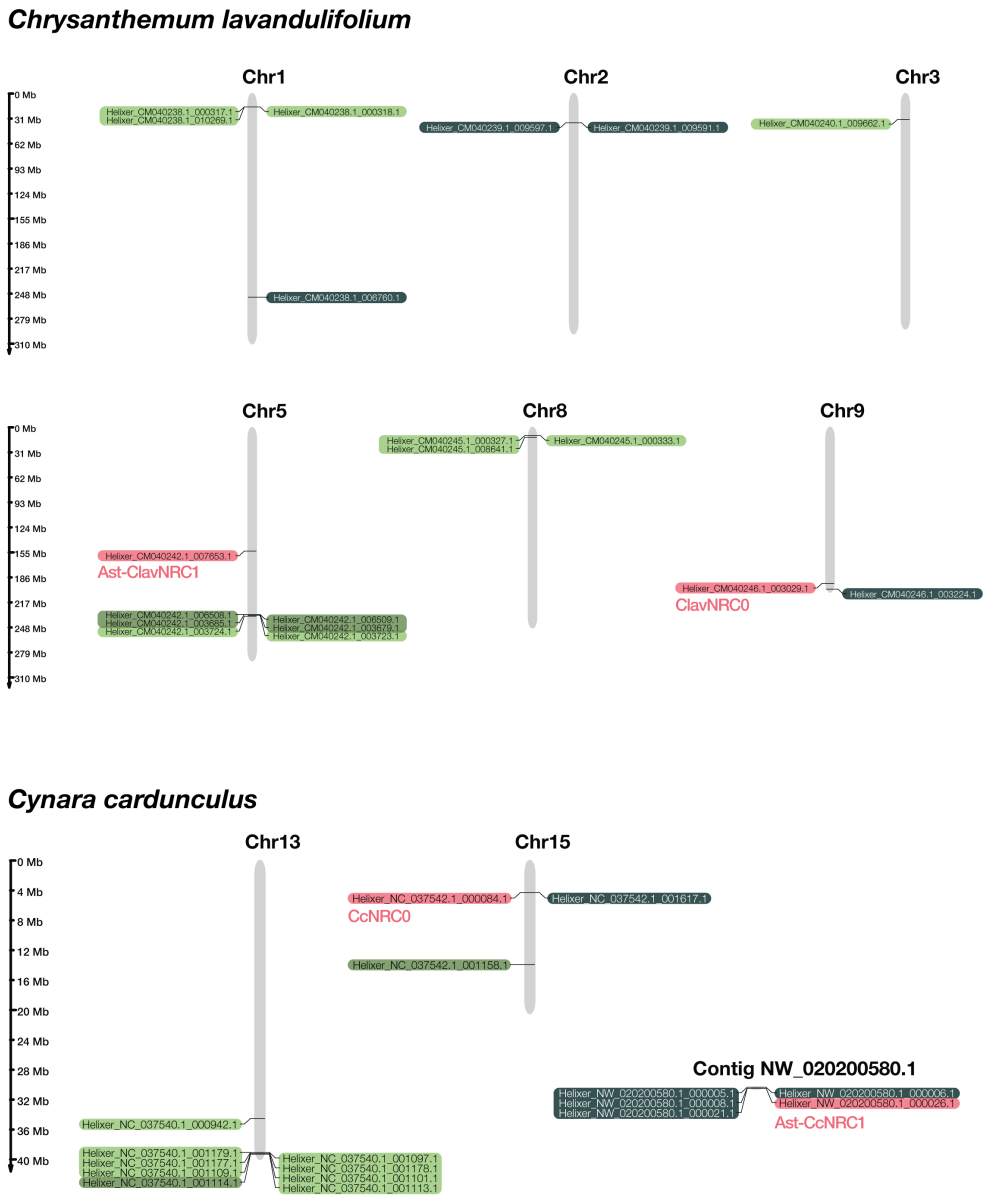


Figure 4.S7: Physical map of *Chrysanthemum lavandulifolium* and *Cynara cardunculus* NRC sequences.

In *C. cardunculus*, CcNRC0 lies close to a group 1 sensor and Ast-CcNRC1 is clustered with four group 1 NRC-S, whereas no helper is clustered with a sensor in *C. lavandulifolium*. This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

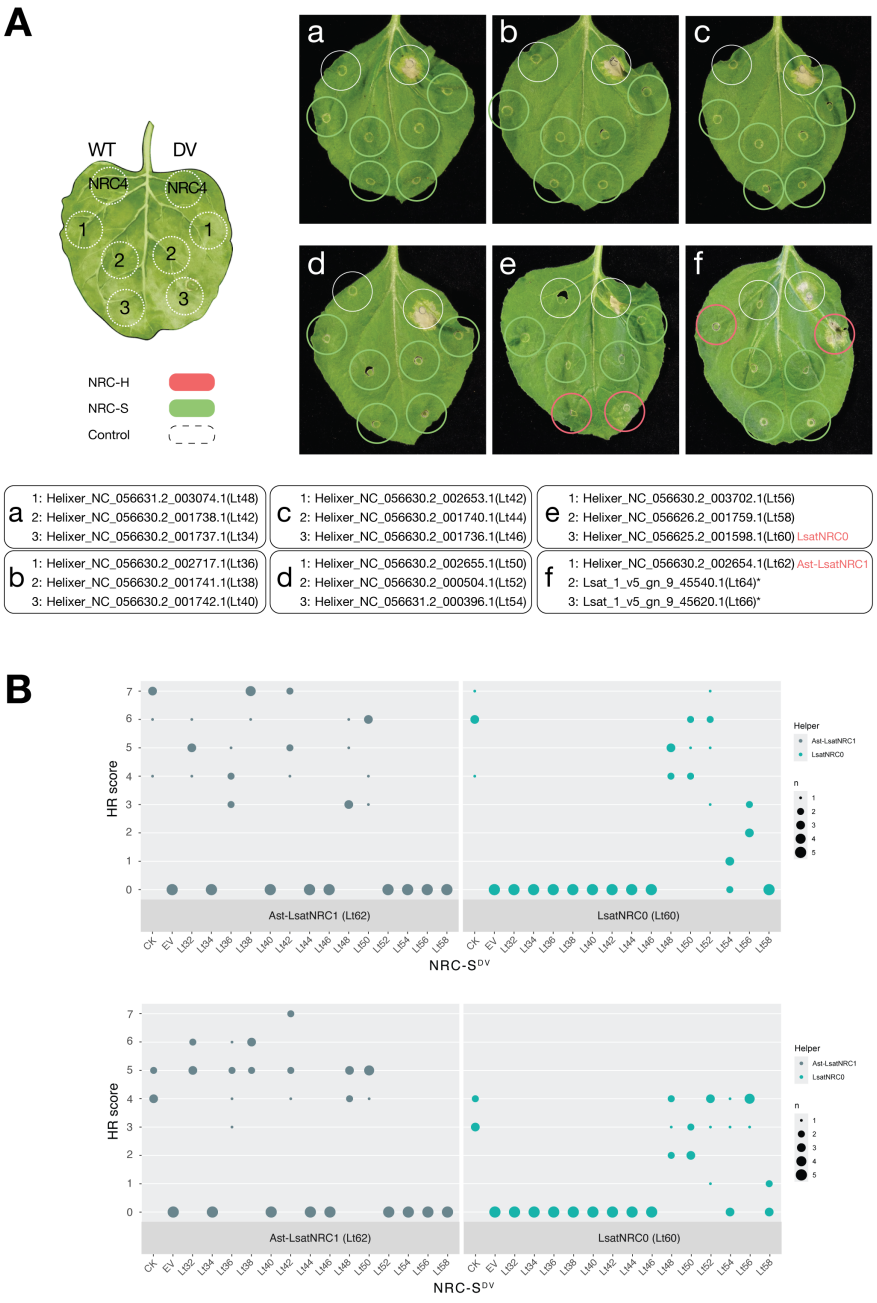


Figure 4.S8: Hypersensitive response (HR) cell death assays of NRC-H and NRC-S. Autoactive (MHD D-to-V) mutants of the lettuce NRC-H and NRC-S were expressed alone and in sensor–helper combinations in the *nrc2/3/4* knockout line of *Nicotiana benthamiana*, with the corresponding HR scoring. This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

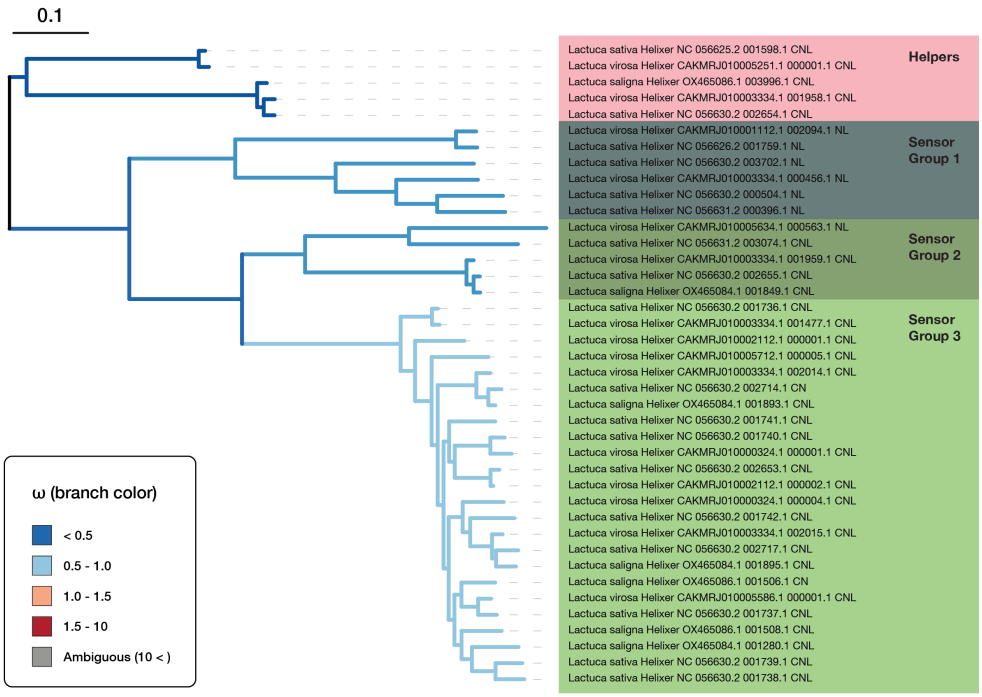


Figure 4.S9: Multi-ratio model branch-specific dN/dS (ω) values across the *Lactuca* NRC phylogeny.

Each phylogroup was assigned its own ω parameter under the branch model in CodeML and compared against the one-ratio model (Yang, 1997). This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

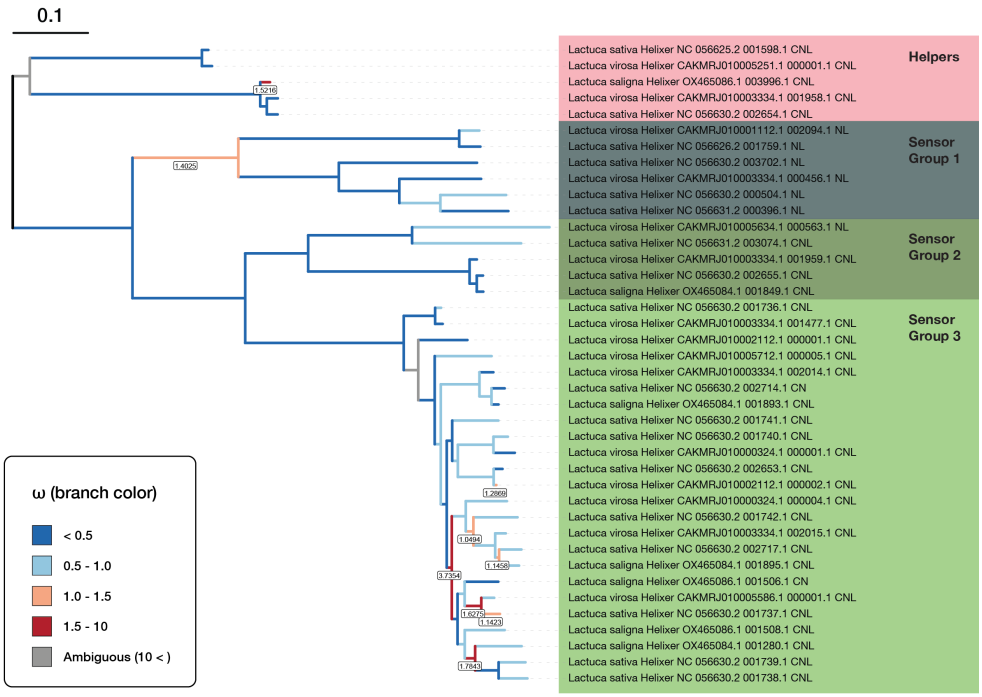


Figure 4.S10: Free-ratio model branch-specific dN/dS (ω) values across the *Lactuca* NRC phylogeny.

Per-branch ω estimates are shown descriptively rather than as a formal test; branches with $\omega \geq 10$ reflect artifacts of low synonymous substitution rates rather than meaningful estimates. This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

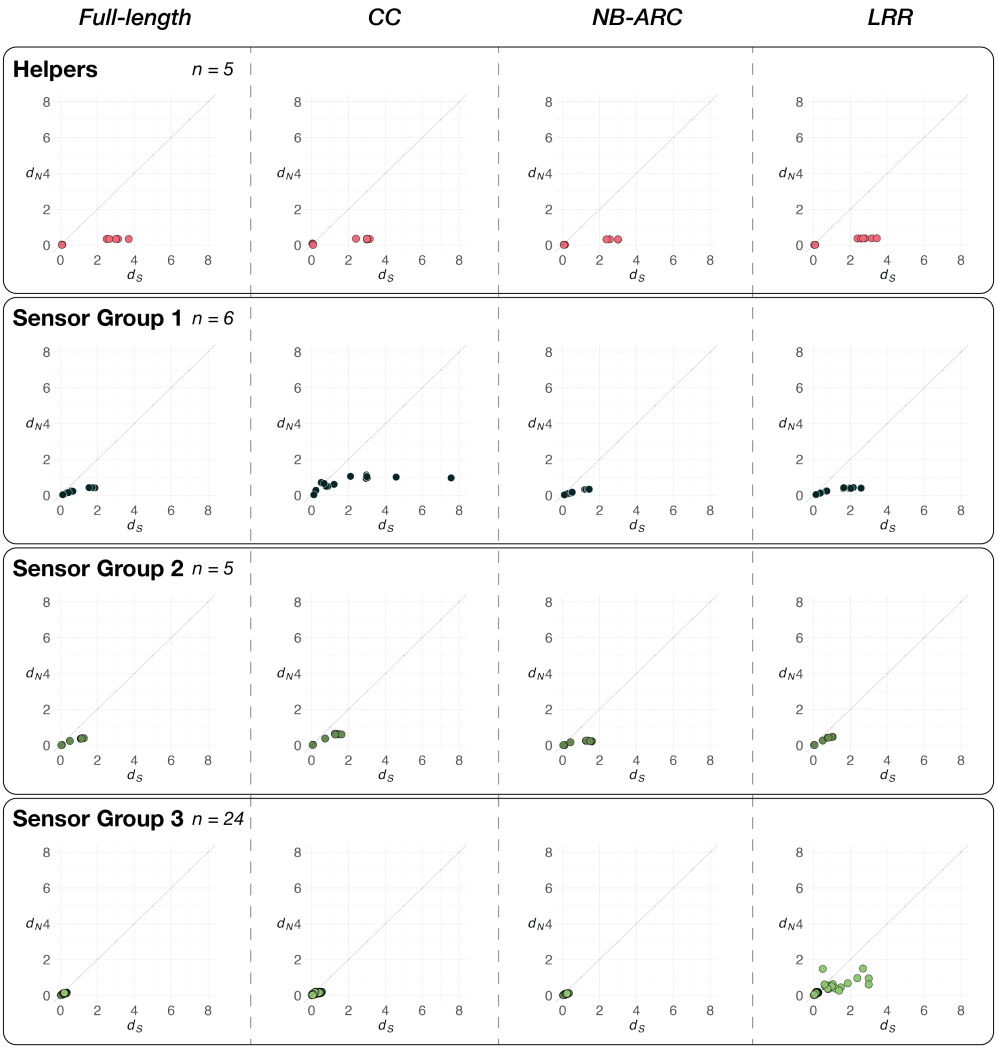


Figure 4.S11: Pairwise dN and dS values for full-length sequences and the CC, NB-ARC, and LRR domains of NLRs within the *Lactuca* NRC phylogroups.
Values were estimated using the Nei and Gojobori method (Nei and Gojobori, 1986). This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

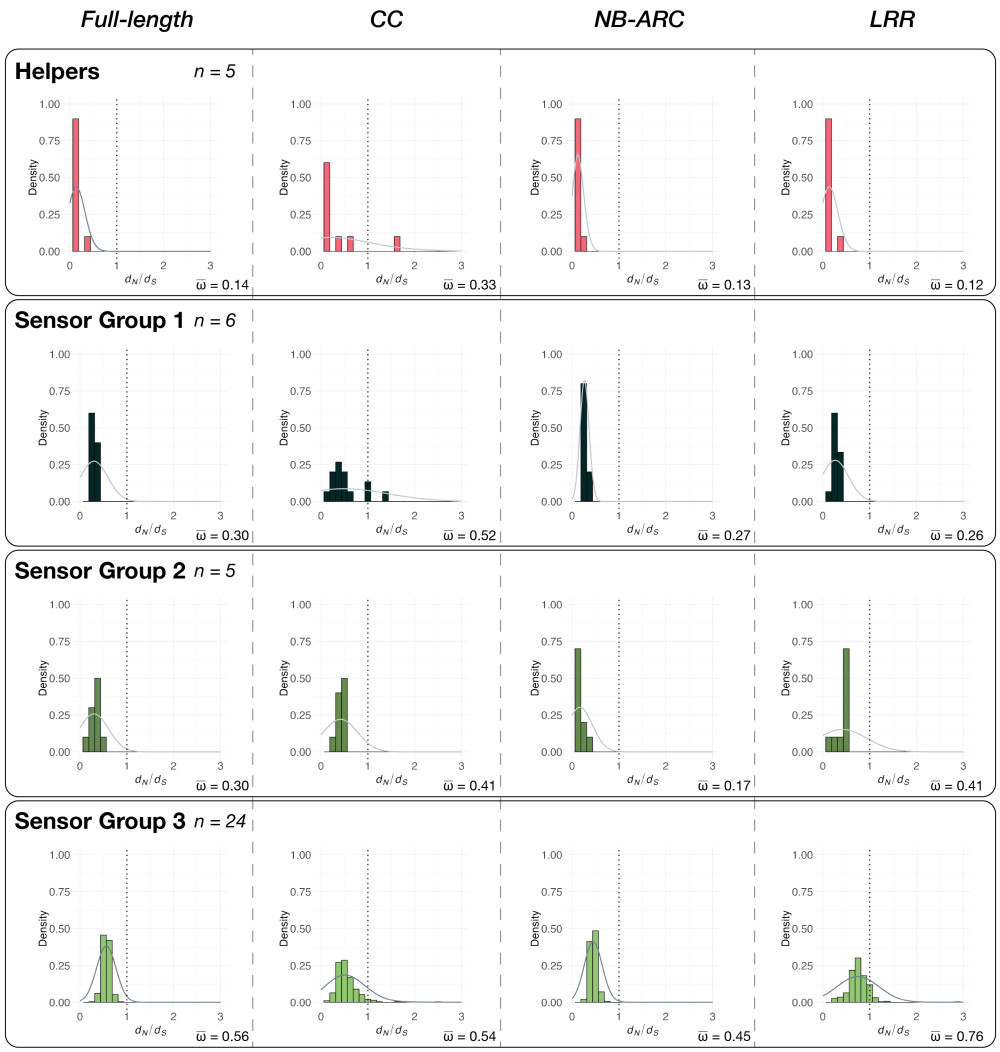


Figure 4.S12: Histogram of pairwise dN/dS ratios for full-length sequences and the CC, NB-ARC, and LRR domains of NLRs within the *Lactuca* NRC phylogroups.
Helpers show the lowest average ω across all comparisons, while the CC domain of group 1 NRC-S returns $\omega > 1$. This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

Chapter 5: Computational structural modeling of NLR resistosomes reveals structural diversity in plant CC-type NLRs

Part of the work described in this chapter has been published in Madhuprakash et al. (Madhuprakash et al., 2024) and appears here with permission. Cryo-EM structure determination was carried out by Jogi Madhuprakash and Michael Webster. Biochemical and cell death assays were performed by Mauricio Contreras and Jogi Madhuprakash. My contribution comprised the computational structural modeling and the comparative analyses described below.

5.1 Introduction

The structural basis of NLR-mediated immunity has been transformed over the past few years by a succession of cryo-EM structures of activated receptor complexes. The first of these, the *Arabidopsis thaliana* CC-NLR AtZAR1, was shown to assemble upon activation into a wheel-shaped pentamer, termed a resistosome, in which the N-terminal α1 helices of the five protomers project outwards to form a funnel that inserts into the plasma membrane and acts as a calcium-permeable channel (Wang et al., 2019b; Wang et al., 2019a; Bi et al., 2021). The wheat CC-NLR Sr35 was subsequently found to adopt the same pentameric architecture (Förderer et al., 2022; Zhao et al., 2022), while the TIR-NLR resistosomes ROQ1 and RPP1 assemble as tetramers with an entirely different signaling output (Martin et al., 2020; Ma et al., 2020). For CC-NLRs specifically, the pentamer was regarded as the canonical activated state.

In Chapter 3, we showed that the helper NLR NRC2 rests as a homodimer (Selvaraj et al., 2024). Upon sensor-mediated activation, NRC2 assembles into a hexameric resistosome rather than a pentamer (Madhuprakash et al., 2024). Structural studies on the related helper NRC4 independently confirmed this hexameric architecture (Liu et al., 2023a). Because NRC helpers are activated indirectly by NRC-S that do not incorporate into the resistosome (Contreras et al., 2022; Ahn et al., 2023), these findings overturned the assumption of a universal pentamer. They raised the broader question of how widely oligomeric states vary across plant CC-NLRs, and whether pentamers or hexamers represent the more common active form.

Testing this experimentally across dozens of receptors remains a bottleneck by protein expression, purification, and cryo-EM data collection. Computational structural mod-

eling offers a scalable alternative. AlphaFold 2 established high-throughput and high-quality monomer prediction (Jumper et al., 2021). AlphaFold 2 multimer enabled high-confidence protein complex predictions (Evans et al., 2021), and AlphaFold 3 introduced modeling proteins with ligand, ions, and small molecules (Abramson et al., 2024). Modeling lipids is particularly useful for CC-NLR resistosomes. The CC-NLR resistosome executes cell death via a membrane-inserting funnel and typically mediated by the α1 helix, which is not resolved in most cryo-EM resistosome models.

In this chapter, I use AlphaFold 3 to address three core questions. First, I test whether computational modeling recapitulates the experimentally resolved NRC2 hexamer, evaluating how predictions hold across different seeds, model versions, and lipid ligands. Second, I benchmark whether model confidence metrics can identify the correct oligomeric state using the established ZAR1 and Sr35 pentamers. Third, I map oligomerization propensity across the NRC helper clade and diverse CC-NLRs across the plant phylogeny.

5.2 AlphaFold 3 predicts a lipid-bound NRC2 resistosome with high confidence

Activated CC-NLR resistosomes insert into the host plasma membrane via a funnel formed by their N-terminal α1 helices. In the NbNRC2 cryo-EM structure, these MADA-containing helices lacked resolved density, leaving the pore unmodeled. I therefore used AlphaFold to predict this missing region and complete the structural model (Abramson et al., 2024).

Using the cryo-EM hexamer as a reference, I compared the performance of AlphaFold 2, AlphaFold 3, and AlphaFold 3 supplemented with 50 oleic acid molecules to simulate the membrane environment (Jumper et al., 2021; Abramson et al., 2024). Given the large size of the NLR oligomer and GPU memory limits, I restricted modeling to the CC-NB-ARC module containing the oligomerization interfaces. I evaluated the resulting predictions using pLDDT, pTM, ipTM, predicted aligned error, minimum per-chain pTM, and chain-pair ipTM.

All three approaches produced high-confidence models with pTM and ipTM values above 0.7, aligning to the experimental structure with RMSD values below 1.8 Å (Figures 5.1A and 5.S1). However, protomer orientations differed fundamentally between versions. In the AlphaFold 2 model, the N-terminal α1 helix pointed inward toward the central pore. In contrast, both AlphaFold 3 models oriented the helix outward, assembling a complete funnel. Adding oleic acid molecules further improved model quality. The α1 helix formed

confident contacts with the surrounding lipids, yielding higher local confidence scores than predictions run without lipids (Figures 5.1B and 5.S1). Including a membrane proxy therefore directly guides the outward orientation of the pore-forming helix missing from the cryo-EM map.

I next examined whether the models recovered the experimental oligomerization contacts. Using a 4.5 Å contact cutoff, the AlphaFold 3 model successfully captured most inter-protomer contacts from the cryo-EM structure (Figure 5.1D). Prediction accuracy varied across individual domains. Adjacent CC domains aligned closely with the experimental contacts, whereas adjacent NBD interfaces diverged from the cryo-EM model (Figures 5.1E and 5.S3). AlphaFold 3 thus reliably recovers the overall quaternary architecture, even when specific inter-domain contacts and specific side-chains show local differences.

5.3 Model confidence is highest for the hexameric stoichiometry

After confirming that AlphaFold 3 reliably predicts the hexamer, I tested how varying stoichiometry influences prediction confidence. I modeled NbNRC2 across stoichiometries ranging from four to eight protomers, running ten random seeds per stoichiometry in the presence of 50 oleic acid molecules (Figures 5.1C and 5.S2).

Hexamer models achieved the highest median pTM and ipTM scores, aligning to the experimental structure with RMSD values below 1.8 Å. Pentamers and heptamers scored lower across all metrics, while tetramers and octamers showed lower confidence and higher variance across runs. All oligomeric forms from pentamer to octamer assembled complete funnel structures. In contrast, tetramer runs produced inconsistent assemblies, alternating between closed funnels and split, arc-like incomplete rings (Figure 5.S3A). Model confidence metrics for NbNRC2 therefore clearly distinguish the experimentally determined hexamer from alternative stoichiometries.

These models provide a structural hypothesis for the unresolved N-terminal region. The α1 helices form an outward-facing bundle at the apex of the complex, superimposing onto the protomer core without disrupting the overall fold (Figure 5.S3B). In the hexamer model, these helices frame a central pore lined by a ring of leucine residues at position L126 (Figure 5.S3C). This arrangement provides a concrete model for the channel-forming executioner region.

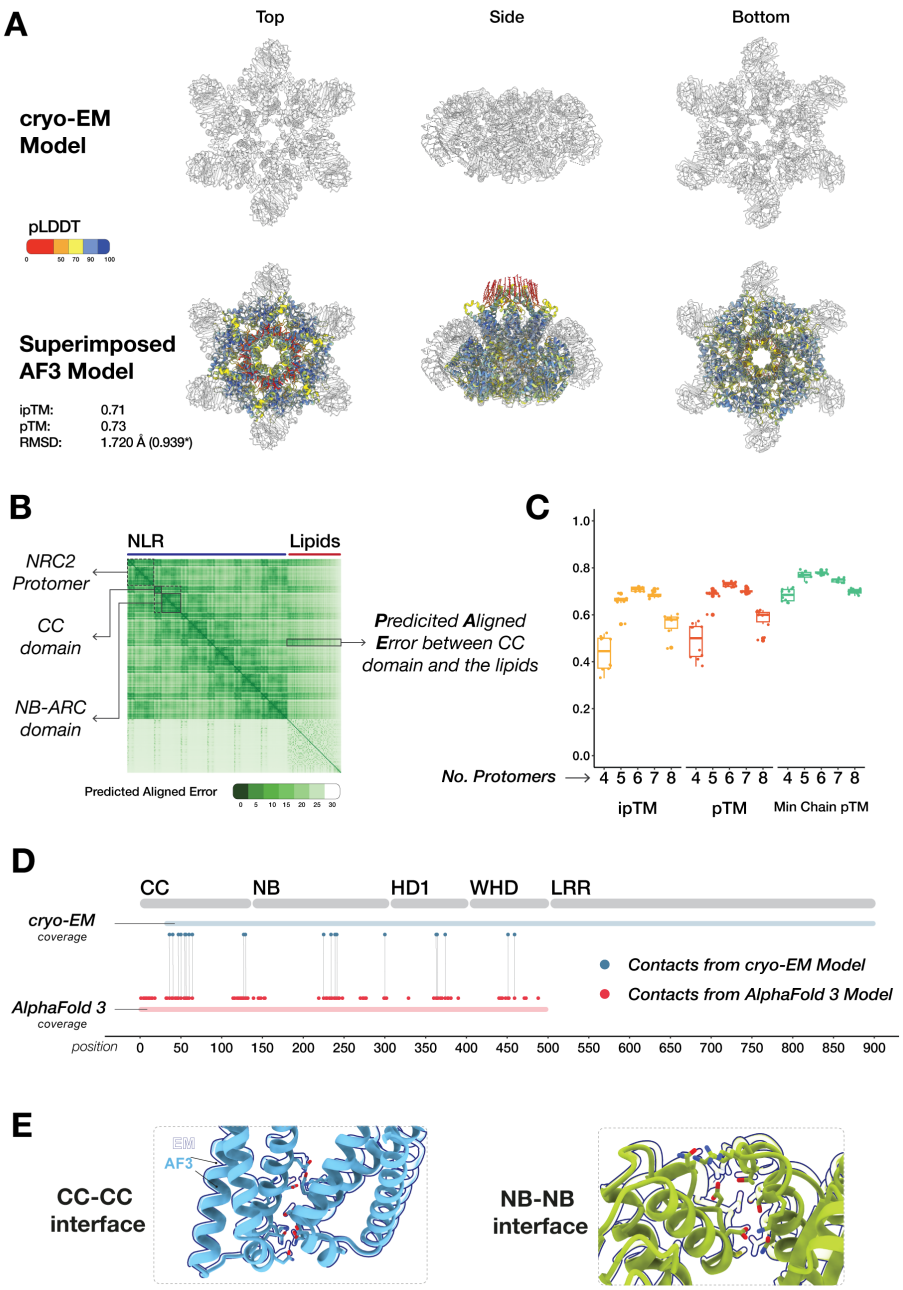


Figure 5.1: AlphaFold 3 recapitulates the NRC2 hexameric resistosome.
(A) Top, side, and bottom views of the cryo-EM model of the NRC2 resistosome (gray) and of the AlphaFold 3 model superimposed on it, colored by pLDDT. Interface pTM (ipTM), pTM, and the root-mean-square deviation from the cryo-EM model are given. (B) Predicted aligned error for the modeled assembly, with the NRC2 protomer, CC domain, and NB-ARC domain indicated, and the elevated error between the CC domains and the lipid molecules highlighted. (C) ipTM, pTM, and minimum per-chain pTM for NRC2 modeled with four to eight protomers. (D) Inter-protomer contacts derived from the cryo-EM model (blue) and from the AlphaFold 3 model (red), mapped onto the NRC2 domain architecture. (E) Superposition of the CC–CC and NBD–NBD interfaces from the cryo-EM (EM) and AlphaFold 3 (AF3) models. This figure has also been published in Madhuprakash et al. (Madhuprakash et al., 2024) and appears here with permission.

5.4 AtZAR1 and TmSr35 are predicted with higher confidence as pentamers

I next asked whether AlphaFold 3 could distinguish the pentameric state of AtZAR1 and TmSr35, the two CC-NLRs with known pentameric resistosome structures (Wang et al., 2019a; Förderer et al., 2022; Zhao et al., 2022). I modeled the CC-NB-ARC modules of both receptors as pentamers and hexamers. While overall confidence scores were lower than those for NbNRC2, both pentamer models aligned closely with their cryo-EM models, yielding RMSDs of 1.51 Å for *Arabidopsis thaliana* AtZAR1 and 1.76 Å for TmSr35 (Figure 5.2A). Crucially, both receptors scored higher as pentamers than as hexamers, matching their experimental assemblies (Figure 5.S6).

These benchmarks show that AlphaFold 3 confidence metrics track the correct oligomeric state across all three structurally characterized CC-NLRs. The helper NRC2 scores highest as a hexamer, while the two singletons score highest as pentamers. However, the score differences between stoichiometries remain modest. These metrics should serve as guides to the most probable state rather than definitive proof of assembly, particularly given known limitations of ipTM in evaluating protein interfaces (Dunbrack, 2025; Lin et al., 2025).

5.5 Modeling across the NRC helper clade

Building on the high-confidence prediction of the NbNRC2 hexamer, I examined structural models of other NRC helpers. I modeled eleven representatives from the major NRC clades described in Chapter 3 as both pentamers and hexamers (Figure 5.S4).

Three of the eleven helpers, specifically SlNRC0, NbNRCX, and NbNRC7a, failed to yield confident models and returned ipTM and pTM scores below 0.6 across both stoichiometries (Figure 5.S5). The remaining eight helpers returned confidence scores of 0.6 or higher in both configurations. For all eleven helpers, hexamer models scored higher than pentamer models, although this difference was often subtle. Every hexamer model except that of NbNRC7a also achieved minimum per-chain pTM scores of 0.7 or higher, a benchmark reached by fewer pentamer models (Figure 5.S5).

The poor prediction scores for NRC0 are particularly notable given its status as the ancestral helper. They align with previous observations of an incomplete N-terminal funnel in modeled NRC0 oligomers (Selvaraj et al., 2024), a feature also observed for the lettuce ortholog LsatNRC0 in Chapter 6. Whether this outcome reflects a genuine divergence in

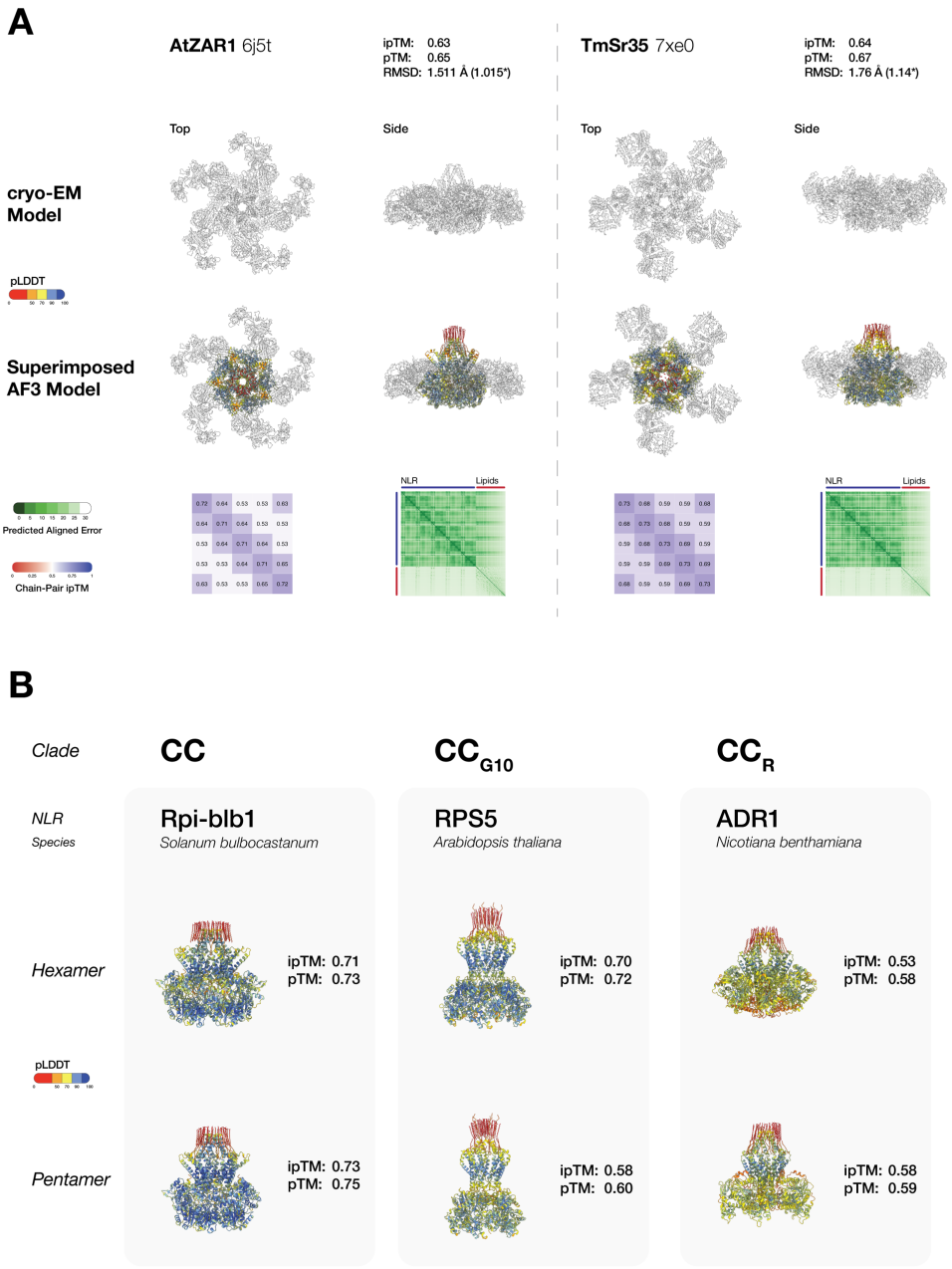


Figure 5.2: AlphaFold 3 modeling of known pentameric resistosomes and of CC-type NLRs from different clades.

(A) Cryo-EM models (gray) and superimposed AlphaFold 3 models, colored by pLDDT, for the *Arabidopsis thaliana* ZAR1 (PDB 6J5T) and *Triticum monococcum* Sr35 (PDB 7XE0) resistosomes, shown in top and side views with ipTM, pTM, and RMSD values. Chain-pair ipTM matrices and predicted aligned error plots are shown beneath each, with the receptor (blue) and lipid (red) regions indicated. (B) AlphaFold 3 models of representative NLRs from the CC, CC_{G10}, and CC_R clades, modeled as both hexamers and pentamers, with ipTM and pTM values: Rpi-blb1 (*Solanum bulbocastanum*), RPS5 (*Arabidopsis thaliana*), and ADR1 (*Nicotiana benthamiana*). This figure has also been published in Madhuprakash et al. (Madhuprakash et al., 2024) and appears here with permission.

resistosome assembly or simply shallower sequence alignments available for divergent helpers remains an open question.

5.6 CC-NLR resistosomes show a diversity of predicted stoichiometries and N-terminal pores

Following the benchmark results for NbNRC2, AtZAR1, and TmSr35, I modeled a phylogenetically diverse panel of plant NLRs with unknown resistosome architectures (Figure 5.S7). I tested eight CC-type, six CC_{G10}-type, and two CC_R-type NLRs as both pentamers and hexamers to evaluate whether their resistosomes could be predicted with high confidence and to determine their preferred stoichiometry (Figure 5.2B).

Both the CC-type and CC_{G10}-type clades contained representatives of hexameric and pentameric assemblies. Every receptor within the CC-NLR clade modeled with ipTM, pTM, and minimum per-chain pTM values of 0.6 or higher (Figures 5.S8–5.S10 and 5.S14). AtHrt1, AtRpm1, and SlPtr1 yielded higher confidence as hexamers than as pentamers. In contrast, OsPit, TaPm2a, StR3a, SbRpi-blb1, and BvRz2 scored higher in the pentameric form (Figures 5.S8–5.S10). In StR3a and BvRz2, the CC domain and N-terminal α1 helix modeled with lower confidence than in the other resistosomes regardless of stoichiometry (Figures 5.S9 and 5.S10).

The CC_{G10}-NLRs showed a similar split in preferred stoichiometry. AtRPS5 returned higher confidence as a hexamer (Figure 5.S12), whereas CaPvr4, LsatRGC2B, GmRsv3, and CmVat scored higher as pentamers (Figures 5.S11 and 5.S12). AtRPS2 was the sole exception in this group, modeling poorly in both stoichiometries with scores below 0.6 (Figures 5.S11 and 5.S14).

The two CC_R-NLRs, NbADR1 and NbNRG1, failed to generate confident models in either configuration and scored below 0.6 throughout (Figures 5.S13 and 5.S14). Even within this low confidence range, both proteins scored slightly higher as pentamers than as hexamers. Because these helper receptors function downstream of TIR-NLR signaling (Castel et al., 2019; Saile et al., 2020; Jacob et al., 2021), these low scores likely reflect current modeling limitations for divergent RPW8-type CC domains rather than an inability to oligomerize.

5.7 Conclusions and discussion

Benchmarked against three experimentally resolved CC-NLR resistosomes, AlphaFold 3 predicted activated NRC hexamers and CC-NLR pentamers with high confidence. The overall fold of the NbNRC2 hexamer was recovered at 1.72 Å RMSD along with key interface residues, while AtZAR1 and TmSr35 pentamers aligned at 1.51 Å and 1.76 Å, respectively. While contacts between adjacent CC domains aligned approximately with experimental interfaces, residues at the NBD–NBD interface showed different side-chain orientations from the cryo-EM structure. The modeling pipeline is therefore best suited to recovering overall resistosome architecture rather than precise atomic contacts across every interface.

The inclusion of membrane lipids proved to be a critical practical factor. Supplying oleic acid molecules enabled high-confidence modeling of the N-terminal funnel. These funnels are frequently unresolved in cryo-EM maps due to non-functional mutations introduced to block cell death or the loss of native lipids during purification. AlphaFold 3 can therefore help fill structural gaps in resistosomes that remain technically challenging to solve experimentally. These findings also demonstrate that providing a lipid proxy substantially improves computational models of membrane-associated complexes.

Model confidence metrics accurately distinguished experimental stoichiometry in all three benchmark cases, favoring a hexamer for NbNRC2 and pentamers for AtZAR1 and TmSr35. This benchmark carries a broader implication. Because published pentameric structures were obtained from proteins expressed in insect cells rather than plants, independent computational support for pentamers suggests that these assemblies are not expression artifacts. Direct purification of these complexes from plant tissue will still be valuable to confirm these states in their native background. While score differences between stoichiometries remain modest, these metrics consistently identify the more probable assembly.

Applied across a broader set of plant NLRs, the pipeline generated confident models for members of both canonical CC and CC_{G10} clades. Both groups contained receptors that favored pentamers alongside others that favored hexamers, demonstrating that stoichiometry is not fixed within a phylogenetic class. Several CC-NLRs and both CC_R-NLRs failed to yield confident assemblies. Within the NRC clade, the three helpers that modeled poorly were SlNRC0, NbNRCX, and NbNRC7a. This is an intriguing pattern, as NRC0 represents the ancestral network helper and NRCX acts as a negative regulator that lacks autonomous cell death activity (Sakai et al., 2024; Adachi et al., 2023a). Building on these observations, we recently developed an expanded computational pipeline to systematically assess oligomeric states across the NRC family. This led to the

discovery of an atypical 11-mer resistosome assembly in NRC7 (Toghani et al., 2026b). This result provides further evidence for structural diversity in CC-NLR resistosomes, demonstrating that stoichiometry can vary even among closely related helper paralogs.

The finding that the conserved NB-ARC domain can support multiple oligomeric states highlights unexpected structural flexibility in this family. Further computational and structural work may identify the sequence determinants within the CC and NB-ARC domains that drive the formation of diverse resistosome assemblies. Additionally, AlphaFold 3 provides a platform for hypothesis generation that complements the phylogenomic and functional analyses developed in Chapters 3 and 4.

5.8 Supplementary figures

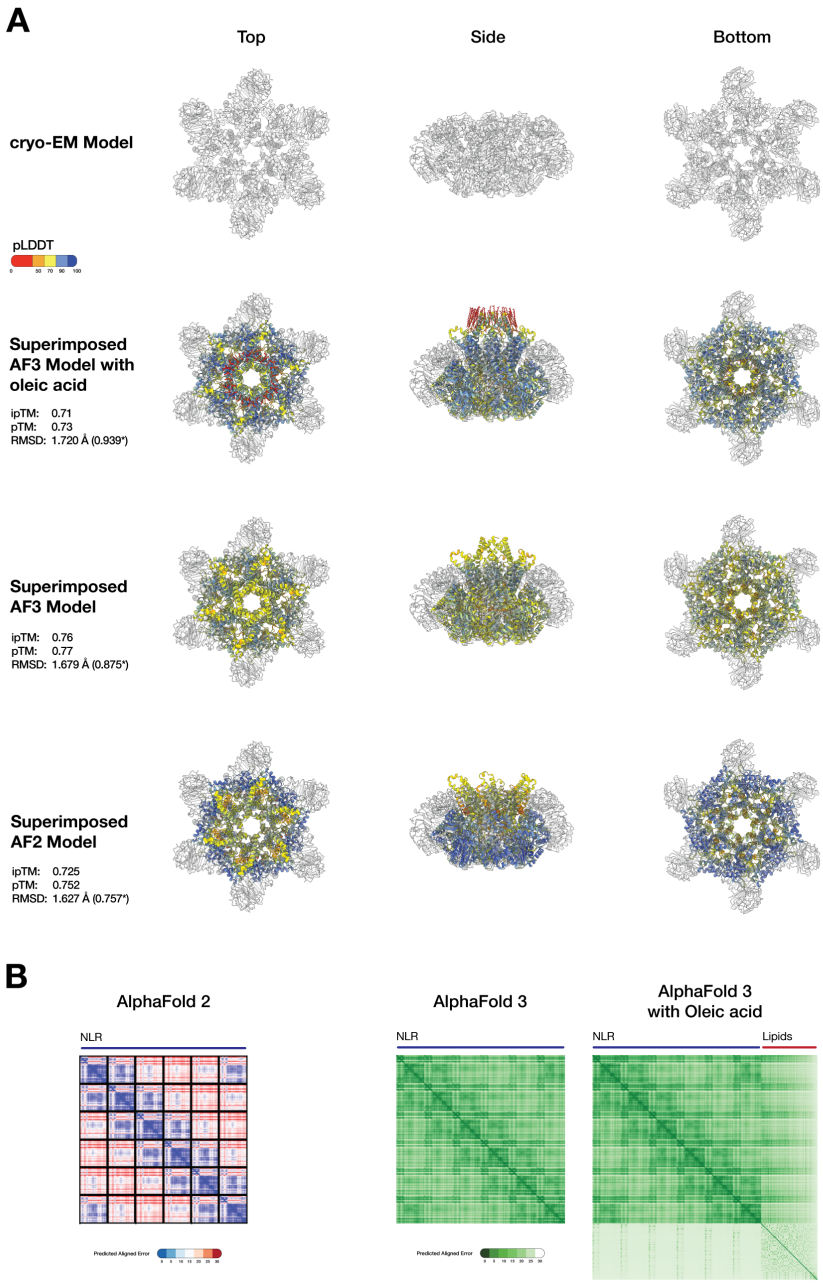


Figure 5.S1: Comparison of AlphaFold 2 and AlphaFold 3 models of the NRC2 hexamer. (A) Top, side, and bottom views of the cryo-EM model (gray) with superimposed models generated by AlphaFold 3 in the presence of oleic acid, by AlphaFold 3 without lipids, and by AlphaFold 2, each colored by pLDDT and annotated with ipTM, pTM, and RMSD. (B) Predicted aligned error plots for the three models; the AlphaFold 2 prediction shows substantially higher error between protomers than either AlphaFold 3 model. This figure has also been published in Madhuprakash et al. (Madhuprakash et al., 2024) and appears here with permission.

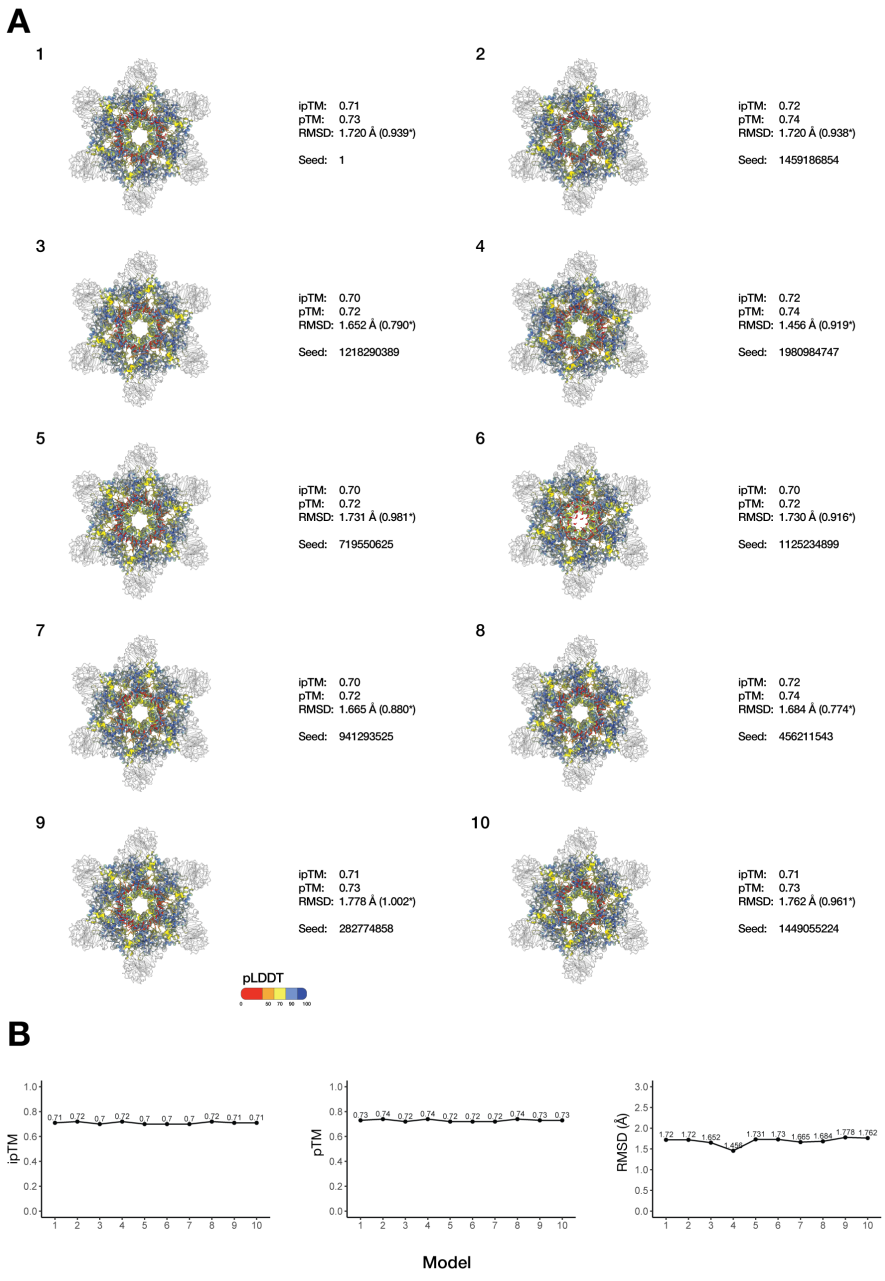


Figure 5.S2: Reproducibility of the NRC2 hexamer prediction across random seeds.
(A) Ten independent AlphaFold 3 predictions of the NRC2 hexamer, each generated with a different random seed, shown in top view and colored by pLDDT, with the corresponding ipTM, pTM, RMSD, and seed value. (B) ipTM, pTM, and RMSD across the ten models. Predictions are tightly clustered, with ipTM between 0.70 and 0.72 and RMSD between 1.46 and 1.78 Å. This figure has also been published in Madhuprakash et al. (Madhuprakash et al., 2024) and appears here with permission.

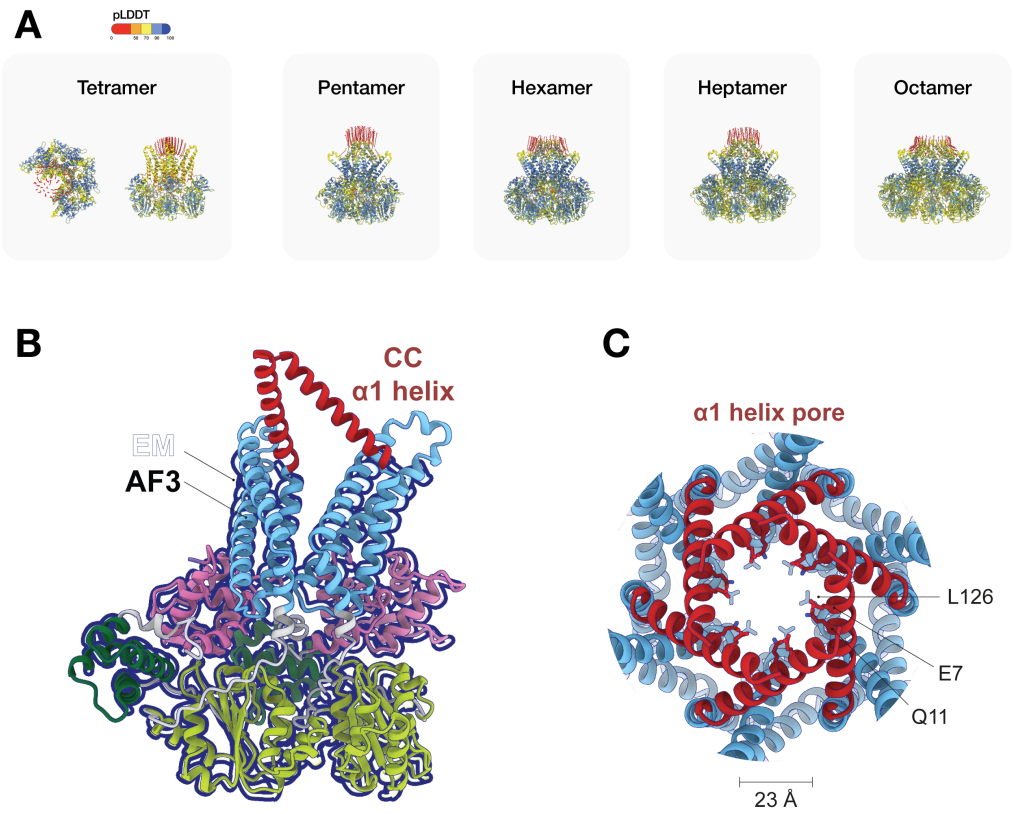


Figure 5.S3: NRC2 modeled across a range of stoichiometries and the predicted N-terminal pore.

(A) AlphaFold 3 models of NRC2 assembled as a tetramer, pentamer, hexamer, heptamer, and octamer, colored by pLDDT. Only the tetramer fails to form a closed, symmetric ring. (B) Superposition of a single protomer from the AlphaFold 3 model (AF3) and the cryo-EM model (EM), showing the position of the CC α1 helix that was unresolved in the experimental density. (C) Top view of the α1 helix bundle in the hexameric model, forming a pore of approximately 23 Å lined by residues including E7, Q11, and L126. This figure has also been published in Madhuprakash et al. (Madhuprakash et al., 2024) and appears here with permission.

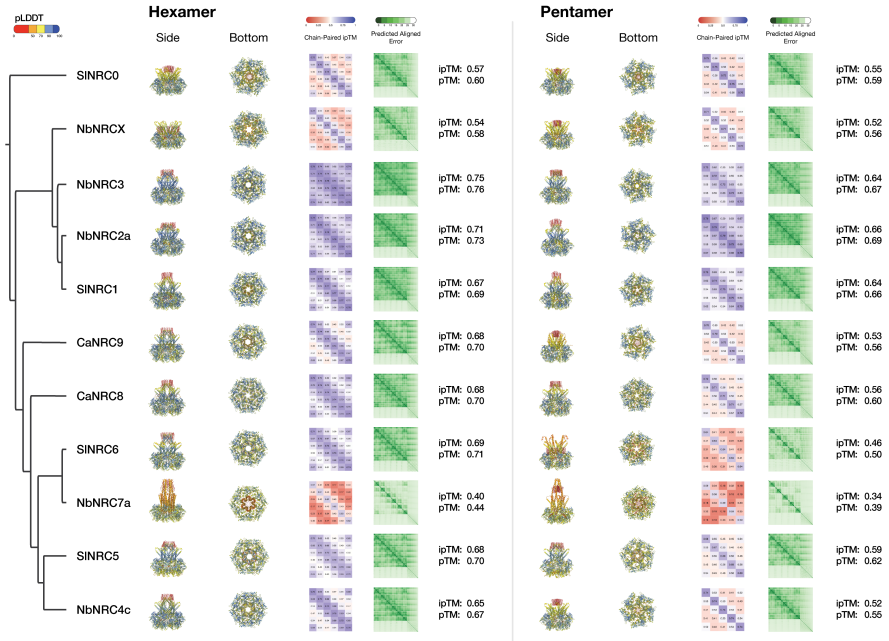


Figure 5.S4: AlphaFold 3 modeling of eleven NRC helpers as hexamers and pentamers.
For each helper, ordered by the phylogeny shown on the left, side and bottom views of the predicted assembly are shown colored by pLDDT, together with the chain-pair ipTM matrix, the predicted aligned error plot, and the overall ipTM and pTM values. Helpers from *Solanum lycopersicum* (Sl), *Nicotiana benthamiana* (Nb), and *Capsicum annuum* (Ca) are included. This figure has also been published in Madhuprakash et al. (Madhuprakash et al., 2024) and appears here with permission.

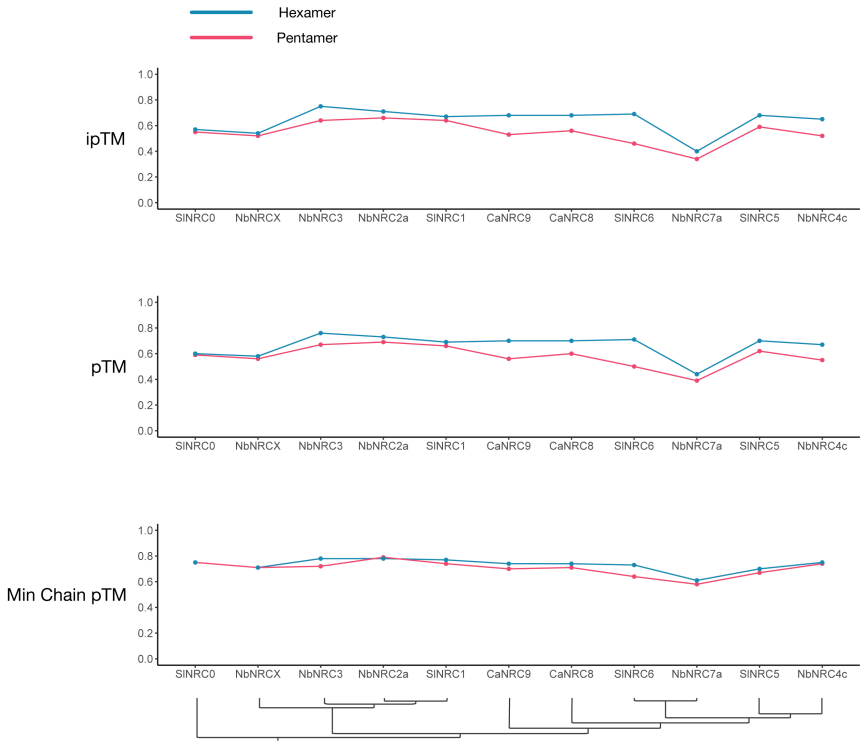


Figure 5.S5: Model confidence metrics for NRC helpers modeled as hexamers and pentamers. ipTM, pTM, and minimum per-chain pTM for each of the eleven NRC helpers, plotted in phylogenetic order and colored by modeled stoichiometry. All helpers score higher as hexamers than as pentamers; NbNRC7a returns the lowest values in every metric. This figure has also been published in Madhuprakash et al. (Madhuprakash et al., 2024) and appears here with permission.

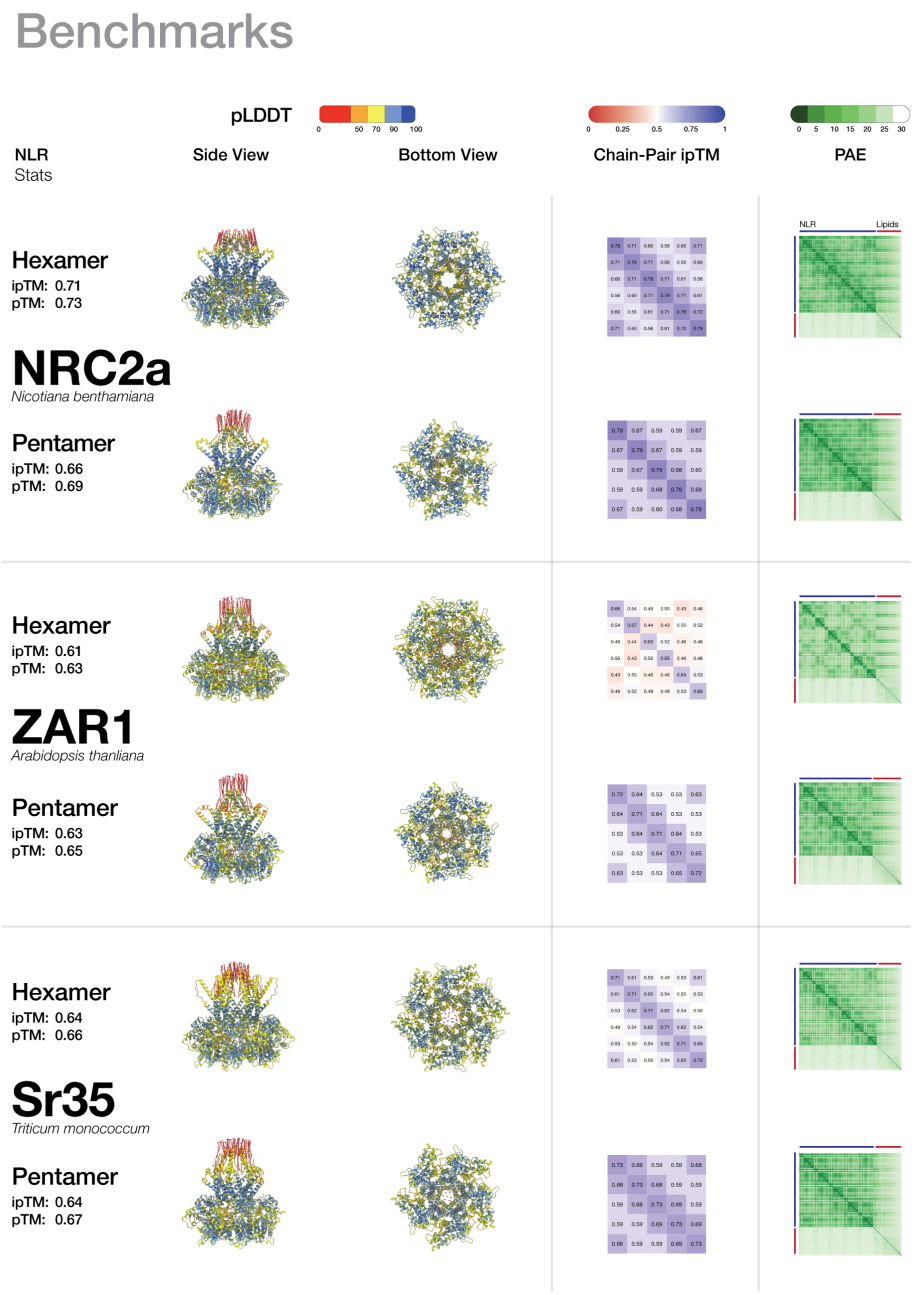


Figure 5.S6: Benchmarking hexameric and pentameric models of receptors with experimentally determined structures.
NRC2a (*Nicotiana benthamiana*), ZAR1 (*Arabidopsis thaliana*), and Sr35 (*Triticum monococcum*) were each modeled as both a hexamer and a pentamer. Side and bottom views colored by pLDDT are shown together with chain-pair ipTM matrices and predicted aligned error plots, in which the receptor (NLR) and lipid regions are indicated. Only NRC2a shows an appreciable preference for the hexameric state. This figure has also been published in Madhuprakash et al. (Madhuprakash et al., 2024) and appears here with permission.

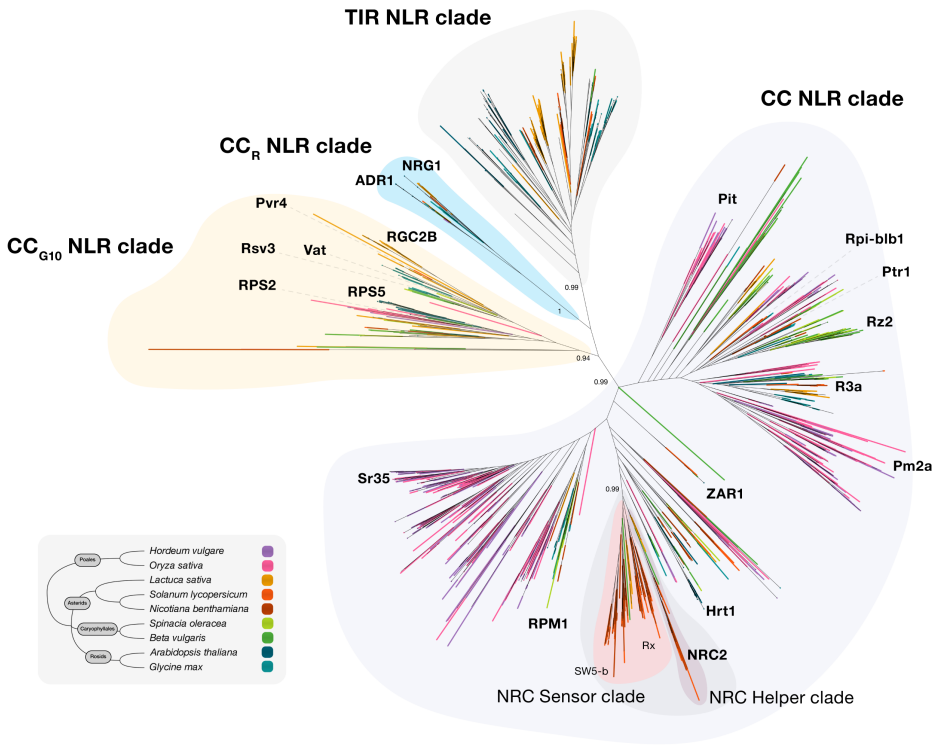


Figure 5.S7: Phylogenetic placement of the CC-type NLRs selected for structural modeling. Phylogenetic tree of plant NLRs with the TIR, CC, CC_R, and CC_{G10} clades indicated, and the NRC sensor and helper clades shown within the CC-NLR clade. Branches are colored by species, with species relationships given in the inset. The receptors modeled in this chapter are labeled, and node values indicate branch support. This figure has also been published in Madhuprakash et al. (Madhuprakash et al., 2024) and appears here with permission.

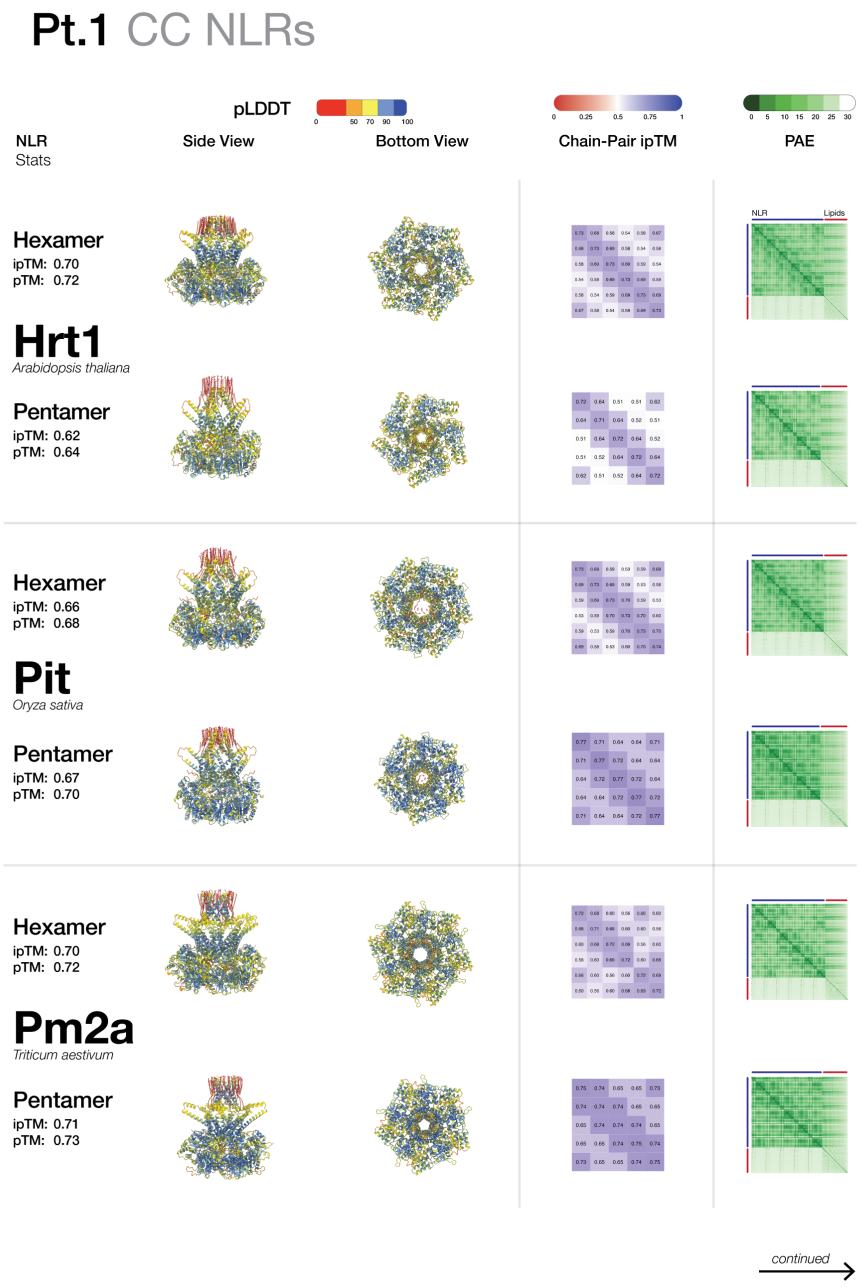


Figure 5.S8: Structural modeling of CC-NLRs (part 1).
Hexameric and pentameric AlphaFold 3 models of Hrt1 (*Arabidopsis thaliana*), Pit (*Oryza sativa*), and Pm2a (*Triticum aestivum*), each shown in side and bottom view colored by pLDDT, with chain-pair ipTM matrices, predicted aligned error plots, and overall ipTM and pTM values. This figure has also been published in Madhuprakash et al. (Madhuprakash et al., 2024) and appears here with permission.

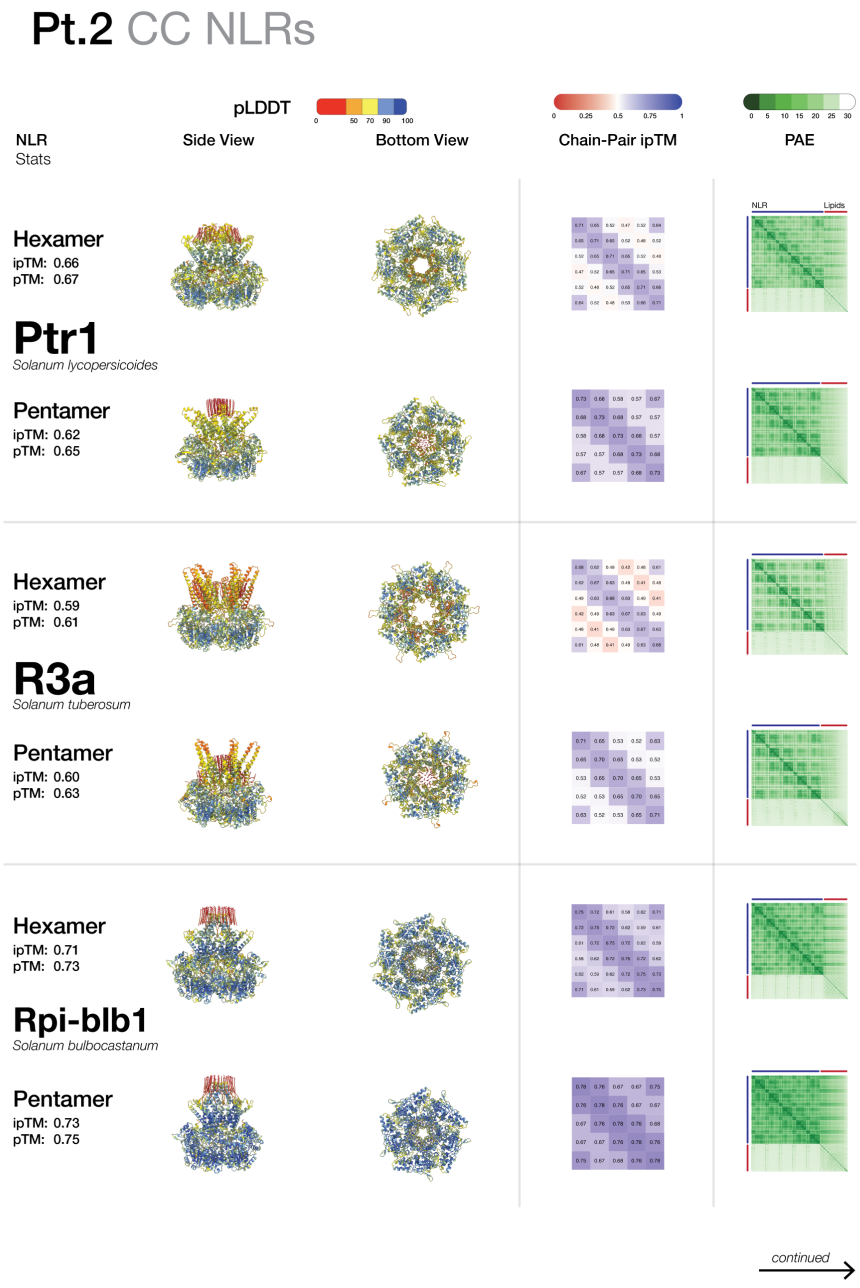


Figure 5.S9: Structural modeling of CC-NLRs (part 2).
Hexameric and pentameric AlphaFold 3 models of Ptr1 (*Solanum lycopersicoides*), R3a (*Solanum tuberosum*), and Rpi-blb1 (*Solanum bulbocastanum*), presented as in Figure 5.S8. This figure has also been published in Madhuprakash et al. (Madhuprakash et al., 2024) and appears here with permission.

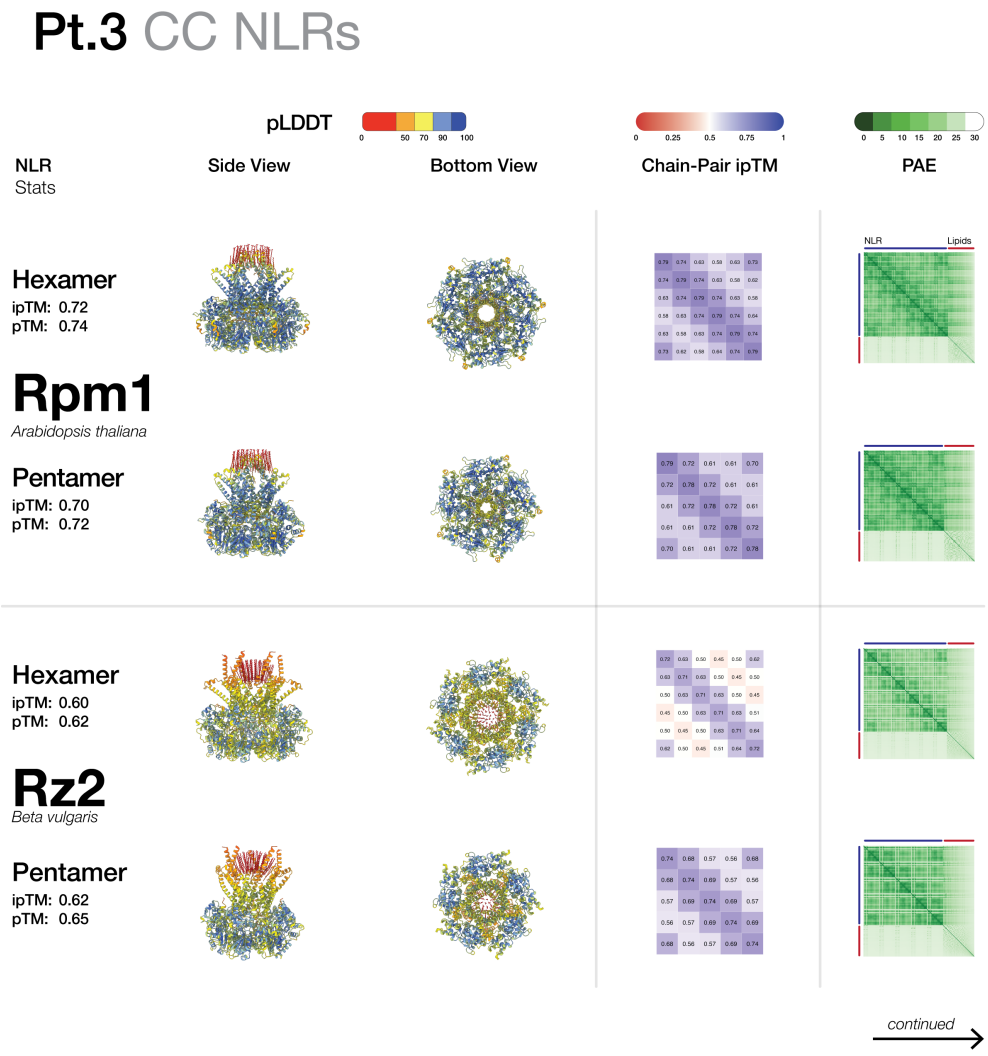


Figure 5.S10: Structural modeling of CC-NLRs (part 3).
Hexameric and pentameric AlphaFold 3 models of Rpm1 (*Arabidopsis thaliana*) and Rz2 (*Beta vulgaris*), presented as in Figure 5.S8. This figure has also been published in Madhuprakash et al. (Madhuprakash et al., 2024) and appears here with permission.

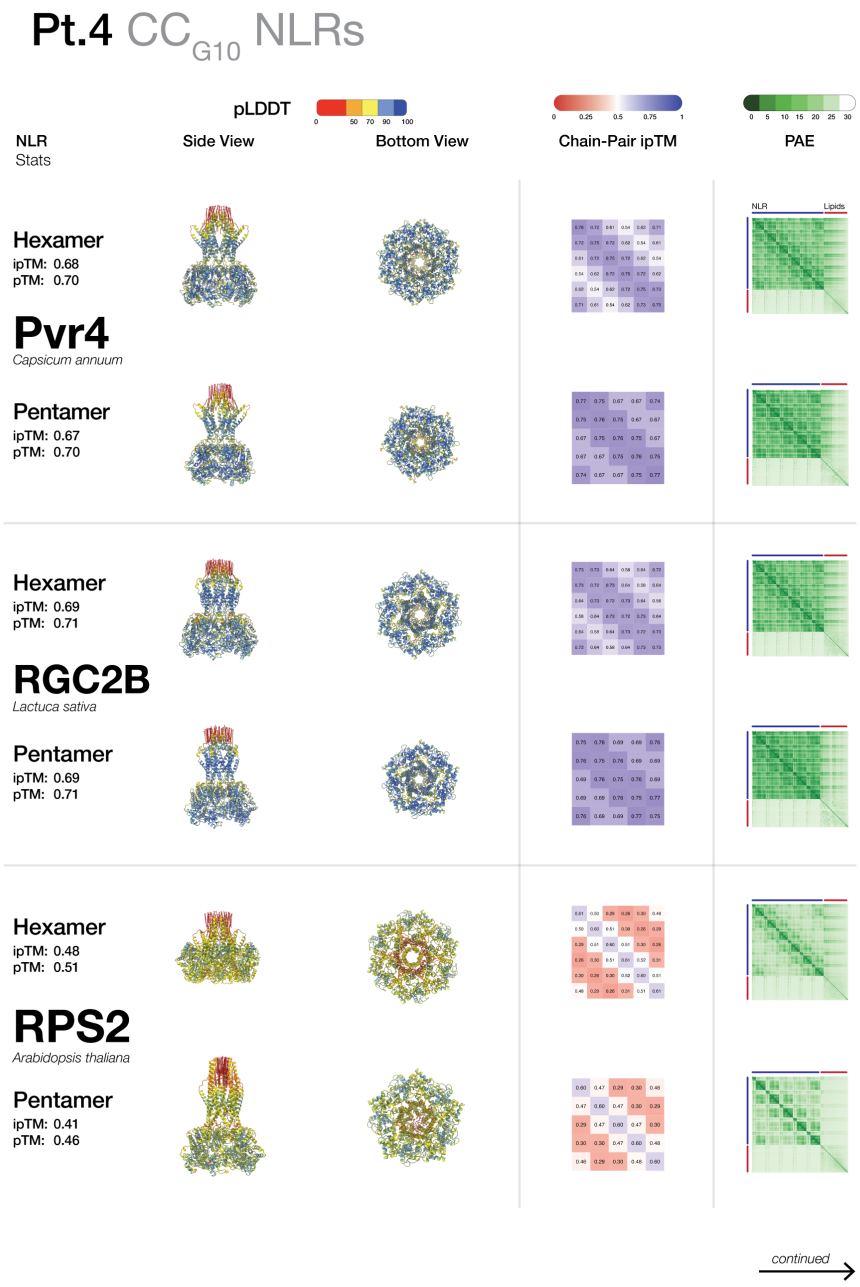


Figure 5.S11: Structural modeling of CC_{G10}-NLRs (part 1).
Hexameric and pentameric AlphaFold 3 models of Pvr4 (*Capsicum annuum*), RGC2B (*Lactuca sativa*), and RPS2 (*Arabidopsis thaliana*), presented as in Figure 5.S8. These receptors adopt a distinctly elongated, two-tiered architecture; RPS2 returns the lowest confidence values of the group. This figure has also been published in Madhuprakash et al. (Madhuprakash et al., 2024) and appears here with permission.

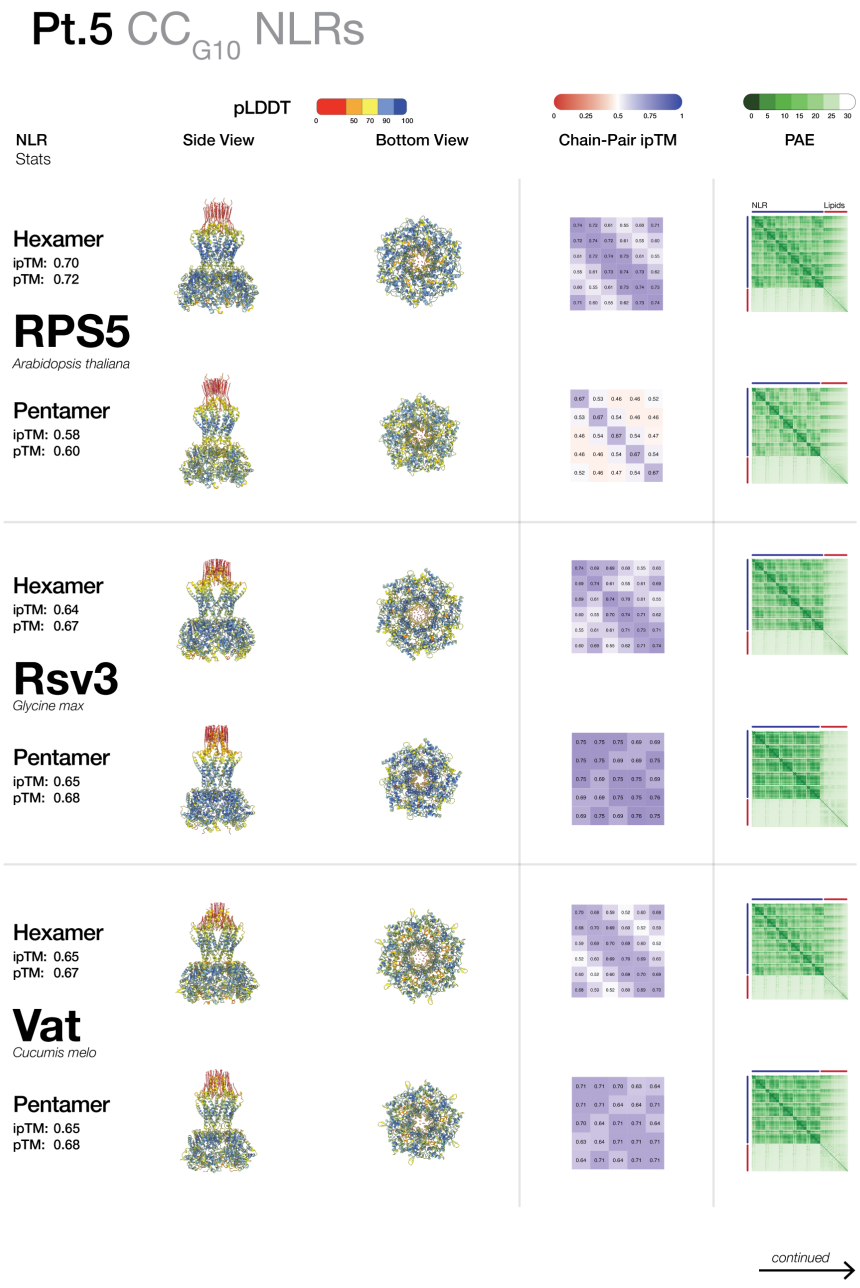


Figure 5.S12: Structural modeling of CC$_{G10}$-NLRs (part 2).
Hexameric and pentameric AlphaFold 3 models of RPS5 (*Arabidopsis thaliana*), Rsv3 (*Glycine max*), and Vat (*Cucumis melo*), presented as in Figure 5.S8. RPS5 shows a marked preference for the hexameric state. This figure has also been published in Madhuprakash et al. (Madhuprakash et al., 2024) and appears here with permission.

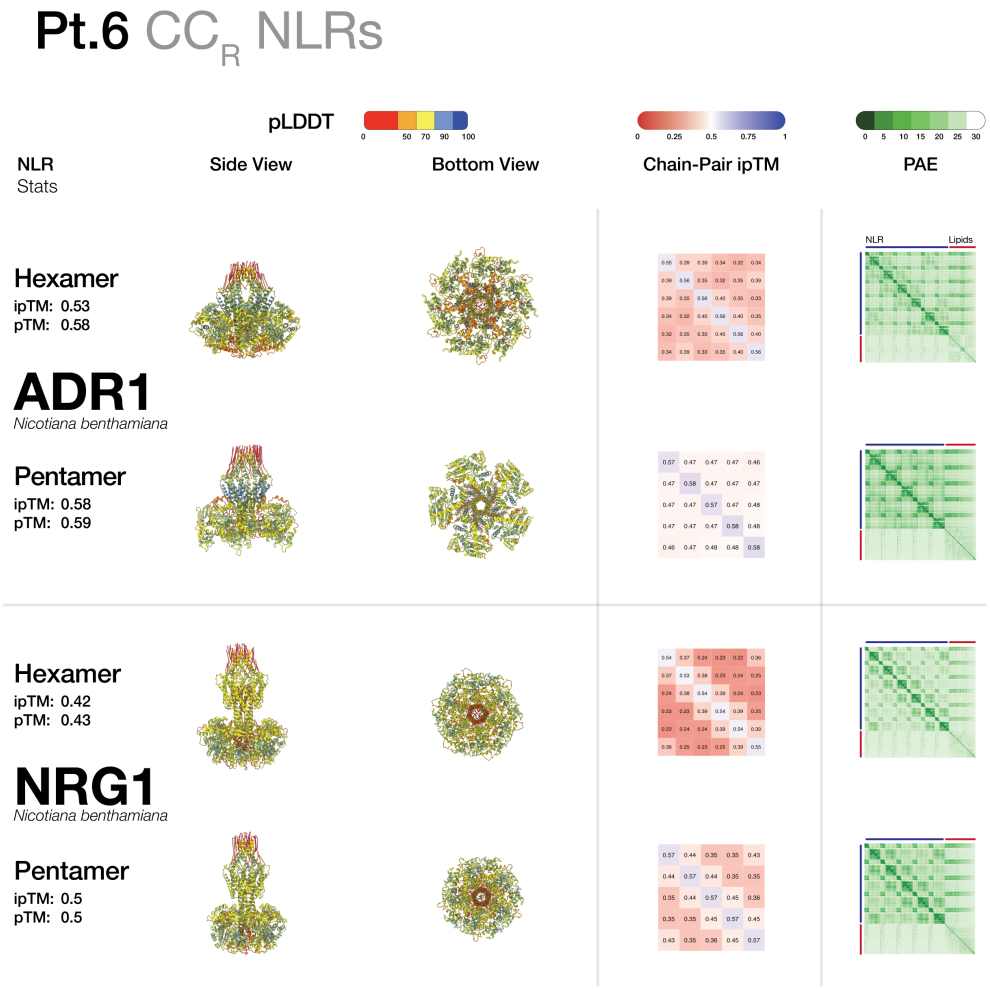


Figure 5.S13: Structural modeling of CC$_R$-NLRs.
Hexameric and pentameric AlphaFold 3 models of ADR1 and NRG1 from *Nicotiana benthamiana*, presented as in Figure 5.S8. Both receptors are modeled with low confidence in either stoichiometry, with chain-pair ipTM matrices dominated by low values. This figure has also been published in Madhuprakash et al. (Madhuprakash et al., 2024) and appears here with permission.

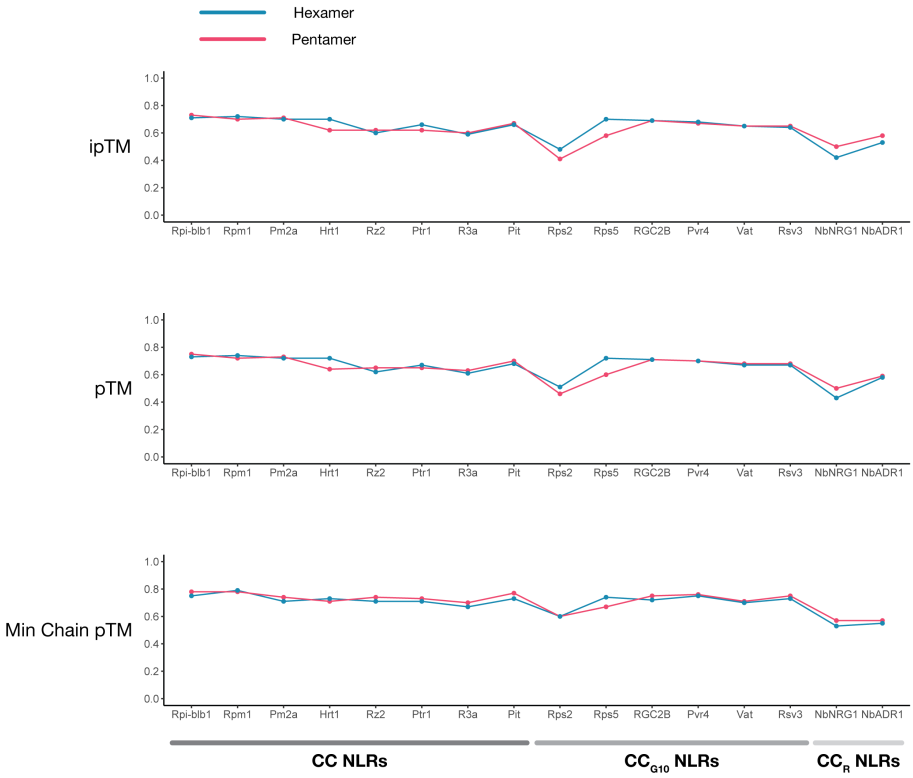


Figure 5.S14: Model confidence metrics for CC-type NLRs modeled as hexamers and pentamers.

ipTM, pTM, and minimum per-chain pTM for the sixteen modeled receptors, grouped by clade into CC, CC_{G10}, and CC_R NLRs and colored by modeled stoichiometry. For most receptors the two stoichiometries are barely distinguishable; RPS2 and NRG1 return the lowest values overall. This figure has also been published in Madhuprakash et al. (Madhuprakash et al., 2024) and appears here with permission.

Chapter 6: Using AI-based modeling to distinguish between Sensor and Helper NLRs

The work described in this chapter has been published in Toghani et al. (Toghani et al., 2025a) and appears here with permission. The analysis of the lettuce NRC network described at the end of the chapter has been published in Pai et al. (Pai et al., 2026) and likewise appears with permission.

6.1 Introduction

NLR receptor proteins are encoded in organisms across all kingdoms of life but are especially diversified in plants (Barragan and Weigel, 2021). Receptors carrying a coiled-coil (CC) domain at their N terminus are phylogenetically the most widespread class. Following pathogen recognition, plant CC-NLR receptors oligomerize into pentameric or hexameric pore-like complexes (Wang et al., 2019a; Förderer et al., 2022; Zhao et al., 2022; Liu et al., 2023a; Madhuprakash et al., 2024). These higher-order complexes, termed resistosomes which are described in Chapter 5, represent the defining feature of NLRs that execute immune responses. Upon activation and complex formation, CC-NLRs translocate to cellular membranes to form pores which allow calcium influx and subsequent cell death (Duggan et al., 2021; Contreras et al., 2022; Ibrahim et al., 2026). The prevailing model indicates that the funnel-shaped resistosome inserts into the membrane to execute cell death via the N-terminal domain (Wang et al., 2019a; Adachi et al., 2019b). The N-terminal α1 helix forms the tip of the funnel and represents a structurally dynamic region that remains difficult to resolve by cryo-EM (Förderer et al., 2022; Zhao et al., 2022; Liu et al., 2023a; Madhuprakash et al., 2024).

As outlined in Chapters 3 and 4, NLRs can function as singletons, pairs, or networks (Adachi et al., 2019b; Contreras et al., 2023a). While singletons can detect pathogens and execute cell death within a single protein, paired and networked NLR have subfunctionalized into pathogen-detecting sensors and downstream signaling helpers with distinct biochemical roles. Paired NLRs often originate from distinct phylogenetic clades yet function together, which makes it challenging to determine NLR pairs based on sequence phylogeny alone, as compared to NLR networks which often fall into neighboring phylogenetic clades (Kourelis et al., 2021; Contreras et al., 2023a).

Two sequence-based criteria are conventionally used to generally separate sensors from helpers. (I) Sensor and helper pairs often locate within shared genetic clusters. (II)

Sensors can carry non-canonical integrated domains (IDs) that act as pathogen baits which are absent from helpers (Białas et al., 2018; Marchal et al., 2018). The presence of an ID therefore provides a useful sequence diagnostic. Additionally, approximately 20% of plant CC-NLRs carry the conserved MADA motif in their N-terminal α1 helix, which this motif has degenerated in some Solanaceous sensor clades (Adachi et al., 2019a). Both criteria remain useful but incomplete, leaving the structural basis underlying sensor and helper specialization unresolved.

AlphaFold 3 might offer a computational strategy to address this gap (Abramson et al., 2024). As shown in Chapter 5, modeling protein complexes with oleic acid molecules simulates the membrane environment and resolves regions that are difficult to capture experimentally, including the N-terminal α1 helix and the resistosome funnel (Madhuprakash et al., 2024). In this chapter, I use AlphaFold 3 to evaluate the functional roles of genetically linked NLRs computationally. I test whether sensors and helpers exhibit distinct structural signatures in modeled oligomers and examine whether structural modeling provides a more robust classification framework than existing sequence-based criteria.

6.2 Helper NLRs produce higher confidence scores than their paired sensors when modeled as resistosomes

I assembled a curated set of CC-NLR pairs comprising experimentally validated pairs in rice (Pikm, Pii, and Pia), their orthologs (PIK5/6-NP, Pi5-3/1, and Pias), and two cloned pairs in barley (RPG5/HvRGA1 and RGH2/3) (Figure 6.S1). Following the approach used in Chapter 5 (Madhuprakash et al., 2024), I extracted the oligomerizing region of each protein from the N terminus through the NB-ARC domain. I then modeled pentameric and hexameric stoichiometries with 50 oleic acid molecules across three random seeds in AlphaFold 3.

Model confidence scores followed a clear, consistent pattern across both pentameric and hexameric configurations. Known helper NLRs yielded higher pTM scores than their respective paired sensor in every case (Figure 6.1a,b, and Figures 6.S2 and 6.S3). When plotting helper against sensor scores, every pair fell above the diagonal line, and ipTM scores showed the same distribution (Figure 6.S4), outlining that this method can be used to distinguish known sensor and helper pairs. For instance, RGH3 reached a pTM of 0.74 compared with 0.31 for its partner RGH2, and Pii-1 reached 0.71 against 0.54 for Pii-2. The score difference was narrowest for RPG5/HvRGA1, where hexameric pTM values were nearly identical (Figure 6.S3). Additionally, pentameric and hexameric scores correlated

strongly, as did pTM and ipTM within each stoichiometry (Figure 6.S2), showing that this pattern holds across different scoring metrics, outlining that modeling does not require detailed a priori knowledge of the real oligimeric state.

Upon closer evaluation, the modeled structures showed clear architectural differences between the two classes (Figure 6.1c, and Figures 6.S5 and 6.S6). Helper NLRs assembled into funnel-shaped oligomers that projected their N-terminal α1 helices outward contacting the lipids. In contrast, sensor NLRs failed to form these ordered funnels. This difference aligns with evidence that sensor NLRs cannot trigger immune signaling independently and may have lost the ability to assemble functional resistosomes (Adachi et al., 2019b; Contreras et al., 2022). While sensors occasionally assembled into ring-like structures, these models consistently returned low confidence scores, especially compared to their paired helper (Figure 6.1). AlphaFold 3 predictions therefore reliably distinguish the structural properties of helpers from those of paired sensors.

6.3 Structure-based classification outperforms the MADA motif as a determinant of sensor/helper NLR predictions in paired NLRs

I next tested whether the presence/absence of the MADA motif alone could be used to classify sensors and helpers of paired NLRs in this curated set using the established MADA hidden Markov model (HMM) (Adachi et al., 2019a). Only the helpers of Pikm and PIK5/6-NP paired NLR system contained a motif that exceeded the HMM score cutoff of 10, outlining the presence of a MADA motif. The remaining sensors and helpers lacked detectable motifs. Therefore, the used sequence analysis of the MADA motif only classified two of the eight helpers within the NLR pairs successfully. Because structural modeling separated all eight pairs into non-resistosome sensors and resistosome forming helpers, the structure-based metric provides a substantially more sensitive classification framework than sequence motifs alone.

6.4 Structure-based classification can be used to predict putative paired NLRs in rice

I next extended this structural analysis to predict putative paired NLRs in rice. Taking the rice pan-genome analysis of Stein and colleagues (Stein et al., 2018), I extracted ten rice CC-NLR pairs that are (I) genetically linked in head-to-head orientation, (II) belong

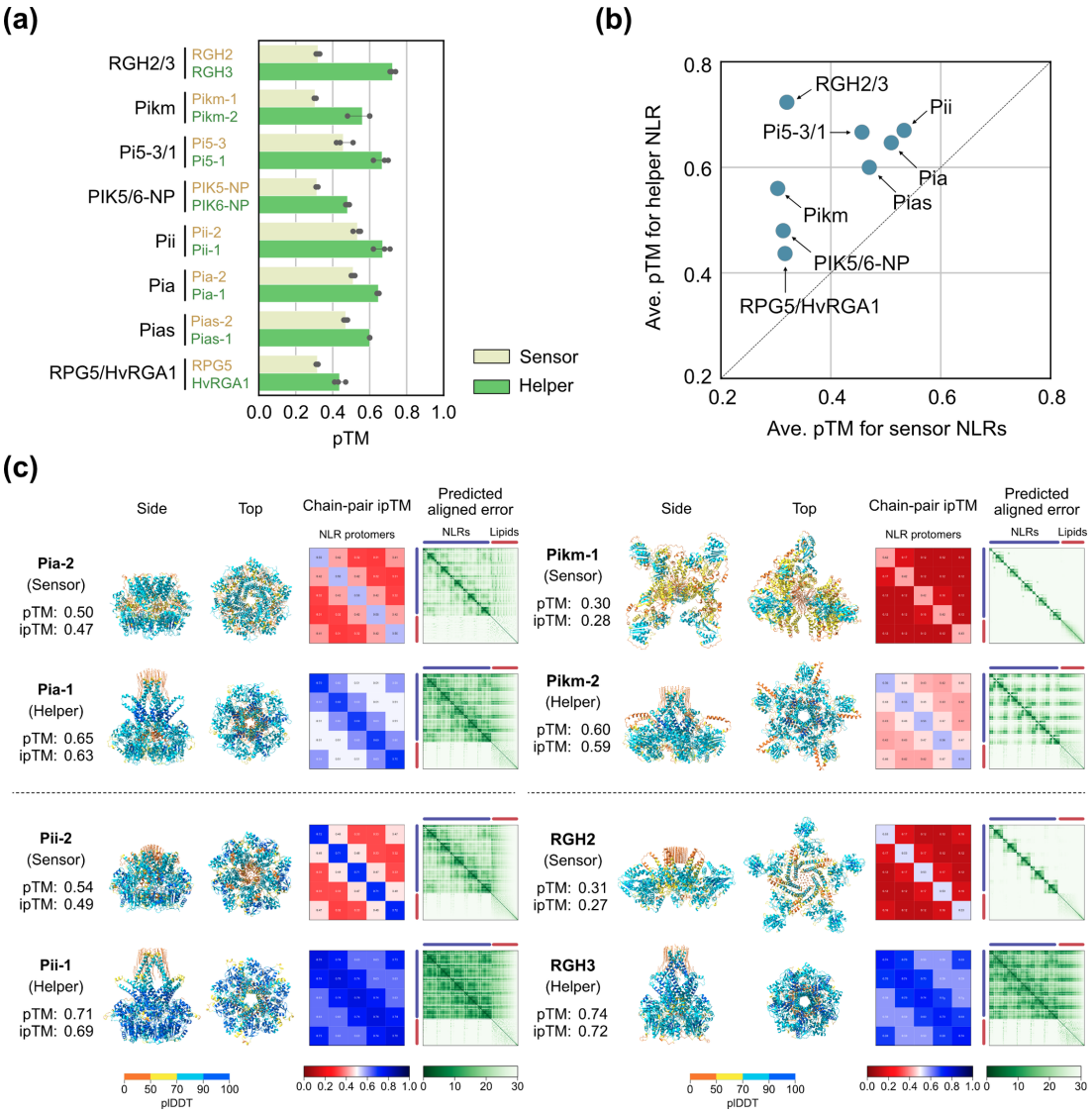


Figure 6.1: Helper NLRs produce higher AlphaFold 3 (AF3) confidence scores than their paired sensors.

(a) Bar plot comparing sensor and helper pTM scores in pentameric AF3 predictions. The amino acid sequences of the oligomerizing domains, from the N terminus to the end of the NB-ARC domain, were used for modeling. Pentameric structures were modeled with 50 oleic acid molecules using three different seed values. (b) Scatter plot comparing sensor and helper pTM scores in pentameric AF3 predictions, averaged across the three seeds. (c) Pentameric AF3 structures of four representative NLR pairs, shown in side and top views colored by pLDDT, with chain-pair ipTM matrices and predicted aligned error plots. Structures predicted using seed value 1 are shown. This figure has also been published in Toghani et al. (Toghani et al., 2025a) and appears here with permission.

to distinct phylogenetic clades, and (III) carry a complete N-terminal CC domain (Figure 6.2a and Figure 6.S1). In five of these ten pairs, one partner carries an ID annotation, identifying it as the putative sensor.

For these five pairs, putative helpers lacking an ID annotation returned higher pTM and ipTM scores than the respective putative sensors, which carry ID annotations (Figure 6.2b,c, and Figures 6.S7–6.S9). These results suggest that AlphaFold 3 can reliably predict sensor and helper identity in uncharacterized pairs, similar to predictions obtained for experimentally validated pairs.

Pair-13 specifically highlights the predictive power of this approach. This pair corresponds to PIK5/6-NP, where PIK5-NP functions as a sensor carrying an integrated domain that escapes detection by InterProScan (Białas et al., 2021; Kourelis et al., 2021). Despite the lack of an annotated ID, AlphaFold 3 prediction correctly identified PIK5-NP as the sensor based on low confidence scores alone. Structural modeling can therefore overcome false negatives in sequence-based domain annotations.

6.5 Validation on an experimentally characterized wheat NLR pair

After the initial submission of this study, two genetically linked NLR genes in wheat, *Pm5e* and *RXL* (*Rx*-CC-like), arranged head-to-head, were reported to function together as a paired CC-NLR module to confer resistance to powdery mildew (Guo et al., 2025). This pair provides a rigorous test case because *Pm5e* lacks an annotated ID and both proteins display non-canonical domain architectures, including an atypical CC domain in Pm5e and a truncated NB-ARC domain in RXL.

I tested whether AlphaFold 3 could be used to distinguish between sensor and helper NLR in this newly characterized pair. Consistent with previous predictions, the experimentally validated helper RXL scored higher in both pTM and ipTM across pentameric and hexameric models compared to the validated sensor Pm5e. This result demonstrates that structural modeling successfully predicts helper and sensor pairs that lack distinguishing sequence motifs or canonical domain features.

6.6 Extension to a dicot NLR network

To test whether this structural dichotomy extends beyond monocot pairs into dicot networks, I used AlphaFold 3 to model and predict sensor and helper NLRs in the lettuce NRC

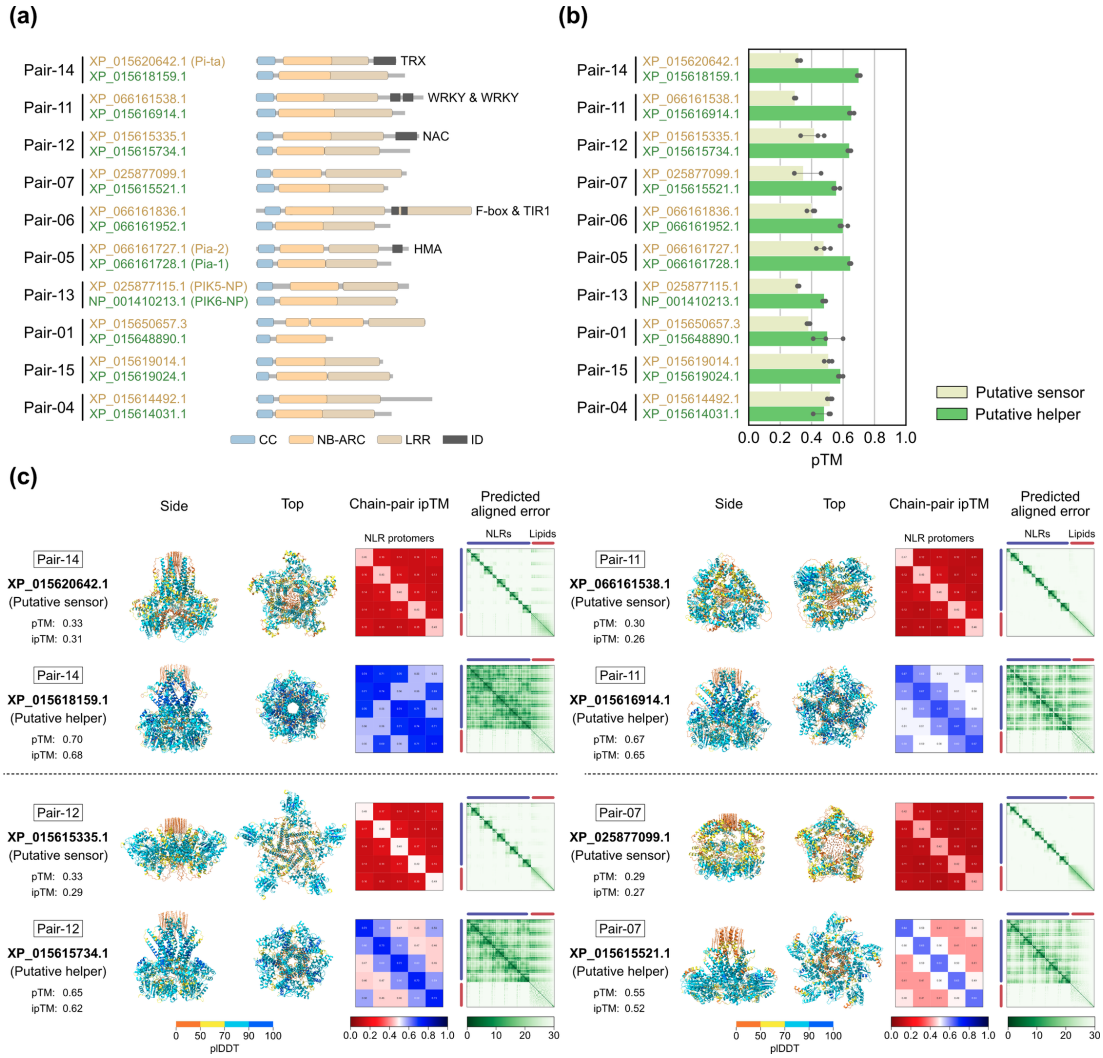


Figure 6.2: NLRs without an integrated domain (ID) exhibit higher AlphaFold 3 (AF3) confidence scores than NLRs with an ID in rice putative paired NLRs.

(a) Domain architectures of ten rice NLR pairs described by Stein and colleagues (Stein et al., 2018), annotated with NLRtracker (Kourelis et al., 2021). (b) Bar plot comparing putative sensor and helper pTM scores in pentameric AF3 predictions, modeled as in Figure 6.1. Putative sensors and helpers were assigned on the basis of the average pTM scores of pentameric and hexameric structures, with the lower-scoring member assigned as the putative sensor. (c) Pentameric AF3 structures of four representative pairs, predicted using seed value 1. This figure has also been published in Toghani et al. (Toghani et al., 2025a) and appears here with permission.

network characterized in Chapter 4 (Pai et al., 2026). I modeled each of the two lettuce helpers and the 16 NRC-S as hexamers across five independent replicates with 50 oleic acid molecules. Subsequently, the assemblies were evaluated using pLDDT, pTM, ipTM, and per-chain pTM.

Both helpers assembled into resistosome-like oligomers, though LsatNRC0 failed to form a complete N-terminal funnel, mirroring the behavior of Solanaceous NRC0 in Chapter 5 (Selvaraj et al., 2024). None of the NRC-S across all three phylogenetic groups formed a resistosome assembly (Figure 6.3). In accordance with that, sensor models returned consistently low confidence scores, with high predicted aligned error, pTM and ipTM values below 0.5, and per-chain pTM values below 0.7 (Figure 6.S10). In contrast, the helpers scored above 0.5 for both pTM and ipTM in every replicate, with per-chain pTM averaging 0.75 (Figure 6.3).

The structural distinction observed in monocot pairs therefore extends to dicot networks where NRC-S lack integrated domains. Importantly, cell death assays in Chapter 4 experimentally validated these functional relationships. In those assays, autoactive sensor mutants did not trigger cell death alone, but activated wild-type helpers to execute cell death. AlphaFold 3 predictions are able to directly capture this functional asymmetry.

6.7 Conclusions and discussion

The lower confidence scores for oligomeric resistosome predictions observed for sensor and higher scores observed for helper NLRs within NLR pairs align with evolutionary models of NLR subfunctionalization (Adachi et al., 2019b). In this model, ancestral singletons capable of both perception and execution subfunctionalized into specialized sensors and helpers (Adachi et al., 2019b; Contreras et al., 2023a). Sensors lost execution activity while helpers lost pathogen detection. Because assembling a funnel-shaped resistosome, followed by integration into the membrane to form a channel is required to trigger cell death, the retention of stable resistosome funnels in helpers and their loss in sensors structurally further supports this division of labor. Given that AlphaFold 3 failed to model these properties for sensor NLRs, these findings further suggest that sensor NLRs have lost the intrinsic capacity to assemble into autonomous resistosomes, consistent with experimental observations for the Solanaceous sensor Rx (Contreras et al., 2022).

Notably, the helpers Pia-1, Pias-1, and RGH3 lack detectable MADA motifs yet assembled outward-facing funnels in the AlphaFold models, when modeled as penta- or hexamers. This raises the possibility that alternative sequence patterns drive membrane insertion in these non-MADA helpers. Structural determination by cryo-EM combined with large-

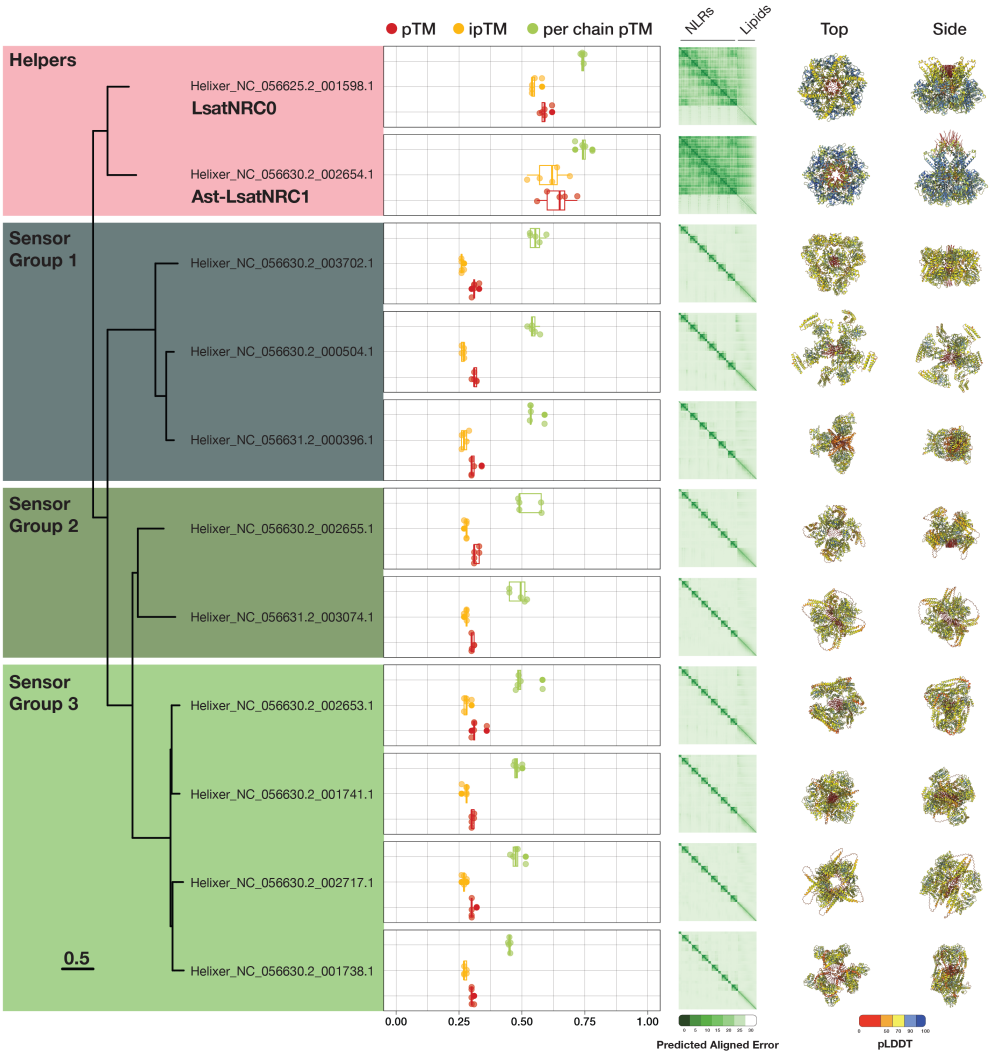


Figure 6.3: In contrast to NRC helpers, NRC-S failed to form resistosome-like structures when modeled with AlphaFold 3.

Lettuce NRC-H and NRC-S were modeled as hexamers including 50 oleic acid molecules to approximate the plasma membrane. Structural metrics, pLDDT (per-residue local confidence), pTM, ipTM, and per-chain pTM, are plotted for each protein ($n = 5$). Representative predicted structures from the fourth replicate are shown in top and side views, colored by pLDDT. This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

scale modeling will be required to define the motifs driving funnel formation in these divergent helpers.

This computational framework has three primary limitations. First, these generated models focus exclusively on active resistosome assemblies and do not evaluate resting states, which range from monomers and homodimers to higher-order oligomers as shown in Chapter 3 (Wang et al., 2019a; Selvaraj et al., 2024). Second, unlike helper NLRs we still do not have any structural evidence for how sensor NLRs, especially NRC-S, look like in their resting or activated states. Third, these simulations test homo-oligomers and do not assess whether paired NLRs assemble into hetero-complexes. The wheat pair RXL/Pm5e highlights this consideration, as these proteins preferentially form hetero-complexes (Guo et al., 2025) that homo-oligomeric modeling cannot capture.

The primary aim of this work was to evaluate AlphaFold as a predictive tool for classifying NLRs into functional sensors and helpers, which are highly diverse, often number in the hundreds per plant genome, and collectively account for approximately 1% of all plant genes, as quantified in Chapter 2 (Toghani and Kamoun, 2024). Combined with the evolutionary analyses in Chapters 3 and 4 and the structural modeling in Chapter 5, these results show that structural prediction can characterize NLR function within genetically paired NLR clusters and prioritize candidate sensor and helper genes for experimental validation. Such structural insights will help elucidate how the distinct architectures of sensors and helpers drive their functional specialization. However, the molecular determinants of sensor NLR emergence remains elusive. Moreover, given the reliance of AlphaFold 3 on multiple sequence alignments and evolutionary information, the differences in the oligomer prediction of sensor and helper NLRs suggests that AlphaFold 3 is detecting a signal in the amino acid sequence of these proteins. This may help uncover the molecular determinants of sensor-NLR evolution in future studies.

6.8 Supplementary figures

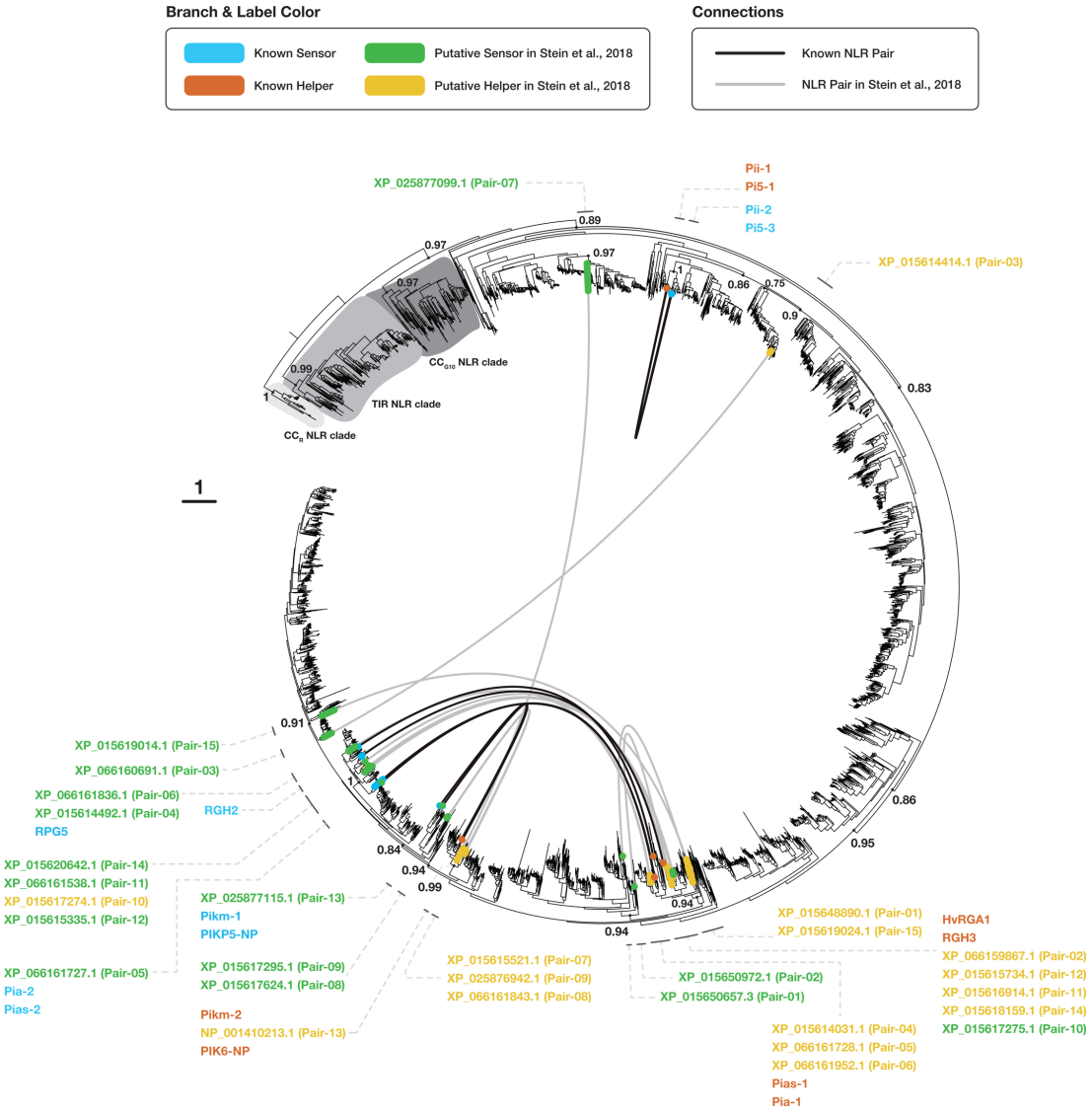


Figure 6.S1: Phylogenetic tree of the NLRs used for structural modeling.
Circular phylogenetic tree of NLRs with branches and labels colored by category: known sensors, known helpers, and the putative sensors and helpers of the rice pairs described by Stein and colleagues (Stein et al., 2018). Chords across the center connect the members of each pair, distinguishing experimentally characterized pairs from those proposed by Stein and colleagues. The TIR-NLR, CC$_R$-NLR, and CC$_{G10}$-NLR clades are indicated, and node values give branch support. This figure has also been published in Toghani et al. (Toghani et al., 2025a) and appears here with permission.

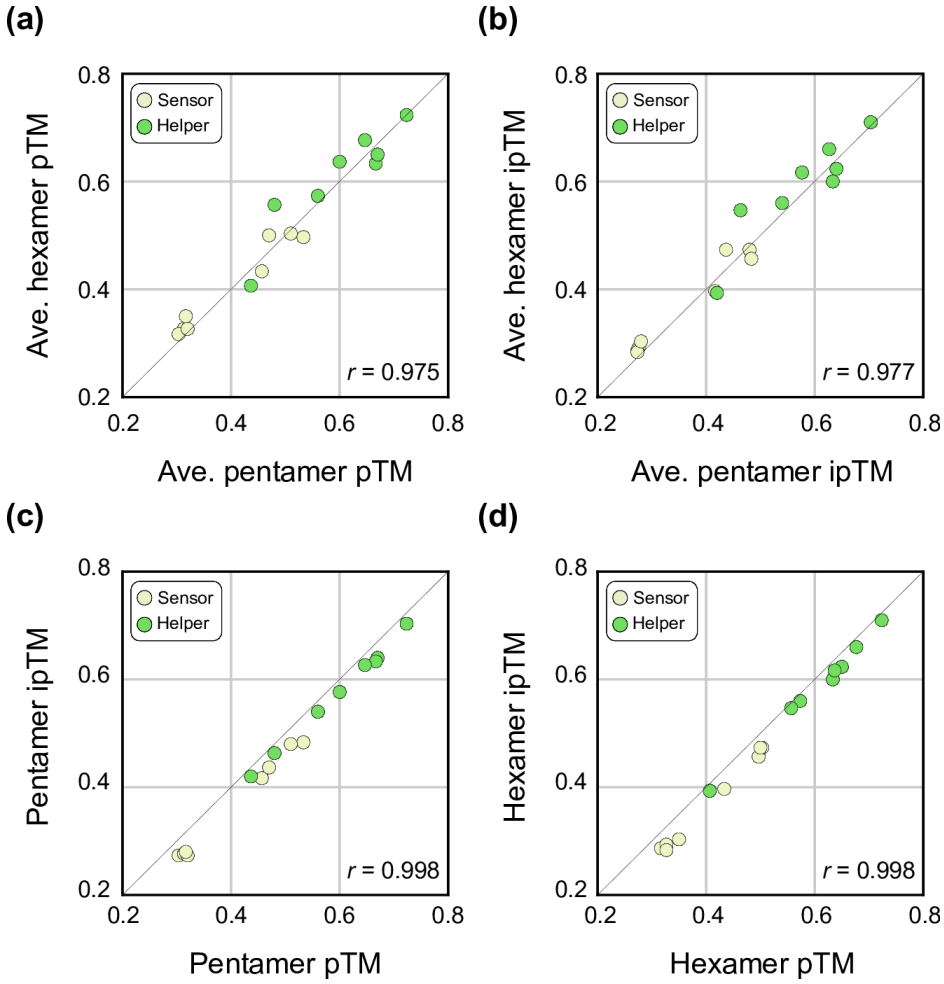


Figure 6.S2: Correlation between AlphaFold 3 confidence metrics across stoichiometries. (a) Average hexameric against average pentameric pTM. (b) Average hexameric against average pentameric ipTM. (c) Pentameric ipTM against pentameric pTM. (d) Hexameric ipTM against hexameric pTM. Points are colored by sensor or helper identity, the diagonal indicates equality, and Pearson correlation coefficients are given in each panel. This figure has also been published in Toghani et al. (Toghani et al., 2025a) and appears here with permission.

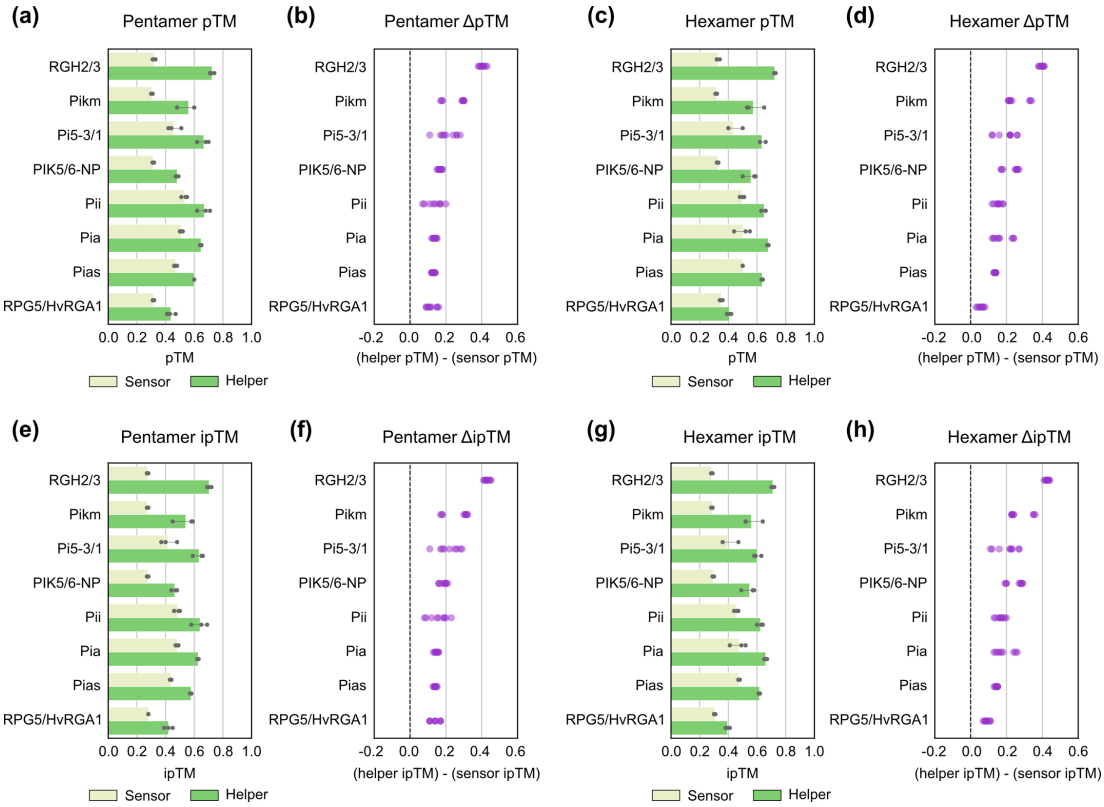


Figure 6.S3: Per-replicate confidence scores and sensor–helper differences for the eight characterized NLR pairs.

(a, c, e, g) Paired bar plots of sensor and helper pTM (a, c) and ipTM (e, g) for pentameric (a, e) and hexameric (c, g) predictions, with individual replicate values overlaid. (b, d, f, h) Corresponding helper-minus-sensor differences, with the dashed line marking zero. The difference is positive for every pair except RPG5/HvRGA1 in the hexameric pTM comparison, where it approaches zero. This figure has also been published in Toghani et al. (Toghani et al., 2025a) and appears here with permission.

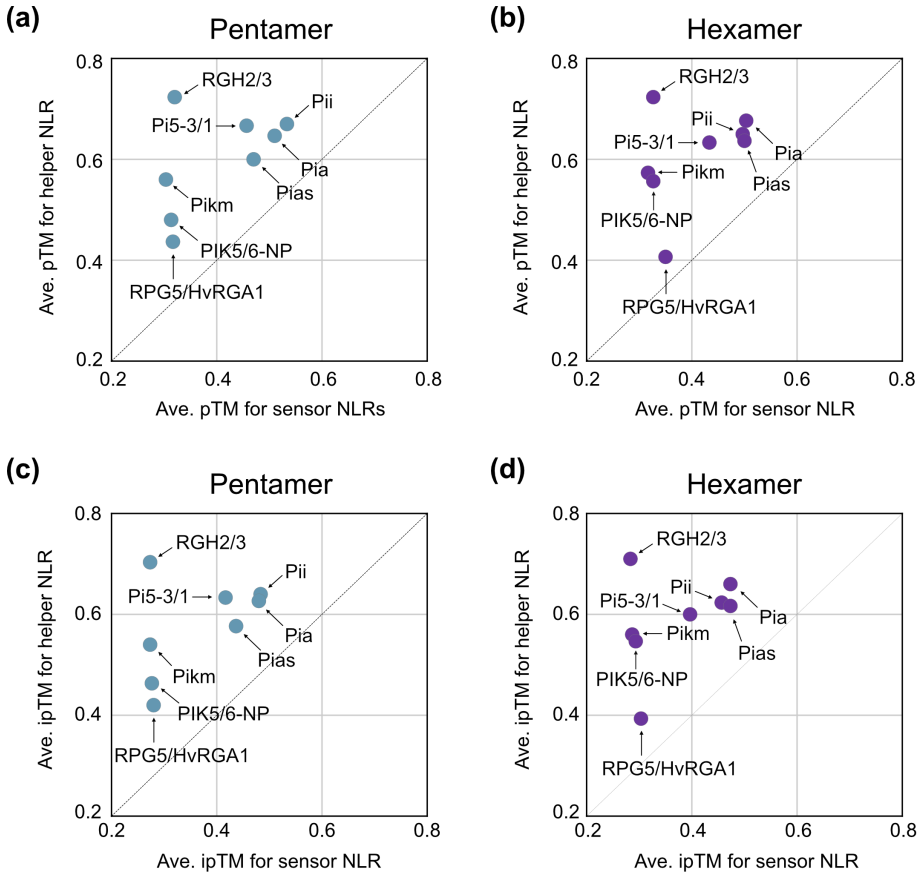


Figure 6.S4: Helper against sensor confidence scores for the eight characterized NLR pairs. (a, b) Average pTM for the helper plotted against the average pTM for the sensor, for pentameric (a) and hexameric (b) predictions. (c, d) The equivalent comparison for ipTM. All eight pairs fall above the diagonal in every comparison. This figure has also been published in Toghani et al. (Toghani et al., 2025a) and appears here with permission.

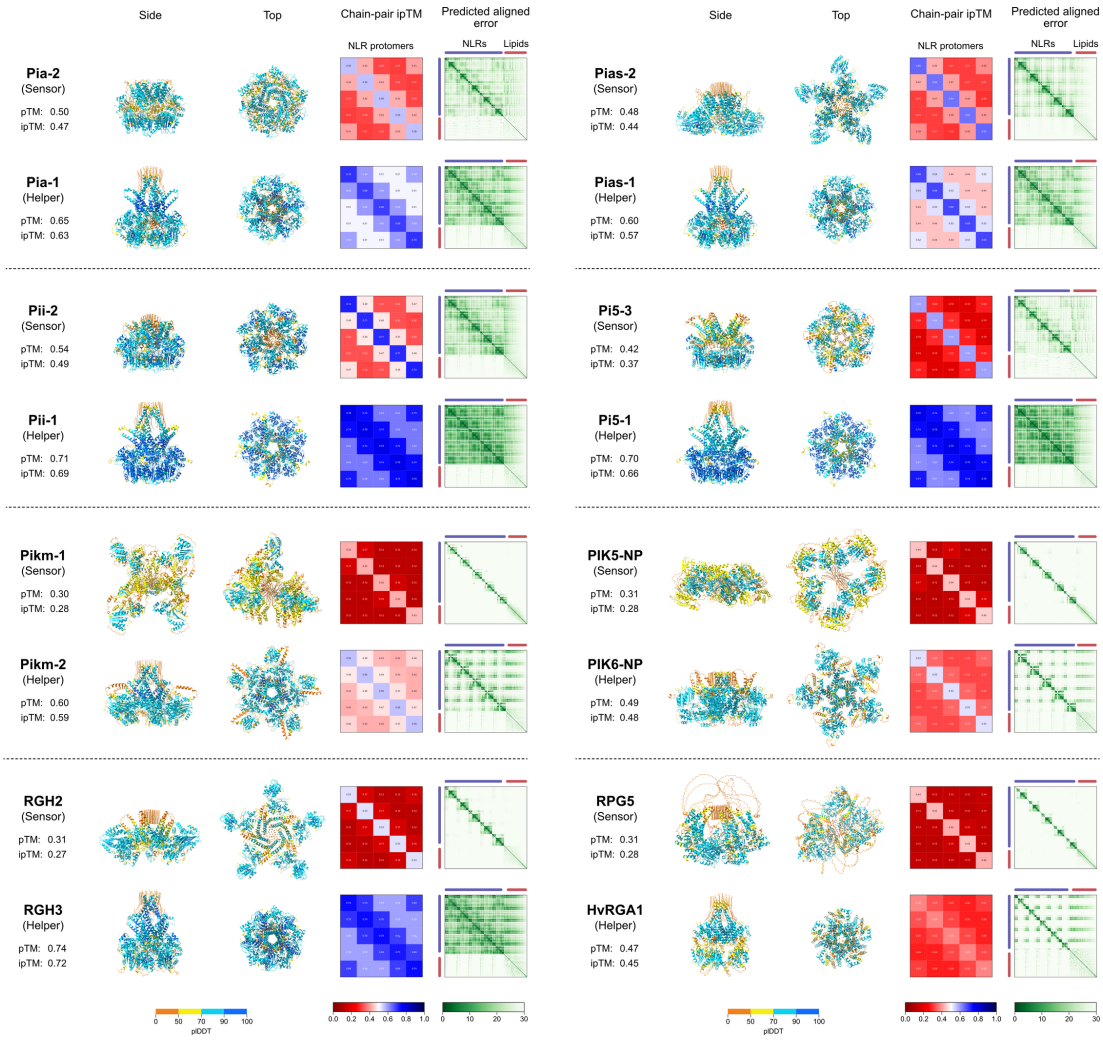


Figure 6.S5: Pentameric AlphaFold 3 models of the eight characterized sensor–helper NLR pairs.

For each protein, side and top views of the predicted pentamer are shown colored by pLDDT, together with the chain-pair ipTM matrix and the predicted aligned error plot, in which the receptor (NLRs) and lipid regions are indicated. Sensors are shown above their paired helpers. Helpers consistently expose the N-terminal $\alpha 1$ helices as a funnel, whereas sensors do not. This figure has also been published in Toghani et al. (Toghani et al., 2025a) and appears here with permission.

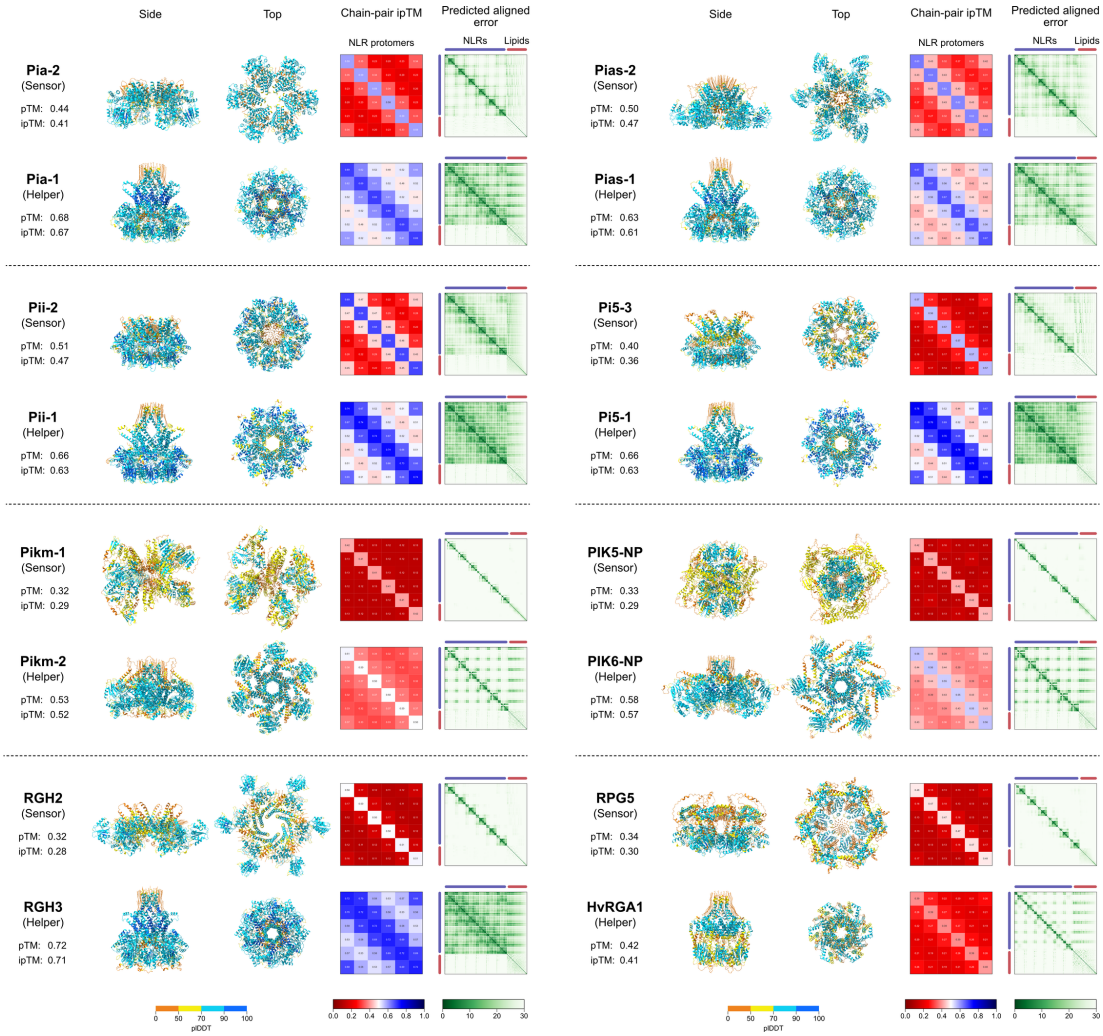


Figure 6.S6: Hexameric AlphaFold 3 models of the eight characterized sensor–helper NLR pairs.

Presented as in Figure 6.S5, with chain-pair ipTM matrices of six protomers. The same sensor–helper asymmetry in confidence and in funnel formation is observed as for the pentameric predictions. This figure has also been published in Toghani et al. (Toghani et al., 2025a) and appears here with permission.

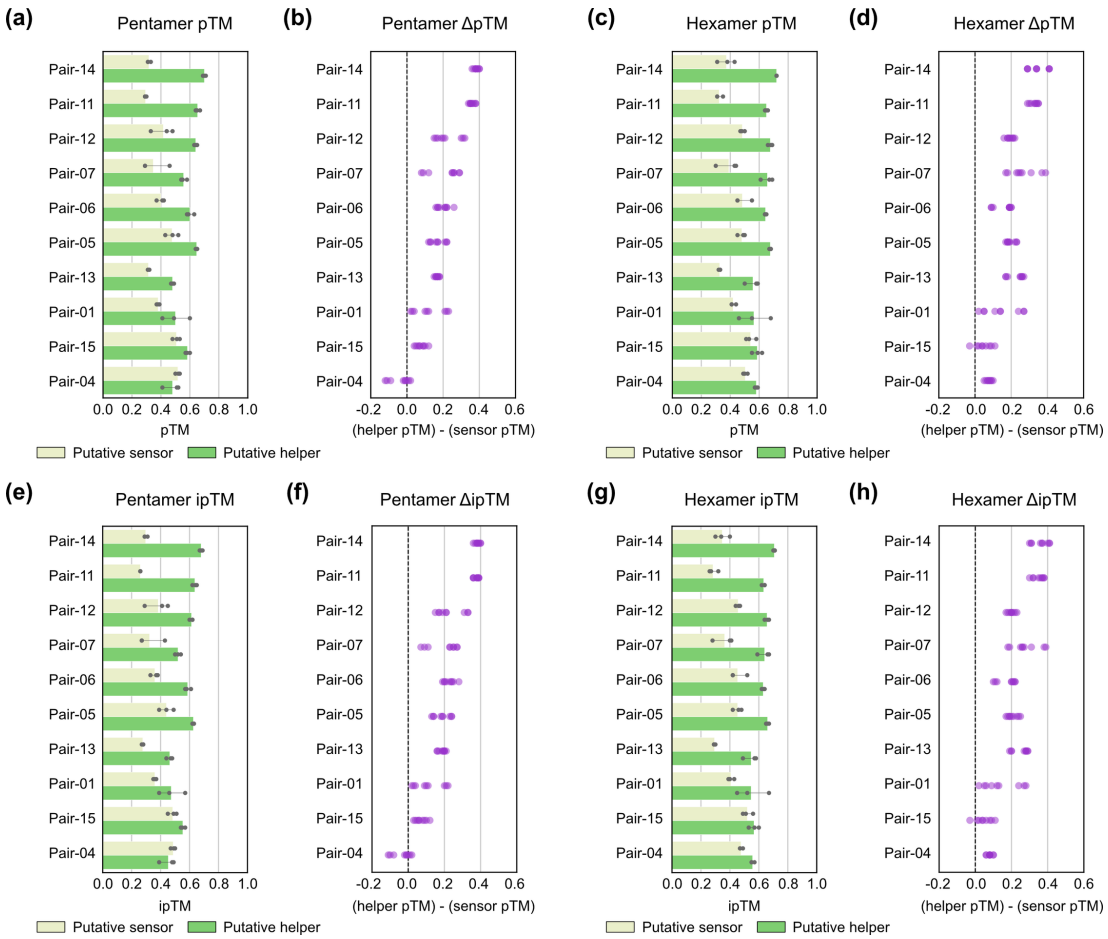


Figure 6.S7: Per-replicate confidence scores and differences for the ten rice putative NLR pairs.

(a, c, e, g) Paired bar plots of putative sensor and putative helper pTM (a, c) and ipTM (e, g) for pentameric (a, e) and hexameric (c, g) predictions, with individual replicate values overlaid. (b, d, f, h) Corresponding helper-minus-sensor differences, with the dashed line marking zero. This figure has also been published in Toghani et al. (Toghani et al., 2025a) and appears here with permission.

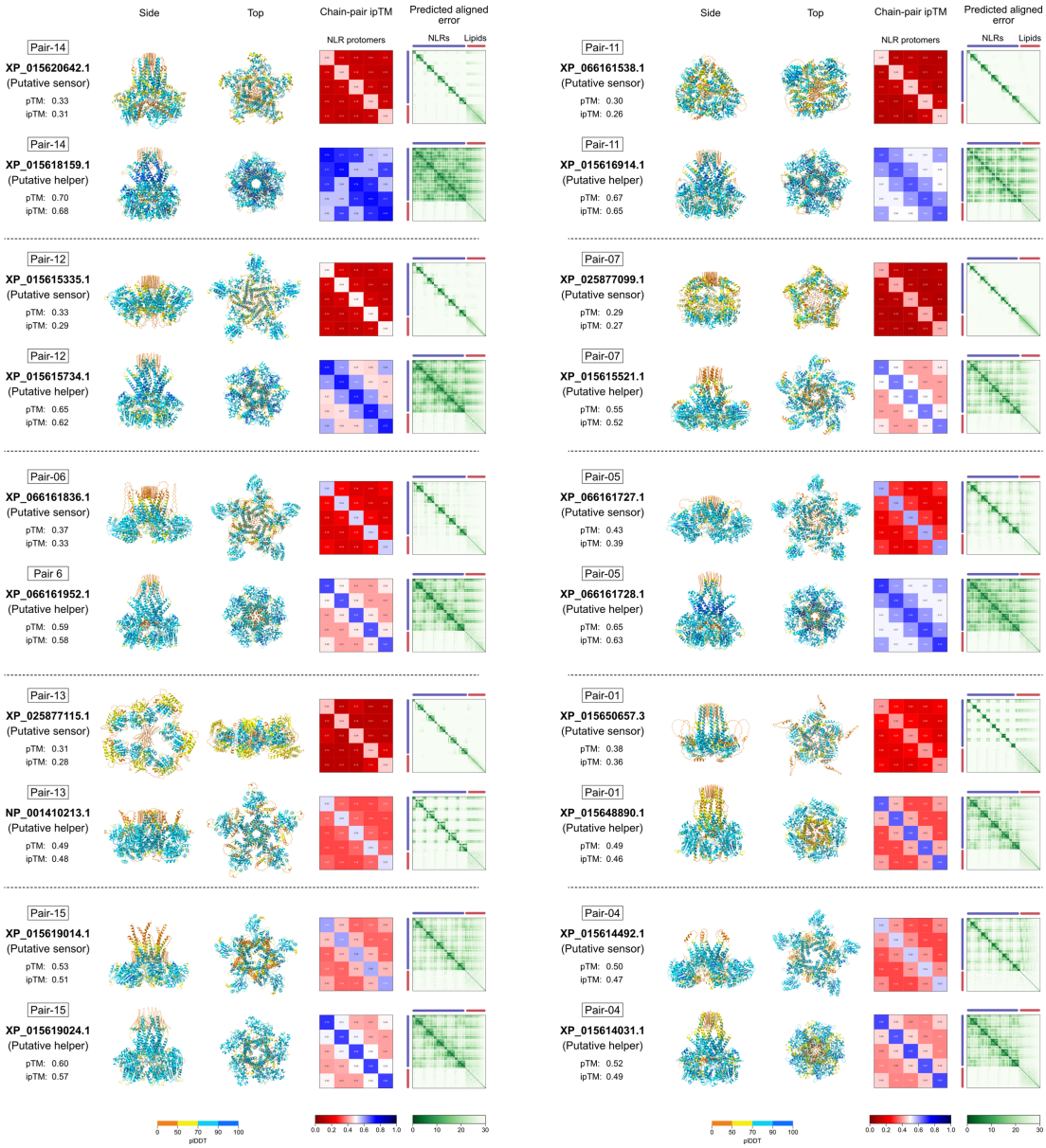


Figure 6.S8: Pentameric AlphaFold 3 models of the ten rice putative NLR pairs.
For each protein, identified by its accession, side and top views of the predicted pentamer are shown colored by pLDDT, together with the chain-pair ipTM matrix and the predicted aligned error plot. Putative sensors are shown above their putative helpers. This figure has also been published in Toghani et al. (Toghani et al., 2025a) and appears here with permission.



Figure 6.S9: Hexameric AlphaFold 3 models of the ten rice putative NLR pairs.
Presented as in Figure 6.S8, with chain-pair ipTM matrices of six protomers. This figure has also been published in Toghani et al. (Toghani et al., 2025a) and appears here with permission.

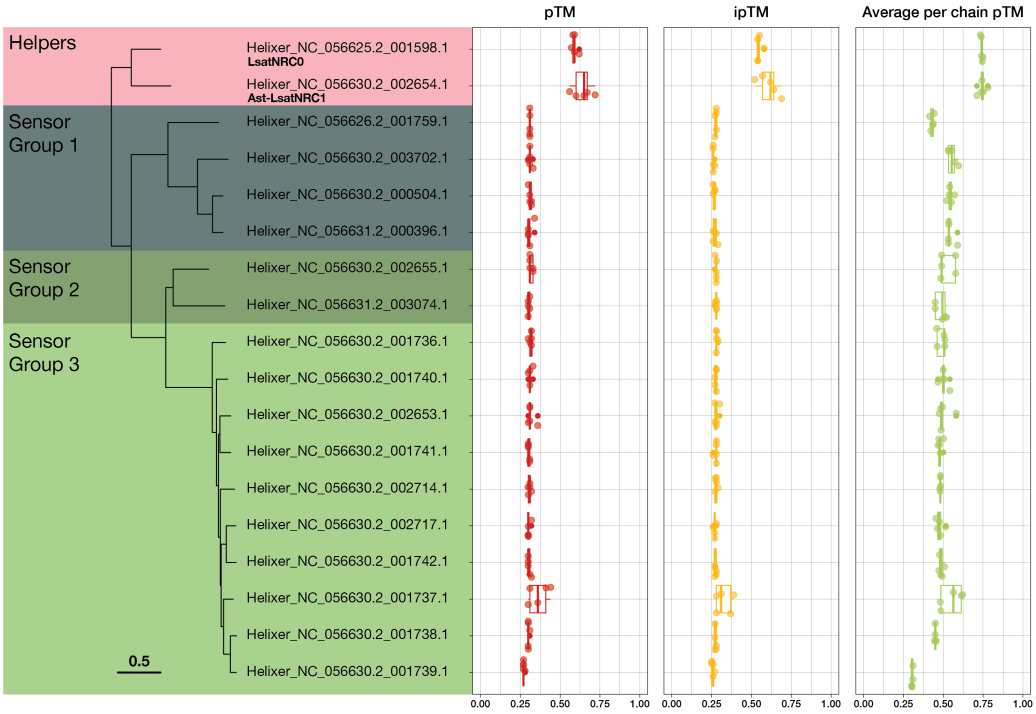


Figure 6.S10: Structural modeling of lettuce NRC-H and NRC-S with AlphaFold 3.
Predicted aligned error matrices and per-replicate confidence metrics for the hexameric models of each lettuce helper and sensor (Abramson et al., 2024). This figure has also been published in Pai et al. (Pai et al., 2026) and appears here with permission.

Chapter 7: General Discussion

The evolution of plant NLR immune receptors is most often framed as an arms race with pathogen effectors. Pathogens deploy effectors, receptors evolve recognition, pathogens mutate to evade detection, and the cycle repeats. This dynamic is well documented, and its molecular signature appears throughout the sensor datasets analyzed in this thesis. Yet the arms-race model primarily explains the recognition surfaces of sensor and singleton NLRs. It reveals little about the rest of the protein, and it does not explain why helper NLRs, which do not recognize pathogen effectors directly, evolve at distinct rates. This thesis investigated the selective pressures that act on NLRs from within the immune receptor network.

Three core conclusions emerge from this work. First, the evolutionary rate of an NLR reflects partner compatibility rather than functional output alone. Second, NLR sequence evolution proceeds under distinct regimes that act on separate domains of the protein. Third, sensor subfunctionalization involves the loss of autonomous oligomerization, which potentially explains why structural modeling can distinguish sensors from helpers where sequence motifs fail (Figure 7.1).

7.1 Diversification structured by internal partner compatibility

Two findings in this thesis indicate that sequence variation is structured by internal partner compatibility rather than by direct pathogen pressure. In lettuce, the rate of sensor diversification correlated with helper dependency, forming a spectrum from receptors dependent on the ancestral NRC0 to those signaling through the lineage-specific derived helper Ast-LsatNRC1 (Chapter 4). Within the Solanaceae NRC helper clade, the interfaces mediating helper homodimerization diverged between clades while remaining conserved within them (Chapter 3).

Consequently, the classification of an NLR as a sensor or a helper does not fully define the selective pressure acting on it. Instead, evolutionary constraint depends on the position of the receptor within the network, reflecting both the number of its interacting partners and their evolutionary conservation. A helper downstream of multiple sensors and a sensor dependent on an ancestral helper both face evolutionary constraint that does not stem directly from pathogen exposure (Figure 7.1).

This model also refines the classic view of sensor–helper subfunctionalization. Special-

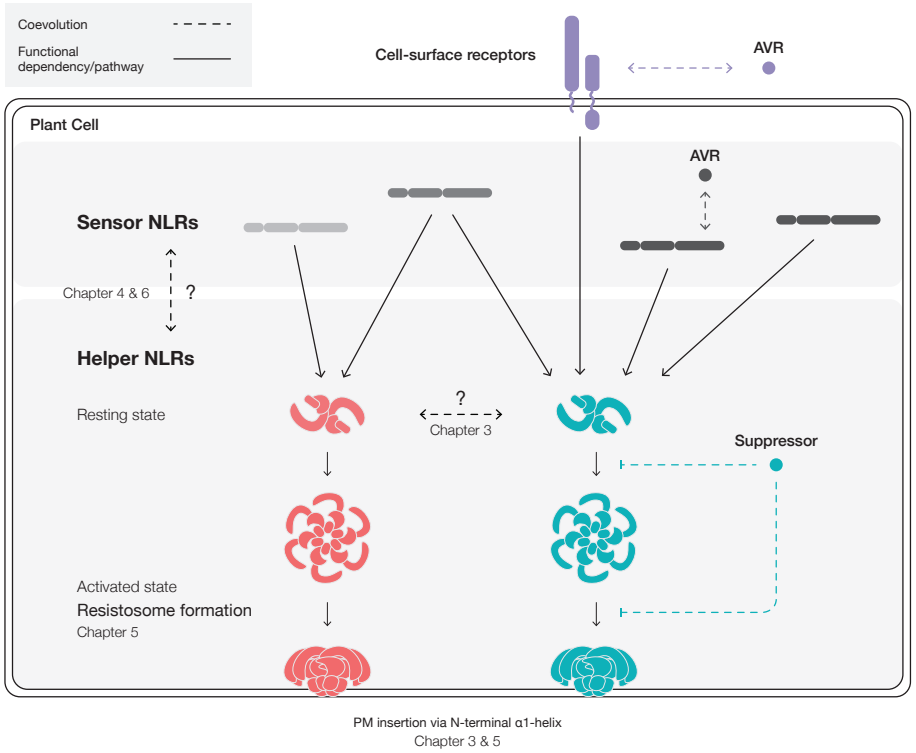


Figure 7.1: A conceptual model of the selective pressures shaping NLR evolution within an immune receptor network.

Sensor NLRs (gray) detect pathogen effectors (AVR) and signal to downstream helper NLRs. Solid lines denote functional dependencies or signaling pathways, and dashed lines denote co-evolutionary relationships. In the classical arms-race model, co-evolution takes place between pathogen effectors and the recognition surfaces of sensor NLRs, and between pathogen-derived molecules and cell-surface receptors (purple), whose signaling can also activate helper NLRs (Kourelis et al., 2022). Pathogen suppressors that target a specific helper at distinct stages of activation impose an additional co-evolutionary pressure on helpers to evade suppression (Derevnina et al., 2021; Contreras et al., 2023b; Sugihara et al., 2025; Seager et al., 2026).

This thesis examined co-evolution beyond R-AVR interactions that arises within the network itself: whether sensors and their helpers co-evolve and how this relates to their signaling dependency (Chapter 4), and whether distinct evolutionary trajectories reflect the ability of sensors and helpers to oligomerize, assessed by structural modeling (Chapter 6). In the resting state, NRC helpers accumulate as homodimers (Selvaraj et al., 2024), and the interfaces that mediate this self-association have diverged between paralogous helpers while remaining conserved within each clade, suggesting selective pressure to insulate their signaling pathways from one another (Chapter 3). Upon activation, helpers assemble into higher-order resistosomes (Chapter 5) that insert into the plasma membrane (PM) through their N-terminal α1 helix (Chapters 3 and 5). Question marks indicate the modes of co-evolution hypothesized in this work.

ization is usually described as relaxed evolutionary pressure on individual receptors that frees sensors to diversify (Adachi et al., 2019b). The results presented here suggest that subfunctionalization can redistribute constraint across the network rather than remove it. Sensors lost the ability to execute cell death but acquired the requirement to signal through a cognate helper. Conversely, helpers were released from pathogen detection but must maintain compatibility with an expanding repertoire of upstream sensors.

One caveat should be noted. The correlation between helper dependency and sensor diversification observed in Chapter 4 is consistent with two alternative mechanisms that current data cannot separate. Sensors dependent on the ancestral NRC0 may be constrained by the need to maintain helper compatibility, or by the host targets they monitor, particularly if they guard conserved host components in a manner analogous to the ZAR1–ZRK system (Adachi et al., 2023b). Nevertheless, these sensors often show signs of diversification in regions not typically implicated in effector recognition, such as the coiled-coil (CC) domain. This could be interpreted as a sign of relaxed selection following subfunctionalization within these sensors. Distinguishing between these possibilities requires identifying receptors that have switched helper dependency while retaining effector recognition specificity. Testing such natural or engineered variants will be essential to establish a direct causal mechanism.

7.2 Selection pressures beyond R-AVR co-evolution

By studying multiple NRC helpers and their sensors across this thesis, I found several distinct patterns of diversification. The most rapidly evolving lettuce sensors carry sites under positive selection concentrated on the concave face of the LRR, the surface implicated in effector binding, as described for highly variable receptors such as RPP1 (Prigozhin and Krasileva, 2021). The most constrained lettuce sensors carry variable sites in the CC domain and the N-terminal portion of the LRR rather than at the concave LRR surface implicated in recognition (Chapter 4). Helper NLRs, in contrast, show a different pattern. The homodimerization interfaces of the NRC helpers are less conserved than the P-loop and MHD motifs across the clade, yet well conserved within individual clades, with polymorphisms unique to each (Chapter 3).

Under the classical R-AVR co-evolution model, sustained diversification is favored because it enables novel recognition of effectors by the receptor, or evasion of recognition by the effector. As discussed throughout this thesis, this pattern is usually observed at recognition interfaces such as the concave side of the LRR domain or the exposed surfaces of Solanaceous or integrated domains. However, I detected two additional patterns of diversification, one in NRC helpers and one in their sensors. In NRC

helpers, stretches in the CC, NB-ARC, and LRR domains mediate homodimerization and, at least in tomato NRC2, homotetramerization of the helpers in their resting state (Selvaraj et al., 2024; Ma et al., 2024). Consistent with biochemical, microscopy, and genetic data, I hypothesized that the diversification of these stretches leads to insulation of the helper signaling pathways, which may contribute to the fine-tuning and robustness of the network. My analysis of NRC helpers is consistent with this model, which points to a selection pressure that acts within the network, independent of the classical R-AVR co-evolution paradigm that mostly applies to sensor NLRs. Helper NLRs have also been reported to co-evolve directly with pathogen effectors, not for recognition, but to evade suppression (Sugihara et al., 2025). This model has been dubbed "Gene-for-Gene 2.0" (Contreras, 2023; Sugihara et al., 2025).

Moreover, I found that not all sensor NLRs evolve in the same way. While the main evolutionary pressure considered for sensor NLRs relates to effector recognition, NRC0-dependent sensors showed little or no diversification on the concave side of their LRR domain. Instead, their diversified regions were located in the CC domain. Whether these patterns relate to sensor–helper communication, or whether these sensors function differently from others, remains an open question.

Taken together, the conclusions of this thesis expand these evolutionary models into a multi-layered framework in which effector recognition, evasion of suppression, receptor–receptor communication, and the maintenance of robust signaling all shape NLR evolution (Figure 7.1).

7.3 Subfunctionalization and the loss of the ability to oligomerize

The degeneration of N-terminal motifs in sensor NLRs is conventionally explained by a use-it-or-lose-it model, in which sensors that rely on helpers for cell death lost the MADA motif (Adachi et al., 2019a). It has also been shown that NRC-dependent sensor NLRs do not oligomerize or become part of the helper resistosome (Contreras et al., 2022; Ahn et al., 2023). However, the loss of oligomerization cannot be attributed only to the loss of the MADA motif, because the motif is also absent from the helpers of many sensor–helper pairs (Chapter 6). To address this limitation, I used AlphaFold 3 modeling of NLR homo-oligomeric resistosomes to assess whether AlphaFold 3 can distinguish between sensor and helper NLRs. Based on confidence metrics alone, sensor NLRs were generally modeled with lower confidence than their helper counterparts across several sensor–helper systems, including monocot paired NLRs and the superasterid

NRC network. From these data, I concluded that the loss of oligomerization is likely a product of sensor NLR subfunctionalization. This suggests that AlphaFold 3 detects a feature of sensor sequences that prevents them from being modeled with high confidence as resistosome-like oligomers. This feature likely resides within the CC or NB-ARC domains of sensor NLRs, where the oligomerization interfaces have undergone regressive evolution and degenerated.

However, one caveat is that the mechanism of action of many sensor–helper pairs, particularly the monocot paired NLRs, is still unknown (Pai et al., 2025b). Although NRC-dependent sensors are known to activate their helpers transiently, the activation mechanism may differ for other sensor–helper pairs. For instance, homo-oligomer modeling is inconclusive for sensor–helper pairs that may form a hetero-oligomeric resistosome complex (Guo et al., 2025).

Furthermore, poor modeling results do not necessarily define an NLR as a sensor. Poor modeling confidence can also point to neofunctionalization, as in NRCX (Adachi et al., 2023a), or to the formation of atypical assemblies, as in NRC7 (Toghani et al., 2026b). The modeling outcome should therefore be interpreted in its biological context and treated as a hypothesis.

7.4 How many protomers in a resistosome?

The computational modeling in Chapter 5 preceded the public release of NRC hexamer structures and was prompted by the determination of the NbNRC2 hexamer in our laboratory. Using AlphaFold 3, I showed that structural modeling can predict resistosome assemblies across diverse CC-type NLRs and reconstruct unresolved regions such as the membrane-inserting N-terminal α1 helix. In addition, AlphaFold 3 confidence metrics distinguished the correct oligomeric stoichiometry for the NbNRC2, AtZAR1, and TmSr35 resistosomes, although the score differences between stoichiometries were modest.

At the start of this project, the pentameric resistosome was regarded as the canonical activated state of plant CC-NLRs. Subsequent work revealed that NRC helpers assemble as hexamers (Liu et al., 2023a; Madhuprakash et al., 2024). Recent structural studies further showed that the wheat CC_{G10}-NLR WAI3 and *Arabidopsis thaliana* RPS2 form octamers (Guo et al., 2026). Most recently, our analysis revealed that the helper NRC7 assembles into an atypical 11-mer resistosome (Toghani et al., 2026b). Together, these discoveries show that plant CC-type NLRs form assemblies of five, six, eight, and eleven protomers, which raises the question of why NLRs adopt distinct oligomeric states.

Although the NB-ARC domain is one of the most conserved modules in plant

NLRs, it supports assemblies ranging from four protomers in TIR-NLR resistosomes (Ma et al., 2020; Martin et al., 2020) to eleven in NRC7. The structural determinants that govern stoichiometry remain unresolved. One possibility is that the NB-ARC domain provides a flexible structural scaffold, while the N-terminal executioner domain or the C-terminal LRR dictates the final number of protomers. For example, the loss of the canonical EDVID motif has been proposed to drive the divergent coiled-coil packing observed in octameric CC_{G10}-NLRs (Sulkowski et al., 2025; Guo et al., 2026).

The biological significance of this architectural variation is also unclear. In structurally resolved resistosomes, higher protomer counts expand the pore diameter. While known resistosomes trigger calcium influx and cell death upon activation, a larger pore may allow larger molecules to cross the host membrane. Furthermore, consistent with the sequence diversification of N-terminal executioner domains, activated NLRs can target distinct cellular compartments, including chloroplasts (Ibrahim et al., 2026) and vacuoles (Sunil et al., 2024). As in animal innate immunity, plant NLR signaling may encompass diverse execution mechanisms that converge on programmed cell death. The selective pressures that drive the evolution of distinct stoichiometries, and their precise physiological outputs, are yet to be determined.

7.5 Outlook

In this thesis, I investigated NLR biology from evolutionary and structural perspectives. In addition to classical arms-race dynamics with pathogens, these receptors experience distinct selective pressures that insulate and optimize immune signaling networks (Figure 7.1). Helper NLRs appear to evolve to keep their activation pathways separate and to prevent hetero-association with paralogs. Furthermore, while sensor NLRs diversify more rapidly than helpers overall, their evolutionary rates vary with their helper dependency and their potential recognition targets.

Using computational modeling, I showed that AlphaFold 3 generates high-confidence resistosome models that can accelerate the structural biology of plant NLRs. This is particularly useful given the sequence diversity and stoichiometric variation of NLR complexes across green plants. More broadly, structural modeling can distinguish helpers from sensors based on their predicted capacity to assemble into homo-oligomers. This structural criterion complements sequence-based annotation pipelines and could help predict sensor–helper relationships and map genetic dependencies from genomic datasets, with potential applications in crop breeding.

Throughout this thesis, I integrated phylogenomics, computational structural biology,

and *in planta* functional assays to generate mechanistic hypotheses and test evolutionary principles. Combining AI-guided workflows with targeted functional experiments should speed up the study of immune receptor evolution and support the rational engineering of durable disease resistance.

Chapter 8: Materials and Methods

The methods below are organized by the chapter in which the corresponding results are presented. Where a method was used in more than one chapter it is described in full at first mention and cross-referenced thereafter.

Chapter 2: Curation of harmonized plant NLRome datasets on multiple taxonomic resolutions

Genome assemblies and annotation

High-quality publicly available genome assemblies were compiled for each of the three datasets. For the 180 RefSeq reference plant proteomes, annotated proteomes were downloaded directly from the NCBI RefSeq database (Pruitt et al., 2007), and no de novo gene prediction was performed. For the 230 superasterid and 346 Solanaceae genome collections, harmonized annotations were generated de novo. Helixer v0.3.2 (Stiehler et al., 2021; Holst et al., 2023) was run using Singularity on the genome FASTA files with the option `--lineage land_plant`, which applies the default land-plant model (`land_plant_v0.3_a_0080.h5`). Coding sequences (CDS) and protein FASTA files were extracted from the resulting GFF files using GffRead v0.12.7 (Pertea and Pertea, 2020) with the `-x` and `-y` options, respectively.

NLR annotation

The extracted protein sequences were analyzed using NLRtracker (Kourelis et al., 2021), which integrates InterProScan for domain identification (Blum et al., 2025), to detect NLR and NLR-associated sequences. NLRs were defined according to the NLR definition established for the RefPlantNLR dataset (Kourelis et al., 2021).

Assessment of annotation completeness

BUSCO scores were generated using BUSCO v5.5.0 with the options `-m protein --lineage_dataset viridiplantae_odb10` for the superasterid dataset, and BUSCO v6.0.0 with `-m protein --lineage_dataset solanales_odb12` for the Solanaceae dataset (Manni et al., 2021). For the RefSeq dataset, annotation BUSCO scores provided by NCBI were used where available. Genome annotations were organized by phylogenetic

order following the APG IV classification (The Angiosperm Phylogeny Group, 2016). Summary statistics were computed and plotted in Python; the scripts and source data are archived with the thesis.

Chapter 3: Functional and molecular evolution of the NLR proteins beyond R-AVR co-evolution

Phylogenetic analyses of the NRC network

NLRtracker (Kourelis et al., 2021) output from 123 genome assemblies representing 38 species across 4 genera of the Solanaceae (Sugihara et al., 2023) was imported into R, from which 66,451 full-length NLR sequences were extracted. Of these, the 23,872 sequences with domain architectures of "CNL" or "CN" as determined by NLRtracker were retained, and their NB-ARC domains extracted. The NB-ARC sequences of RefPlantNLR (Kourelis et al., 2021), together with tomato (Wu et al., 2017) and pepper (Lee et al., 2021) NRC helpers, were added to this set. Using the NB-ARC domain boundaries reported by NLRtracker, only sequences with NB-ARC domains between 300 and 400 amino acids in length were kept, so that truncated or unusually long sequences were discarded. This left 19,760 sequences, which were aligned with MAFFT v7.505 `[--anysymbol]` (Katoh and Standley, 2013) and used to generate a phylogenetic tree with FastTree v2.1.11 `[-lg]` (Price et al., 2010).

The NRC network superclade, comprising 10,665 sequences including both helpers and sensors, was extracted from the CC-NLR tree by selecting the major branch containing both the reference NRC helpers and sensors, using Dendroscope v3.8.10 (Huson and Scornavacca, 2012). These sequences were realigned and used to build a new tree, from which the NRC helper clade was extracted in the same way. The resulting 1,796 NRC helper sequences, excluding the reference sequences, were deduplicated with CD-HIT v4.8.1 `[-c 1.00]` (Fu et al., 2012). Deduplication was performed to capture the greatest diversity without biasing the analysis through over-represented sequences.

The deduplicated sequences were then filtered by length, retaining only those between 750 and 950 amino acids, which left 993 sequences. This threshold was chosen on the basis of the length distribution of the NRC helper set and the lengths of the reference NRC sequences. The NB-ARC domains of these 993 sequences were combined with the length-filtered NB-ARCs (300–400 amino acids) of reference NRC helpers from *N. benthamiana* (16 sequences), tomato (13), pepper (10), potato (42), and RefPlantNLR (14), giving 1,088 sequences in total, and aligned with MAFFT. The alignment was used to

construct a further tree with FastTree, from which NRC helper subclades were delimited on the basis of reference NRCs falling within well-supported major branches. Sequences falling outside or between major clades were not assigned to a subclade. Clades were extracted with Dendroscope as before.

The full set of 1,092 NRC sequences comprises the 993 filtered full-length sequences together with the unfiltered reference sequences from *N. benthamiana* (16), tomato (13), pepper (14), potato (42), and RefPlantNLR (14). These were aligned with MAFFT `[--localpair]` in order to locate the dimerization stretches across all NRC helpers, and the alignment was trimmed with ClipKIT v2.0.1 `[-m gappy]` to remove sites with 90% gaps (Steenwyk et al., 2020).

Mapping of dimerization stretches and conserved motifs

The dimerization stretches and conserved sequence motifs were mapped onto the alignment of all NRC helpers on the basis of NbNRC2 and the NLRtracker output, and extracted using custom R scripts. Sequence logo plots for the stretches and motifs were generated with the `ggseqlogo` R package (Wagih, 2017).

Shannon entropy analysis

Shannon entropy values were calculated for all NRC helpers together and for each NRC clade separately, using the `entropy` R package (Hausser and Strimmer, 2009). Entropy ranges from zero for a fully conserved position to a maximum of 4.32 ($\log_2 20$) where all 20 amino acids occur with equal frequency. A threshold of 1.5 was used to identify highly variable positions, consistent with previous NLR studies (Prigozhin and Krasileva, 2021).

Identification of clade-unique residues

To identify residues unique to individual NRC clades, the relative amino acid ratio was calculated among the NRC2, NRC3, NRC4a, NRC4b, and NRC4other clades. Sequences were first divided into query and subject groups. Using the trimmed alignment of NRC helpers, the amino acid frequency at each position was calculated for both groups, and the frequency in the query group divided by that in the subject group at the same position. Only positions at which an amino acid frequency exceeded 80% within the query group were retained, and positions with a relative amino acid ratio greater than 50 were regarded as unique to the query clade.

Extraction of CC domains and N-terminal sequences

From the same NLRtracker output, the CC domain of each NLR was extracted from the start of the sequence to the beginning of the NB-ARC domain as determined by NLRtracker. The CC domain sequences were used for motif analysis with MEME (Bailey et al., 2015). To extract the N-terminal $\alpha 1$ helix, the CC domain sequences were first processed with NetSurfP-3.0 (Høie et al., 2022) to determine helix boundaries; the $\alpha 1$ helix was then extracted using a custom R script and used for clustering. Unbiased clustering of the N-terminal sequences was performed with MCL (Van Dongen, 2008), and the resulting clusters were mapped onto the NRC helper phylogeny. MADA motifs were identified using the MADA hidden Markov model (Adachi et al., 2019a) with HMMER (Eddy, 2011).

Transient protein expression and cell death assays

Proteins of interest were transiently expressed in *N. benthamiana* following previously described methods (Wu et al., 2017). In brief, leaves of 4–5-week-old plants were infiltrated with suspensions of *Agrobacterium tumefaciens* GV3101 pMP90 transformed with the relevant expression vectors. The final OD_{600} of all suspensions was adjusted in infiltration buffer (10 mM MES, 10 mM $MgCl_2$, and 150 μM acetosyringone, pH 5.6), and an OD_{600} of 0.3 was used for each construct. Autoactive variants were generated by introducing a D-to-V substitution within the MHD motif (Adachi et al., 2019a). Assays were performed in wild-type and in *nrc2/3/4* CRISPR knockout *N. benthamiana* lines (Wu et al., 2017).

Chapter 4: Contrasting patterns of evolution in sensor and helper NLRs

Extraction of NRC network sequences

A total of 29,261 NLRs were compiled from the NLRtracker output of 39 Solanales and 29 Asterales de novo annotated high-quality genome assemblies obtained from NCBI, together with a chromosome-level *Solanum melongena* assembly from the GWH Genome Warehouse. Annotation followed the Helixer-based workflow described for Chapter 2. On the basis of the simplified NLRtracker domain annotation, only NLRs with the domain architectures “CNL”, “CNLO”, “CN”, “OCNL”, “CONL”, “NL”, “NLO”, “ONL”, “BCNL”, “BNL”, “BCN”, “BCCNL”, “BNLO”, “BOCNL”, “RNL”, “TN”, “TNL”, “TNLO”, and “TNLJ” were retained, leaving 25,042 sequences. These were deduplicated with a custom R script re-

moving redundancy at 100% sequence identity, giving 24,518 non-redundant sequences. Finally, entries whose NB-ARC domain was shorter than 250 or longer than 400 amino acids were removed, yielding the 21,232 sequences used for phylogenetic analysis.

To construct a phylogenetic tree of the extracted NLRs, the NB-ARC domain sequences were aligned together with the RefPlantNLR NB-ARCs using FAMSA v2.2.2 with default options (Deorowicz et al., 2016), and the alignment used to build a tree with FastTree v2.1.11 `[-lg]` (Price et al., 2010; Kourelis et al., 2021). The NRC network, together with CcRPP1 and its orthologs, was extracted as a well-supported branch containing both NRC helpers and sensors identified by reference sequences. The NB-ARC domains of the extracted sequences were aligned with MAFFT v7.525 (Katoh and Standley, 2013) and used to construct a new tree as before, which was rooted on the CcRPP1 phylogroup.

Phylogenetic and gene distance analyses

Plant taxonomy trees and taxon divergence times were obtained from the TimeTree 5 database (Kumar et al., 2022). From the dataset, 122 NRC sequences including helpers and sensors from *Lactuca sativa*, *L. saligna*, *L. virosa*, *Cichorium intybus*, *Helianthus annuus*, *Chrysanthemum lavandulifolium*, *Cynara cardunculus*, and *Codonopsis lanceolata* were extracted on the basis of phylogeny. Their NB-ARC sequences were aligned with AtZAR1, NbZAR1, NbNRC2, and SlNRC0 as references using MAFFT v7.525 `[--anysymbol]`, and FastTree v2.1.11 `[-lg]` used to generate the tree of the NRC network in these species. IQ-TREE v2.3.0 `[-B 1000 -m MFP]` (Minh et al., 2020) was then used to generate a phylogenetic tree of the 40 NRC network sequences of the *Lactuca* genus. The resulting tree was divided into helpers and sensor groups 1, 2, and 3 on the basis of branch lengths and reference sequences.

To determine the physical location of genes on the chromosomes, the GFF output from Helixer was first simplified with GffRead using default options (Pertea and Pertea, 2020) and then imported into R with the `rtracklayer` package (Lawrence et al., 2009). Physical distances between genes were calculated with a custom R script, and gene coordinates exported for visualization on chromosomes using the MG2C online tool (Chao et al., 2021). The output was then arranged manually.

Linking previous lettuce NLRome classifications to the new annotation

Lettuce NLRome identifiers from Christopoulou and colleagues (Christopoulou et al., 2015) and the Lettuce V8 genome assembly from Phytozome 13 (Goodstein et al., 2012) were retrieved. BLASTP, as part of BLAST+ v2.16.0 (Camacho et al., 2009), was used

to identify the gene identifiers in the Phytozome 13 annotation corresponding to the new Helixer annotation. The BLAST output was processed in R, and gene coordinates, major resistance locus information, and RGC group assignments retrieved for the new annotation. Corresponding genes between the two annotations were identified from BLAST hits with greater than 98% sequence identity and bit scores above 1500. The same approach was used to retrieve corresponding sequences from the RefSeq reference annotation of Lsat_Salinas_v11 (GCF_002870075.4).

Cloning, agroinfiltration, and cell death assays

The NRC sequences used in the cell death assays were originally retrieved from the Phytozome 13 annotation and cross-checked against the Helixer annotation, with two sequences absent from the latter. Annotated genes were cloned using the Golden Gate Modular Cloning (MoClo) kit and the MoClo Plant Parts kit. In brief, coding sequences were synthesized by GENEWIZ as Golden Gate Level 0 pICH41155 modules and subsequently transferred into the binary vector pICH86988 by BsaI digestion; native BsaI sites in the original sequences were domesticated as required. In addition to wild-type clones, a D-to-V substitution within the MHD motif was introduced to generate autoactive versions of the lettuce NRCs (Adachi et al., 2019a). Verified clones were transformed into *A. tumefaciens* strain GV3101 pMP90 for ectopic expression in *N. benthamiana*.

For the cell death assays, wild-type and *nrc2/3/4 N. benthamiana* plants were grown in a controlled environment chamber at 22–25°C with a 16-hour photoperiod and 45%/65% relative humidity during the light and dark cycles respectively, at a light intensity of approximately 200 μmol m^{-2} s^{-1}. Four- to five-week-old plants were used for agroinfiltration following previously described methods. The final OD_{600} of all *Agrobacterium* suspensions was adjusted to 0.5 in infiltration buffer (10 mM MES, 10 mM $MgCl_2$, and 150 μM acetosyringone, pH 5.6), for both individual expression and co-expression of the NRC helpers and sensors of interest. Cell death phenotypes were imaged 4–5 days post-agroinfiltration.

Sequence polymorphism and selection analyses

Nucleotide sequences of the NRC genes were obtained with a custom Python script from a compiled CDS file of the annotated genomes. Codon-based nucleotide alignments were generated for each NRC helper and sensor phylogroup using MACSE v2.07 `[-prog alignSequences -seq]` (Ranwez et al., 2018). Positive selection tests were conducted on the resulting alignments using CodeML in PAML v4.10.7 (Yang, 1997). For each phylogroup (helpers and sensor groups 1, 2, and 3), two site

models were evaluated: M8a (beta + $\omega = 1$) and M8 (beta + ω). Likelihood ratio tests ($2\Delta L = 2(\ln L_{\text{alternative}} - \ln L_{\text{null}})$) were performed comparing M8a against M8 (df $= 1$; χ^2 critical value 6.63 at the 1% significance level). Sites under positive selection were identified with the Bayes Empirical Bayes method (Yang, 2005) under the M8 model, with posterior probabilities greater than 0.95, and the more stringent 0.99, taken to indicate strong support.

To compare average selective pressure across the NRC subclades, PAML branch model 2 was run in CodeML (`model = 2`, `NSsites = 0`, `CodonFreq = 2`), assigning each of the four subclades its own ω parameter via `$1`–`$4` labels placed at the most recent common ancestor of each subclade in the rooted tree. This four-ratio model was tested against the one-ratio model (M0; `model = 0`) by likelihood ratio test (df $= 4$). Per-subclade tree dS lengths were obtained by summing the M0 branch-level dS estimates over the stem and internal branches of each subclade. The free-ratio model (M1; `model = 1`) was also run for exploratory visualization of branch-level heterogeneity and is not used for hypothesis testing; in this model, branches with $\omega \geq 10$ reflect low synonymous substitution rates ($d\text{S} \approx 0$) and are flagged as ambiguous rather than as evidence of positive selection.

Pairwise nonsynonymous (dN) and synonymous (dS) substitution rates were estimated for full-length sequences and for the individual CC, NB-ARC, and LRR domains within each phylogroup using the Nei and Gojobori method (Nei and Gojobori, 1986). These rates were extracted from the `2NG.dN` and `2NG.dS` outputs of CodeML with a custom R script. Figures were generated with the `ggplot2` R package (Wickham, 2016).

Chapter 5: Computational structural modeling of NLR resistosomes reveals structural diversity in plant CC-type NLRs

Structural modeling and processing

ColabFold v1.5.5 (Mirdita et al., 2022; Jumper et al., 2021) was used to model the NbNRC2 CC-NB-ARC region. The AlphaFold 3 web server (Abramson et al., 2024) was used to model the CC-NB-ARC domains of 11 NRC helpers, 10 CC-NLRs, 6 CC_{G10}-NLRs, and 2 CC_{R}-NLRs, in each case with 50 oleic acid molecules as a proxy for the plasma membrane. The default seed was set to 1; the NbNRC4c and NbNRG1 hexamers were modeled with seed 11. To assess reproducibility of the NbNRC2 oligomers, nine additional models were generated with randomized seed values in addition to the default seed. From each modeling run only the top-ranked model (`model_0`) was

retained and processed.

The ipTM, pTM, chain-pair ipTM, and minimum per-chain pTM values were extracted from the AlphaFold 3 JSON metadata files using custom scripts and plotted in R. The AlphaFold 3 models corresponding to the cryo-EM structures of NbNRC2 (PDB 9FP6), AtZAR1 (6J5T), and TmSr35 (7XE0) were aligned to those structures using the `matchmaker` function of ChimeraX with default options (Pettersen et al., 2021). Contact points were extracted from the AlphaFold structures using the ChimeraX `alphafold contacts` command with a distance cut-off of 4.5 Å. All structures were visualized in ChimeraX and assembled manually.

Chapter 6: Using AI-based modeling to distinguish between Sensor and Helper NLRs

NLR annotation

NLRs were annotated using NLRtracker v1.0.3 (Kourelis et al., 2021) and InterProScan v5.67-99.0 (Blum et al., 2025). InterProScan did not annotate an integrated domain in PIK5-NP, contrary to a previous report (Kourelis et al., 2021); the domain architecture of PIK5-NP was therefore manually changed from "CNL" to "CONL", since it contains an HMA domain between the NB-ARC and LRR domains (Białas et al., 2021). Domain architectures were visualized with `refplantnlR`. MADA motifs were analyzed using the hidden Markov model of Adachi and colleagues (Adachi et al., 2019a) together with HMMER v3.4 (Eddy, 2011) with the `--max` option.

Curation of NLR sequences

The sequences of the curated NLR pairs were obtained either from RefPlantNLR (Kourelis et al., 2021) or from NCBI. For the NLR pairs described by Stein and colleagues (Stein et al., 2018), the protein sequences of *Oryza sativa* cv. Nipponbare were downloaded from the accompanying data release. From those data, NLR pairs that are genetically linked in head-to-head orientation and belong to distinct phylogenetic clades were extracted. Using DIAMOND BLASTP v2.1.9 (Buchfink et al., 2015), the corresponding NLR sequences were identified in the NCBI RefSeq annotation of the rice cultivar Nipponbare genome (GCF_034140825.1). The best-hit sequences were retained and confirmed to be genetically linked in a head-to-head orientation. On the basis of the NLRtracker output, only the CC-NLR pairs carrying a full N-terminal CC domain were

kept. Pair-05, Pair-13, and the sensor of Pair-14 correspond to Pia, PIK5/6-NP, and Pi-ta, respectively.

AlphaFold 3 prediction

Using the NLRtracker output, the amino acid sequences from the N terminus to the end of the NB-ARC domain were extracted. These were modeled with the AlphaFold 3 web server (Abramson et al., 2024) in both pentameric and hexameric configurations, in each case with 50 oleic acid molecules and using three different seed values (1, 2, and 3). For the wheat helper RXL, the full-length amino acid sequence was used as input owing to its atypical domain architecture (Guo et al., 2025). For the NLR pairs described by Stein and colleagues (Stein et al., 2018), the pTM scores of the pentameric and hexameric configurations were averaged, and the members of each pair assigned as putative helpers or sensors according to whether their score was the higher or the lower of the two.

For the lettuce NRC network, the two helpers and 16 sensors were modeled as hexamers with 50 oleic acid molecules. The first model of each sequence was generated with seed 1 and four further replicates with random seeds. Model confidence values and metadata were processed and plotted in R.

Statistical analysis and phylogenetics

Welch's *t*-test was performed using the `ttest_ind` function of the SciPy Python library with the option `equal_var = False`. The tree of monocot paired NLRs was built using RefPlantNLR (Kourelis et al., 2021) together with a set of 4,936 NLR proteins from the NLRtracker output of 13 RefSeq proteomes, comprising two dicot species (***Arabidopsis thaliana*** and ***Solanum lycopersicum***) and eleven monocot species (***Zea mays***, ***Triticum aestivum***, ***Setaria viridis***, ***Phragmites australis***, ***Phoenix dactylifera***, ***Oryza sativa***, ***Musa acuminata***, ***Lolium perenne***, ***Hordeum vulgare*** subsp. ***vulgare***, ***Brachypodium distachyon***, and ***Asparagus officinalis***) (Toghani and Kamoun, 2024).

Data and code availability

The harmonized annotation and NLR datasets described in Chapter 2 are openly available on Zenodo and Dryad (Toghani and Kamoun, 2024; Toghani et al., 2025b; Toghani et al., 2026c). Scripts and derived data associated with the analyses in Chapters 3–6 are available in the corresponding publication repositories (Selvaraj et al., 2024; Pai et al., 2026; Madhuprakash et al., 2024; Toghani et al., 2025a),

and the analysis code and source data for the figures generated for this thesis are archived alongside the relevant chapters.

Declaration of Artificial Intelligence usage

I used Claude (Anthropic) to assist in setting up the initial Git repository and LaTeX project structure, document templates, and syntax. I also consulted Claude to draft preliminary chapter outlines synthesized from my published manuscripts and PhD progress reports. Over the course of writing, these outlines and the structure of each chapter were substantially revised, expanded, and refined independently through my own work. This approach was discussed with my PhD supervisory committee and complies fully with the University of East Anglia policy on the use of artificial intelligence in research and thesis writing. A complete export of all interactive transcripts is openly archived in the accompanying Zenodo repository (Toghani, 2026).

Abbreviations

AAA+ ATPases Associated with diverse cellular Activities

ADP Adenosine Diphosphate

ADPR Adenosine Diphosphate Ribose

AF3 AlphaFold 3

AI Artificial Intelligence

ANL Autonomous NLR

APG IV Angiosperm Phylogeny Group IV classification

ARC Apoptotic ATPase domain shared by APAF-1, R proteins and CED-4

ATP Adenosine Triphosphate

AVR Avirulence protein (pathogen effector recognized by an R protein)

BEB Bayes Empirical Bayes

BED BED-type zinc finger domain

BLAST Basic Local Alignment Search Tool

BLASTP Protein–protein BLAST

BUSCO Benchmarking Universal Single-Copy Orthologs

cAMP Cyclic Adenosine Monophosphate

CC Coiled-Coil

CC-NLR NLR with an N-terminal coiled-coil domain (see CNL)

CC_{G10} G10-type coiled-coil domain

CC_{R} RPW8-type coiled-coil domain

CD-HIT Cluster Database at High Identity with Tolerance

CDS Coding Sequence

cGMP Cyclic Guanosine Monophosphate

CNL CC-NLR (NLR with an N-terminal coiled-coil domain)

CodeML Codon-based maximum likelihood program of the PAML package

CRISPR Clustered Regularly Interspaced Short Palindromic Repeats

cryo-EM Cryogenic Electron Microscopy

df Degrees of freedom

DIAMOND Double Index Alignment of Next-generation sequencing Data (protein aligner)

DNA Deoxyribonucleic Acid

dN/dS Ratio of nonsynonymous to synonymous substitution rates (ω)

DV Aspartate-to-valine substitution in the MHD motif (autoactive mutant)

EDVID Conserved Glu-Asp-Val-Ile-Asp motif in the α3 helix of canonical CC domains

ETI Effector-Triggered Immunity

ETS Effector-Triggered Susceptibility

FAMSA Fast and Accurate Multiple Sequence Alignment

FASTA Sequence file format

Gb Gigabase pairs

GFF, GFF3 General Feature Format (gene annotation file format; GFF3, version 3)

GOF Gain of Function

GPU Graphics Processing Unit

GWH Genome Warehouse

HD1 Helical Domain 1

HMA Heavy Metal-Associated domain

HMM Hidden Markov Model

HMMER Profile hidden Markov model search package

hpi Hours Post Infiltration

HR Hypersensitive Response

ID Integrated Domain

InterProScan Protein domain and family annotation software

IP Immunoprecipitation

ipSAE Interaction prediction Score from Aligned Errors

ipTM Interface predicted Template Modeling score

IQ-TREE Maximum likelihood phylogenetic inference software

iTOL Interactive Tree Of Life

JSON JavaScript Object Notation

kb Kilobase pairs

kDa Kilodalton

LIS Local Interaction Score

$\ln L$ Log-likelihood

LRR Leucine-Rich Repeat

MACSE Multiple Alignment of Coding SEquences

MADA Conserved N-terminal α1-helix motif of CC-NLRs

MADAL MADA-like motif

MAEPL MADA-like N-terminal motif of non-flowering plants

MAFFT Multiple Alignment using Fast Fourier Transform

MAMP Microbe-Associated Molecular Pattern

MCL Markov Cluster algorithm

MEME Multiple Em for Motif Elicitation

MES 2-(*N*-morpholino)ethanesulfonic acid

MFP ModelFinder Plus (IQ-TREE model selection)

MG2C MapGene2Chromosome

MHD Methionine-Histidine-Aspartate motif

MoClo Modular Cloning

MRC Major Resistance Cluster

MSA Multiple Sequence Alignment

MYA Million Years Ago

NAD^+ Nicotinamide Adenine Dinucleotide (oxidized form)

NADase NAD^+-hydrolyzing enzyme

NB-ARC Nucleotide-Binding adaptor shared by APAF-1, R proteins and CED-4

NBD Nucleotide-Binding Domain

NCBI National Center for Biotechnology Information

NetSurfP Protein secondary structure and solvent accessibility predictor

NLR Nucleotide-binding Leucine-rich Repeat receptor

NLRome Complete NLR repertoire of a genome or set of genomes

NLRtracker NLR annotation and classification tool

NOD Nucleotide-binding and Oligomerization Domain

NRC NLR Required for Cell death

NRC-H NRC helper NLR

NRC-S NRC-dependent sensor NLR

NTPase Nucleoside Triphosphatase

OD Optical Density

PAE Predicted Aligned Error

PAML Phylogenetic Analysis by Maximum Likelihood

PAMP Pathogen-Associated Molecular Pattern

PDB Protein Data Bank

pDockQ Predicted DockQ score

pLDDT Predicted Local Distance Difference Test

P-loop Phosphate-binding loop (Walker A motif)

PM Plasma Membrane

PPR Pentatricopeptide Repeat

pRib-AMP/ADP Phosphoribosyl-AMP/ADP

PRR Pattern Recognition Receptor

PS Ponceau Stain

PSIPRED PSI-BLAST-based secondary structure prediction

PTI Pattern-Triggered Immunity

pTM Predicted Template Modeling score

R gene Resistance gene

RefPlantNLR Curated collection of experimentally validated plant NLRs

RefSeq NCBI Reference Sequence database

RGC Resistance Gene Candidate

RLCK Receptor-Like Cytoplasmic Kinase

RLK Receptor-Like Kinase

RLP Receptor-Like Protein

RMSD Root Mean Square Deviation

RNL RPW8-type CC_R-NLR

RPW8 RESISTANCE TO POWDERY MILDEW 8

SD Solanaceous Domain

SSFR Superstructure-Forming Repeat

STAND Signal Transduction ATPases with Numerous Domains

TIR Toll/Interleukin-1 Receptor domain

TIR-NLR NLR with an N-terminal TIR domain (see TNL)

TNL TIR-NLR (NLR with an N-terminal TIR domain)

TSL The Sainsbury Laboratory

UTR Untranslated Region

WHD Winged-Helix Domain

WRKY WRKY transcription factor domain

WT Wild Type

ZMBP Center for Plant Molecular Biology (Zentrum für Molekularbiologie der Pflanzen), University of Tübingen

List of Figures

1.1 Domain architecture of plant NLR immune receptors 12

1.2 NLR configurations and phylogenetic organization 16

1.3 The MADA motif is a functionally replaceable unit at the N-terminus of CC-NLRs 18

1.4 Structural diversity of plant NLR resistosomes 23

1.5 Higher-order assemblies of nucleotide-binding receptors across the domains of life 25

2.1 Workflow for compiling harmonized NLR datasets 39

2.2 Overview of the 180-genome RefSeq NLRome 41

2.3 Overview of the 230-genome superasterid NLRome 45

2.4 Overview of the 346-genome Solanaceae NLRome 47

3.1 NRC sequences form a monophyletic clade 53

3.2 The N-terminal MADA motif has diversified in NRC helpers 54

3.3 NRCX has a distinct threonine at the fourth position of its MADA motif . 55

3.4 NRC4t sequences form a distinct phylogenetic clade 56

3.5 NRC4t sequences induce cell death as autoactive mutants 57

3.6 The helper NLR NRC2 forms a homodimer in planta 58

3.7 The NRC homodimerization interface has diversified through NRC evolution 60

3.8 The SlNRC2 tetramer and its interdimeric interface 63

3.S1 Sequence logos and Shannon entropy for all NRC clades 64

3.S2 Full-length Shannon entropy across NRC clades 65

4.1 The NRC network has expanded less in Asterales than in Solanales . . . 71

4.2 Ast-LsatNRC1, unlike LsatNRC0, is genetically linked to NRC-S 73

4.3 LsatNRC0 and Ast-LsatNRC1 mediate cell death for distinct sensor groups 75

4.4 Ast-LsatNRC1-dependent NRC-S are under stronger positive selection . 77
4.5 Evolutionary constraints and diversification in NRC-H and NRC-S . . . 80
4.S1 Phylogenetic tree of Solanales and Asterales NB-ARC domains 82
4.S2 NLR and NRC network counts per species 83
4.S3 Phylogenetic tree of the common lettuce NLRome 84
4.S4 Physical map of wild lettuce NRC sequences 85
4.S5 Phylogenetic tree of NRC sequences across selected Asterales 86
4.S6 Physical map of Codonopsis, Helianthus, and Cichorium NRC sequences 87
4.S7 Physical map of Chrysanthemum and Cynara NRC sequences 88
4.S8 Hypersensitive response cell death assays of NRC-H and NRC-S 89
4.S9 Multi-ratio model branch-specific dN/dS across the Lactuca NRC phylogeny 90
4.S10 Free-ratio model branch-specific dN/dS across the Lactuca NRC phylogeny 91
4.S11 Pairwise dN and dS values within the Lactuca NRC phylogroups 92
4.S12 Histogram of pairwise dN/dS ratios within the Lactuca NRC phylogroups 93
5.1 AlphaFold 3 modeling of the NRC2 hexameric resistosome 97
5.2 AlphaFold 3 modeling of pentameric resistosomes and CC-type NLRs . 99
5.S1 Comparison of AlphaFold 2 and AlphaFold 3 models of the NRC2 hexamer 103
5.S2 Reproducibility of the NRC2 hexamer prediction across random seeds . 104
5.S3 NRC2 modeled across stoichiometries and the predicted N-terminal pore 105
5.S4 AlphaFold 3 modeling of NRC helpers as hexamers and pentamers . . . 106
5.S5 Model confidence metrics across the NRC helper clade 107
5.S6 Benchmarking hexameric and pentameric models against known structures 108
5.S7 Phylogenetic placement of the CC-type NLRs selected for modeling . . . 109
5.S8 Structural modeling of CC-NLRs, part 1 110
5.S9 Structural modeling of CC-NLRs, part 2 111

5.S10 Structural modeling of CC-NLRs, part 3 112

5.S11 Structural modeling of CCG10-NLRs, part 1 113

5.S12 Structural modeling of CCG10-NLRs, part 2 114

5.S13 Structural modeling of CCR-NLRs 115

5.S14 Model confidence metrics across CC-type NLRs 116

6.1 Helper NLRs produce higher AlphaFold 3 confidence scores than paired sensors 120

6.2 NLRs without integrated domains give higher confidence scores in rice putative pairs 122

6.3 NRC-S fail to form resistosome-like structures in AlphaFold 3 models . . 124

6.S1 Phylogenetic tree of the NLRs modeled in this chapter 126

6.S2 Correlation between confidence metrics and stoichiometries 127

6.S3 Per-replicate confidence scores for the characterized NLR pairs 128

6.S4 Helper against sensor confidence scores for the characterized pairs . . . 129

6.S5 Pentameric models of the characterized NLR pairs 130

6.S6 Hexameric models of the characterized NLR pairs 131

6.S7 Per-replicate confidence scores for the rice putative pairs 132

6.S8 Pentameric models of the rice putative NLR pairs 133

6.S9 Hexameric models of the rice putative NLR pairs 134

6.S10 AlphaFold 3 structural modeling of lettuce NRC-H and NRC-S 135

7.1 A conceptual model of the selective pressures shaping NLR evolution within an immune receptor network 137

List of Tables

3.1 Constructs used in the NRC4t cell death assays 66

References

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, June 2024. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-024-07487-w.

Hiroaki Adachi, Mauricio P Contreras, Adeline Harant, Chih-hang Wu, Lida Derevnina, Toshiyuki Sakai, Cian Duggan, Eleonora Moratto, Tolga O Bozkurt, Abbas Maqbool, Joe Win, and Sophien Kamoun. An N-terminal motif in NLR immune receptors is functionally conserved across distantly related plant species. *eLife*, 8:e49956, November 2019a. ISSN 2050-084X. doi: 10.7554/eLife.49956.

Hiroaki Adachi, Lida Derevnina, and Sophien Kamoun. NLR singletons, pairs, and networks: Evolution, assembly, and regulation of the intracellular immunoreceptor circuitry of plants. *Current Opinion in Plant Biology*, 50:121–131, August 2019b. ISSN 13695266. doi: 10.1016/j.pbi.2019.04.007.

Hiroaki Adachi, Toshiyuki Sakai, Adeline Harant, Hsuan Pai, Kodai Honda, AmirAli Toghani, Jules Claeys, Cian Duggan, Tolga O. Bozkurt, Chih-hang Wu, and Sophien Kamoun. An atypical NLR protein modulates the NRC immune receptor network in nicotiana benthamiana. *PLOS Genetics*, 19(1):e1010500, January 2023a. ISSN 1553-7404. doi: 10.1371/journal.pgen.1010500.

Hiroaki Adachi, Toshiyuki Sakai, Jiorgos Kourelis, Hsuan Pai, Jose L Gonzalez Hernandez, Yoshinori Utsumi, Motoaki Seki, Abbas Maqbool, and Sophien Kamoun. Jurassic NLR: Conserved and dynamic evolutionary features of the atypically ancient immune receptor ZAR1. *Plant Cell*, 35(10):3662–3685, September 2023b. ISSN 1040-4651, 1532-298X. doi: 10.1093/plcell/koad175.

Hee-Kyung Ahn, Xiao Lin, Andrea Carolina Olave-Achury, Lida Derevnina, Mauricio P Contreras, Jiorgos Kourelis, Chih-Hang Wu, Sophien Kamoun, and Jonathan D G Jones. Effector-dependent activation and oligomerization of plant NRC class helper

NLRs by sensor NLR immune receptors rpi-amr3 and rpi-amr1. *EMBO Journal*, 42 (5):EMBJ2022111484, March 2023. ISSN 0261-4189, 1460-2075. doi: 10.15252/embj.2022111484.

Youwei Ai, Yutong Meng, Bo Yan, Qinyi Zhou, and Xiaodong Wang. The biochemical pathways of apoptotic, necroptotic, pyroptotic, and ferroptotic cell death. *Molecular Cell*, 84(1):170–179, January 2024. ISSN 10972765. doi: 10.1016/j.molcel.2023.11.040.

Giuseppe Andolfo, Antimo Di Donato, Pasquale Chiaiese, Antonino De Natale, Antonino Pollio, Jonathan D G Jones, Luigi Frusciante, and Maria Raffaella Ercolano. Alien domains shaped the modular structure of plant NLR proteins. *Genome Biology and Evolution*, 11(12):evz248, November 2019. ISSN 1759-6653. doi: 10.1093/gbe/evz248.

Liudmila Andreeva, Liron David, Shaun Rawson, Chen Shen, Teerithveen Pasricha, Pablo Pelegrin, and Hao Wu. NLRP3 cages revealed by full-length mouse NLRP3 structure control pathway activation. *Cell*, 184(26):6299–6312.e22, December 2021. ISSN 0092-8674. doi: 10.1016/j.cell.2021.11.011.

L. Aravind, Gianlucca G. Nicastro, Lakshminarayan M. Iyer, and A. Maxwell Burroughs. The prokaryotic roots of eukaryotic immune systems. *Annual Review of Genetics*, 58(1):365–389, November 2024. ISSN 0066-4197, 1545-2948. doi: 10.1146/annurev-genet-111523-102448.

Ikuo Ashikawa, Nagao Hayashi, Hiroko Yamane, Hiroyuki Kanamori, Jianzhong Wu, Takashi Matsumoto, Kazuko Ono, and Masahiro Yano. Two adjacent nucleotide-binding site–leucine-rich repeat class genes are required to confer *Pikm* -specific rice blast resistance. *Genetics*, 180(4):2267–2276, December 2008. ISSN 1943-2631. doi: 10.1534/genetics.108.095034.

Erin L. Baggs, J. Grey Monroe, Anil S. Thanki, Ruby O'Grady, Christian Schudoma, Wilfried Haerty, and Ksenia V. Krasileva. Convergent loss of an EDS1/PAD4 signaling pathway in several plant lineages reveals coevolved components of plant immunity and drought response. *Plant Cell*, 32(7):2158–2177, July 2020. ISSN 1532-298X, 1040-4651. doi: 10.1105/tpc.19.00903.

Timothy L. Bailey, James Johnson, Charles E. Grant, and William S. Noble. The MEME suite. *Nucleic Acids Research*, 43(W1):W39–W49, July 2015. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkv416.

A Cristina Barragan and Detlef Weigel. Plant NLR diversity: The known unknowns of pan-NLRomes. *Plant Cell*, 33(4):814–831, May 2021. ISSN 1040-4651, 1532-298X. doi: 10.1093/plcell/koaa002.

Inigo Barrio-Hernandez, Jingi Yeo, Jürgen Jänes, Milot Mirdita, Cameron L. M. Gilchrist, Tanita Wein, Mihaly Varadi, Sameer Velankar, Pedro Beltrao, and Martin Steinegger. Clustering predicted structures at the scale of the known protein universe. *Nature*, 622(7983):637–645, October 2023. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-023-06510-w.

Kristian Barry, Christopher Murphy, and Ashley Mansell. NLRP1- a CINDERELLA STORY: A perspective of recent advances in NLRP1 and the questions they raise. *Communications Biology*, 6(1):1274, December 2023. ISSN 2399-3642. doi: 10.1038/s42003-023-05684-3.

Adam M. Bayless, Sisi Chen, Sam C. Ogden, Xiaoyan Xu, John D. Sidda, Mohammad K. Manik, Sulin Li, Bostjan Kobe, Thomas Ve, Lijiang Song, Murray Grant, Li Wan, and Marc T. Nishimura. Plant and prokaryotic TIR domains generate distinct cyclic ADPR NADase products. *Science Advances*, 9(11):eade8487, March 2023. ISSN 2375-2548. doi: 10.1126/sciadv.ade8487.

Adam M. Bayless, Lijiang Song, Mitchell Sorbello, Sam C. Ogden, Tyler S. Todd, Alice Flint, Natsumi Maruta, Jedidiah Tulu, Mikhail Drenichev, Vardis Ntoukakis, Thomas Ve, Mehdi Mobli, Li Wan, Qingli Liu, Jeffery L. Dangl, Bostjan Kobe, Murray Grant, and Marc T. Nishimura. The *Arabidopsis* TIRome informs the design of artificial TIR (toll/interleukin-1 receptor) domain proteins. *Proceedings of the National Academy of Sciences*, 122(37):e2505893122, September 2025. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2505893122.

Josh W. Bennett, Yu Sugihara, John F. Haidoulis, Clemence A. Rodney, Rafal Zdrzalek, Enzo Zanchet, Indira Saado, Pirita Paajanen, Paul Nicholson, Soichiro Asuke, and Mark J. Banfield. A wheat immune receptor pair executes cell death through a helper MLKL, August 2026. ISSN 2692-8205.

Andrew F. Bent, Barbara N. Kunkel, Douglas Dahlbeck, Kit L. Brown, Renate Schmidt, Jerome Giraudat, Jeffrey Leung, and Brian J. Staskawicz. RPS2 of *Arabidopsis thaliana*: A leucine-rich repeat class of plant disease resistance genes. *Science*, 265(5180):1856–1860, September 1994. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.8091210.

Aude Bernheim, Jean Cury, and Enzo Z. Poirier. The immune modules conserved across the tree of life: Towards a definition of ancestral immunity. *PLOS Biology*, 22(7):e3002717, July 2024. ISSN 1545-7885. doi: 10.1371/journal.pbio.3002717.

Guozhi Bi and Jian-Min Zhou. Regulation of cell death and signaling by pore-forming resistosomes. *Annual Review of Phytopathology*, 59(1):239–263, August 2021. ISSN 0066-4286, 1545-2107. doi: 10.1146/annurev-phyto-020620-095952.

Guozhi Bi, Min Su, Nan Li, Yu Liang, Song Dang, Jiachao Xu, Meijuan Hu, Jizong Wang, Minxia Zou, Yanan Deng, Qiyu Li, Shijia Huang, Jiejie Li, Jijie Chai, Kangmin He, Yu-hang Chen, and Jian-Min Zhou. The ZAR1 resistosome is a calcium-permeable channel triggering plant immune signaling. *Cell*, 184(13):3528–3541.e12, June 2021. ISSN 00928674. doi: 10.1016/j.cell.2021.05.003.

Aleksandra Białas, Erin K. Zess, Juan Carlos De La Concepcion, Marina Franceschetti, Helen G. Pennington, Kentaro Yoshida, Jessica L. Upson, Emilie Chanclud, Chih-Hang Wu, Thorsten Langner, Abbas Maqbool, Freya A. Varden, Lida Derevnina, Khaoula Belhaj, Koki Fujisaki, Hiromasa Saitoh, Ryohei Terauchi, Mark J. Banfield, and Sophien Kamoun. Lessons in effector and NLR biology of plant-microbe systems. *Molecular Plant-Microbe Interactions®*, 31(1):34–45, January 2018. ISSN 0894-0282, 1943-7706. doi: 10.1094/MPMI-08-17-0196-FI.

Aleksandra Białas, Thorsten Langner, Adeline Harant, Mauricio P Contreras, Clare Em Stevenson, David M Lawson, Jan Sklenar, Ronny Kellner, Matthew J Moscou, Ryohei Terauchi, Mark J Banfield, and Sophien Kamoun. Two NLR immune receptors acquired high-affinity binding to a fungal effector through convergent evolution of their integrated domain, January 2021.

Matthias Blum, Antonina Andreeva, Laise Cavalcanti Florentino, Sara Rocio Chuguransky, Tiago Grego, Emma Hobbs, Beatriz Lazaro Pinto, Ailsa Orr, Typhaine Paysan-Lafosse, Irina Ponamareva, Gustavo A Salazar, Nicola Bordin, Peer Bork, Alan Bridge, Lucy Colwell, Julian Gough, Daniel H Haft, Ivica Letunic, Felipe Llinares-López, Aron Marchler-Bauer, Laetitia Meng-Papaxanthos, Huaiyu Mi, Darren A Natale, Christine A Orengo, Arun P Pandurangan, Damiano Piovesan, Catherine Rivoire, Christian J A Sigrist, Narmada Thanki, Françoise Thibaud-Nissen, Paul D Thomas, Silvio C E Tosatto, Cathy H Wu, and Alex Bateman. InterPro: The protein sequence classification resource in 2025. *Nucleic Acids Research*, 53(D1):D444–D456, January 2025. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkae1082.

Helen J Brabham, Diana Gómez De La Cruz, Vincent Were, Motoki Shimizu, Hiromasa Saitoh, Inmaculada Hernández-Pinzón, Phon Green, Jennifer Lorang, Koki Fujisaki, Kazuhiro Sato, István Molnár, Hana Šimková, Jaroslav Doležel, James Russell, Jodie Taylor, Matthew Smoker, Yogesh Kumar Gupta, Tom Wolpert, Nicholas J Talbot, Ryohei Terauchi, and Matthew J Moscou. Barley MLA3 recognizes the host-specificity effector Pwl2 from *Magnaporthe oryzae*. *Plant Cell*, 36(2):447–470, January 2024. ISSN 1040-4651, 1532-298X. doi: 10.1093/plcell/koad266.

Petr Broz, Pablo Pelegrín, and Feng Shao. The gasdermins, a protein family executing cell death and inflammation. *Nature Reviews Immunology*, 20(3):143–157, March 2020. ISSN 1474-1733, 1474-1741. doi: 10.1038/s41577-019-0228-2.

Benjamin Buchfink, Chao Xie, and Daniel H Huson. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1):59–60, January 2015. ISSN 1548-7091, 1548-7105. doi: 10.1038/nmeth.3176.

Michael Busche and Bradley Laflamme. From algae to apples: The structural and functional conservation of NLRs. *Plant Cell*, 36(7):2453–2454, April 2024. ISSN 1532-298X. doi: 10.1093/plcell/koae131.

Christiam Camacho, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L Madden. BLAST+: Architecture and applications. *BMC Bioinformatics*, 10(1):421–429, December 2009. ISSN 1471-2105. doi: 10.1186/1471-2105-10-421.

Emily J. Capra, Barrett S. Perchuk, Jeffrey M. Skerker, and Michael T. Laub. Adaptive mutations that prevent crosstalk enable the expansion of paralogous signaling protein families. *Cell*, 150(1):222–232, July 2012. ISSN 00928674. doi: 10.1016/j.cell.2012.05.033.

Baptiste Castel, Pok-Man Ngou, Volkan Cevik, Amey Redkar, Dae-Sung Kim, Ying Yang, Pingtao Ding, and Jonathan D. G. Jones. Diverse NLR immune receptors activate defence via the RPW 8- NLR NRG 1. *New Phytologist*, 222(2):966–980, April 2019. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.15659.

Stella Cesari, Maud Bernoux, Philippe Moncuquet, Thomas Kroj, and Peter N. Dodds. A novel conserved mechanism for plant NLR protein pairs: The â€œintegrated decoyâ€ hypothesis. *Frontiers in Plant Science*, 5:1–10, November 2014. ISSN 1664-462X. doi: 10.3389/fpls.2014.00606.

Stella Césari, Hiroyuki Kanzaki, Tadashi Fujiwara, Maud Bernoux, Véronique Chalvon, Yoji Kawano, Ko Shimamoto, Peter Dodds, Ryohei Terauchi, and Thomas Kroj. The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. *EMBO Journal*, 33(17):1941–1959, September 2014. ISSN 0261-4189, 1460-2075. doi: 10.15252/embj.201487923.

Jiangtao Chao, Zhiyuan Li, Yuhe Sun, Oluwaseun Olayemi Aluko, Xinru Wu, Qian Wang, and Guanshan Liu. MG2C: A user-friendly online tool for drawing genetic maps. *Molecular Horticulture*, 1(1):16–19, December 2021. ISSN 2730-9401. doi: 10.1186/s43897-021-00020-x.

Jiapeng Chen, Narayana M. Upadhyaya, Diana Ortiz, Jana Sperschneider, Feng Li, Clement Bouton, Susan Breen, Chongmei Dong, Bo Xu, Xiaoxiao Zhang, Rohit Mago, Kim Newell, Xiaodi Xia, Maud Bernoux, Jennifer M. Taylor, Brian Steffenson, Yue Jin, Peng Zhang, Kostya Kanyuka, Melania Figueroa, Jeffrey G. Ellis, Robert F. Park, and

Peter N. Dodds. Loss of *AvrSr50* by somatic exchange in stem rust leads to virulence for *Sr50* resistance in wheat. *Science*, 358(6370):1607–1610, December 2017. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aao4810.

Kaiwen W. Chen and Petr Broz. Gasdermins as evolutionarily conserved executors of inflammation and cell death. *Nature Cell Biology*, 26(9):1394–1406, September 2024. ISSN 1465-7392, 1476-4679. doi: 10.1038/s41556-024-01474-z.

Yi-Feng Chen, Kuan-Yu Lin, Ching-Yi Huang, Liang-Yu Hou, Enoch Lok Him Yuen, Wei-Che Jimmy Sun, Bing-Jen Chiang, Chin-Wen Chang, Hung-Yu Wang, Tolga Osman Bozkurt, and Chih-Hang Wu. Subcellular calcium dynamics and organelle perturbations in resistosome-mediated cell death. *Proceedings of the National Academy of Sciences*, 123(12):e2523470123, March 2026. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2523470123.

Khong-Sam Chia and Philip Carella. Taking the lead: NLR immune receptor N-terminal domains execute plant immune responses. *New Phytologist*, 240(2):496–501, October 2023. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.19170.

Khong-Sam Chia, Jiorgos Kourelis, Martin Vickers, Toshiyuki Sakai, Sophien Kamoun, and Philip Carella. The N-terminal executioner domains of NLR immune receptors are functionally conserved across major plant lineages, October 2022.

Marilena Christopoulou, Sebastian Reyes-Chin Wo, Alex Kozik, Leah K McHale, Maria-Jose Truco, Tadeusz Wroblewski, and Richard W Michelmore. Genome-wide architecture of disease resistance genes in lettuce. *G3 Genes|Genomes|Genetics*, 5(12):2655–2669, December 2015. ISSN 2160-1836. doi: 10.1534/g3.115.020818.

Mauricio P. Contreras. *Biochemical Basis of Activation and Inhibition of an NLR Immune Receptor Network*. PhD thesis, University of East Anglia, October 2023.

Mauricio P. Contreras, Hsuan Pai, Yasin Tumtas, Cian Duggan, Enoch Lok Him Yuen, Angel Vergara Cruces, Jiorgos Kourelis, Hee-Kyung Ahn, Kim-Teng Lee, Chih-Hang Wu, Tolga O. Bozkurt, Lida Derevnina, and Sophien Kamoun. Sensor NLR immune proteins activate oligomerization of their NRC helpers in response to plant pathogens. *EMBO Journal*, 42(5):EMBJ2022111519, December 2022. ISSN 1460-2075. doi: 10.15252/embj.2022111519.

Mauricio P Contreras, Daniel Lüdke, Hsuan Pai, AmirAli Toghani, and Sophien Kamoun. NLR receptors in plant immunity: Making sense of the alphabet soup. *EMBO Reports*, 24(10):EMBR202357495, October 2023a. ISSN 1469-221X, 1469-3178. doi: 10.15252/embr.202357495.

Mauricio P. Contreras, Hsuan Pai, Muniyandi Selvaraj, AmirAli Toghani, David M. Lawson, Yasin Tumtas, Cian Duggan, Enoch Lok Him Yuen, Clare E. M. Stevenson, Adeline Harant, Abbas Maqbool, Chih-Hang Wu, Tolga O. Bozkurt, Sophien Kamoun, and Lida Derevnina. Resurrection of plant disease resistance proteins via helper NLR bioengineering. *Science Advances*, 9(18):eadg3861, May 2023b. ISSN 2375-2548. doi: 10.1126/sciadv.adg3861.

Mauricio P. Contreras, Hsuan Pai, Rebecca Thompson, Jules Claeys, Hiroaki Adachi, and Sophien Kamoun. The nucleotide binding domain of NRC-dependent disease resistance proteins is sufficient to activate downstream helper NLR oligomerization and immune signaling, December 2023c.

Mauricio P. Contreras, Hsuan Pai, Rebecca Thompson, Clemence Marchal, Jules Claeys, Hiroaki Adachi, and Sophien Kamoun. The nucleotide-binding domain of NRC -dependent disease resistance proteins is sufficient to activate downstream helper NLR oligomerization and immune signaling. *New Phytologist*, 243(1):345–361, July 2024. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.19818.

Edward M. Culbertson and Tera C. Levin. Eukaryotic CD-NTase, STING, and viperin proteins evolved via domain shuffling, horizontal transfer, and ancient inheritance from prokaryotes. *PLOS Biology*, 21(12):e3002436, December 2023. ISSN 1545-7885. doi: 10.1371/journal.pbio.3002436.

Asen Daskalov, Birgit Habenstein, Raimon Sabaté, Mélanie Berbon, Denis Martinez, Stéphane Chaignepain, Bénédicte Coulary-Salin, Kay Hofmann, Antoine Loquet, and Sven J. Saupe. Identification of a novel cell death-inducing domain reveals that fungal amyloid-controlled programmed cell death is related to necroptosis. *Proceedings of the National Academy of Sciences*, 113(10):2720–2725, March 2016. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1522361113.

Juan Carlos De La Concepcion, Marina Franceschetti, Abbas Maqbool, Hiromasa Saitoh, Ryohei Terauchi, Sophien Kamoun, and Mark J. Banfield. Polymorphic residues in rice NLRs expand binding and response to effectors of the blast pathogen. *Nature Plants*, 4 (8):576–585, July 2018. ISSN 2055-0278. doi: 10.1038/s41477-018-0194-x.

Juan Carlos De La Concepcion, Javier Vega Benjumea, Aleksandra Bialas, Ryohei Terauchi, Sophien Kamoun, and Mark J Banfield. Functional diversification gave rise to allelic specialization in a rice NLR immune receptor pair. *eLife*, 10:e71662, November 2021. ISSN 2050-084X. doi: 10.7554/eLife.71662.

Athos Silva De Oliveira, Ivo Koolhaas, Leonardo Silva Boiteux, Octav F. Caldararu, Andrei-Jose Petrescu, Renato Oliveira Resende, and Richard Kormelink. Cell death

triggering and effector recognition by sw-5 SD-CNL proteins from resistant and susceptible tomato isolines to *Tomato spotted wilt virus*. *Molecular Plant Pathology*, 17 (9):1442–1454, December 2016. ISSN 1464-6722, 1364-3703. doi: 10.1111/mpp.12439.

Elke De Schutter, Ria Roelandt, Franck B. Riquet, Guy Van Camp, Andy Wullaert, and Peter Vandenabeele. Punching holes in cellular membranes: Biology and evolution of gasdermins. *Trends in Cell Biology*, 31(6):500–513, June 2021. ISSN 09628924. doi: 10.1016/j.tcb.2021.03.004.

Morris Degen, José Carlos Santos, Kristyna Pluhackova, Gonzalo Cebrero, Saray Ramos, Gytis Jankevicius, Ella Hartenian, Undina Guillerm, Stefania A. Mari, Bastian Kohl, Daniel J. Müller, Paul Schanda, Timm Maier, Camilo Perez, Christian Sieben, Petr Broz, and Sebastian Hiller. Structural basis of NINJ1-mediated plasma membrane rupture in cell death. *Nature*, 618(7967):1065–1071, June 2023. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-023-05991-z.

James F. Denton, Jose Lugo-Martinez, Abraham E. Tucker, Daniel R. Schrider, Wesley C. Warren, and Matthew W. Hahn. Extensive error in the number of genes inferred from draft genome assemblies. *PLOS Computational Biology*, 10(12):e1003998, December 2014. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1003998.

Sebastian Deorowicz, Agnieszka Debudaj-Grabysz, and Adam Gudyś. FAMSA: Fast and accurate multiple sequence alignment of huge protein families. *Scientific Reports*, 6(1): 33964, September 2016. ISSN 2045-2322. doi: 10.1038/srep33964.

Lida Derevnina, Sophien Kamoun, and Chih-hang Wu. Dude, where is my mutant? *Nicotiana benthamiana* meets forward genetics. *New Phytologist*, 221(2):607–610, January 2019. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.15521.

Lida Derevnina, Mauricio P. Contreras, Hiroaki Adachi, Jessica Upson, Angel Vergara Cruces, Rongrong Xie, Jan Sklenar, Frank L. H. Menke, Sam T. Mugford, Dan MacLean, Wenbo Ma, Saskia Hogenhout, Aska Goverse, Abbas Maqbool, Chih-Hang Wu, and Sophien Kamoun. Plant pathogens convergently evolved to counteract redundant nodes of an NLR immune receptor network. *bioRxiv*, page 2021.02.03.429184, February 2021. doi: 10.1101/2021.02.03.429184.

Anne Dievart, Céline Gottin, Christophe Périn, Vincent Ranwez, and Nathalie Chantret. Origin and diversity of plant receptor-like kinases. *Annual Review of Plant Biology*, 71(1):131–156, April 2020. ISSN 1543-5008, 1545-2123. doi: 10.1146/annurev-arplant-073019-025927.

Jingjin Ding, Kun Wang, Wang Liu, Yang She, Qi Sun, Jianjin Shi, Hanzi Sun, Da-Cheng Wang, and Feng Shao. Pore-forming activity and structural autoinhibition of the

gasdermin family. *Nature*, 535(7610):111–116, July 2016. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature18590.

Peter N Dodds, Jian Chen, and Megan A Outram. Pathogen perception and signaling in plant immunity. *Plant Cell*, 36(5):1465–1481, May 2024. ISSN 1040-4651, 1532-298X. doi: 10.1093/plcell/koae020.

Oliver Xiaoou Dong, Kevin Ao, Fang Xu, Kaeli C. M. Johnson, Yuxiang Wu, Lin Li, Shitou Xia, Yanan Liu, Yan Huang, Eleazar Rodriguez, Xuejin Chen, She Chen, Yuelin Zhang, Morten Petersen, and Xin Li. Individual components of paired typical NLR immune receptors are regulated by distinct E3 ligases. *Nature Plants*, 4(9):699–710, August 2018. ISSN 2055-0278. doi: 10.1038/s41477-018-0216-8.

Joram A. Dongus, Deepak D. Bhandari, Eva Penner, Dmitry Lapin, Sara C. Stolze, Anne Harzen, Monika Patel, Lani Archer, Lucas Dijkgraaf, Jyoti Shah, Hirofumi Nakagami, and Jane E. Parker. Cavity surface residues of PAD4 and SAG101 contribute to EDS1 dimer signaling specificity in plant immunity. *Plant Journal*, 110(5):1415–1432, April 2022. ISSN 1365-313X. doi: 10.1111/tpj.15747.

Cian Duggan, Eleonora Moratto, Zachary Savage, Eranthika Hamilton, Hiroaki Adachi, Chih-Hang Wu, Alexandre Y. Leary, Yasin Tumtas, Stephen M. Rothery, Abbas Maqbool, Seda Nohut, Toby Ross Martin, Sophien Kamoun, and Tolga Osman Bozkurt. Dynamic localization of a helper NLR at the plant–pathogen interface underpins pathogen recognition. *Proceedings of the National Academy of Sciences*, 118(34): e2104997118, August 2021. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2104997118.

Roland L. Dunbrack. *Rēs ipSAE loquuntur*: What's wrong with AlphaFold's *ipTM* score and how to fix it, February 2025.

Zane Duxbury, Shanshan Wang, Craig I. MacKenzie, Jeannette L. Tenthorey, Xiaoxiao Zhang, Sung Un Huh, Lanxi Hu, Lionel Hill, Pok Man Ngou, Pingtao Ding, Jian Chen, Yan Ma, Hailong Guo, Baptiste Castel, Panagiotis N. Moschou, Maud Bernoux, Peter N. Dodds, Russell E. Vance, and Jonathan D. G. Jones. Induced proximity of a TIR signaling domain on a plant-mammalian NLR chimera activates defense in plants. *Proceedings of the National Academy of Sciences*, 117(31):18832–18839, August 2020. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2001185117.

Zane Duxbury, Chih-hang Wu, and Pingtao Ding. A comparative overview of the intracellular guardians of plants and animals: NLRs in innate immunity and beyond. *Annual Review of Plant Biology*, 72(1):155–184, June 2021. ISSN 1543-5008, 1545-2123. doi: 10.1146/annurev-arplant-080620-104948.

Witold Dyrka, Virginie Coustou, Asen Daskalov, Alons Lends, Thierry Bardin, Mélanie Berbon, Brice Kauffmann, Corinne Blancard, Bénédicte Salin, Antoine Loquet, and Sven J. Saupe. Identification of NLR-associated amyloid signaling motifs in bacterial genomes. *Journal of Molecular Biology*, 432(23):6005–6027, November 2020. ISSN 00222836. doi: 10.1016/j.jmb.2020.10.004.

Sean R. Eddy. Accelerated profile HMM searches. *PLOS Computational Biology*, 7(10): e1002195, October 2011. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002195.

Kow Essuman, Jeffrey Milbrandt, Jeffery L. Dangl, and Marc T. Nishimura. Shared TIR enzymatic functions regulate cell death and immunity across the tree of life. *Science*, 377(6605):eabo0001, July 2022. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abo0001.

Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, and Demis Hassabis. Protein complex prediction with AlphaFold-multimer, October 2021.

Joanna M Feehan, Baptiste Castel, Adam R Bentham, and Jonathan Dg Jones. Plant NLRs get by with a little help from their friends. *Current Opinion in Plant Biology*, 56(Cc): 99–108, August 2020. ISSN 13695266. doi: 10.1016/j.pbi.2020.04.006.

Joanna M. Feehan, Junli Wang, Xinhua Sun, Jihyeon Choi, Hee-Kyung Ahn, Bruno Pok Man Ngou, Jane E. Parker, and Jonathan D. G. Jones. Oligomerization of a plant helper NLR requires cell-surface and intracellular immune receptor activation. *Proceedings of the National Academy of Sciences*, 120(11):e2210406120, March 2023. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2210406120.

Xing-Yu Feng, Qian Li, Yang Liu, Yan-Mei Zhang, and Zhu-Qing Shao. Evolutionary and immune-activating character analyses of NLR genes in algae suggest the ancient origin of plant intracellular immune receptors. *Plant Journal*, 119(5):2316–2330, July 2024. ISSN 1365-313X. doi: 10.1111/tpj.16919.

Alexander Förderer, Ertong Li, Aaron W. Lawson, Ya-nan Deng, Yue Sun, Elke Logemann, Xiaoxiao Zhang, Jie Wen, Zhifu Han, Junbiao Chang, Yuhang Chen, Paul Schulze-Lefert, and Jijie Chai. A wheat resistosome defines common principles of immune receptor channels. *Nature*, 610(7932):532–539, October 2022. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-022-05231-w.

Donald T. Fox, Douglas E. Soltis, Pamela S. Soltis, Tia-Lynn Ashman, and Yves Van De Peer. Polyploidy: A biological force from cells to ecosystems. *Trends in Cell Biology*, 30(9): 688–694, September 2020. ISSN 09628924. doi: 10.1016/j.tcb.2020.06.006.

Jianing Fu, Kate Schroder, and Hao Wu. Mechanistic insights from inflammasome structures. *Nature Reviews Immunology*, 24(7):518–535, July 2024. ISSN 1474-1733, 1474-1741. doi: 10.1038/s41577-024-00995-w.

Limin Fu, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23):3150–3152, December 2012. ISSN 1367-4803, 1367-4811. doi: 10.1093/bioinformatics/bts565.

Koki Fujisaki, Yoshiko Abe, Akiko Ito, Hiromasa Saitoh, Kentaro Yoshida, Hiroyuki Kanzaki, Eiko Kanzaki, Hiroe Utsushi, Tetsuro Yamashita, Sophien Kamoun, and Ryohei Terauchi. Rice Exo70 interacts with a fungal effector , AVR -pii, and is required for AVR -pii-triggered immunity. *Plant Journal*, 83(5):875–887, September 2015. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.12934.

Lars Gabriel, Felix Becker, Katharina J Hoff, and Mario Stanke. Tiberius: End-to-end deep learning with an HMM for gene prediction. *Bioinformatics*, 40(12):btae685, November 2024. ISSN 1367-4811. doi: 10.1093/bioinformatics/btae685.

Johannes Gantner, Jana Ordon, Carola Kretschmer, Raphaël Guerois, and Johannes Stuttmann. An EDS1-SAG101 complex is essential for TNL-mediated immunity in *Nicotiana benthamiana*. *Plant Cell*, 31(10):2456–2474, October 2019. ISSN 1040-4651, 1532-298X. doi: 10.1105/tpc.19.00099.

Linyi Alex Gao, Max E. Wilkinson, Jonathan Strecker, Kira S. Makarova, Rhiannon K. Macrae, Eugene V. Koonin, and Feng Zhang. Prokaryotic innate immunity through pattern recognition of conserved viral proteins. *Science*, 377(6607):eabm4096, August 2022. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abm4096.

Maria Laura Gaspar and Teresa E. Pawlowska. Innate immunity in fungi: Is regulated cell death involved? *PLOS Pathogens*, 18(5):e1010460, May 2022. ISSN 1553-7374. doi: 10.1371/journal.ppat.1010460.

Dongdong Ge, Fausto Andres Ortiz-Morea, Yingpeng Xie, In-Cheol Yeo, Qiaochu Shen, Yulu Zhou, Guangchao Liu, Liang Kong, Libo Shan, and Ping He. Assembly of helper NLR resistosome clusters upon activation of a coiled-coil NLR. *Nature*, 652(8108):251–258, March 2026. ISSN 1476-4687. doi: 10.1038/s41586-026-10215-1.

Foong-Jing Goh, Ching-Yi Huang, Lida Derevnina, and Chih-Hang Wu. NRC immune receptor networks show diversified hierarchical genetic architecture across plant

lineages. *Plant Cell*, 36(9):3399–3418, September 2024. ISSN 1040-4651, 1532-298X. doi: 10.1093/plcell/koae179.

Diana Gómez De La Cruz, Thomas Ingram, Rafał Zdrzałek, Jodie Taylor, Aleksandra Wawryk-Khamdavong, Kinga Bachowska, Mark J. Banfield, Nicholas J. Talbot, and Matthew J. Moscou. Molecular mimicry of a pathogen virulence target by a plant immune receptor. *Science*, 392(6802):1050–1055, June 2026. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aef9946.

Yihan Gong, Lei Tian, Ilias Kontos, Josh Li, and Xin Li. Plant immune signaling network mediated by helper NLRs. *Current Opinion in Plant Biology*, 73:102354, June 2023. ISSN 13695266. doi: 10.1016/j.pbi.2023.102354.

David M. Goodstein, Shengqiang Shu, Russell Howson, Rochak Neupane, Richard D. Hayes, Joni Fazo, Therese Mitros, William Dirks, Uffe Hellsten, Nicholas Putnam, and Daniel S. Rokhsar. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research*, 40(D1):D1178–D1186, January 2012. ISSN 1362-4962, 0305-1048. doi: 10.1093/nar/gkr944.

Jason Greenwald, Carolin Buhtz, Christiane Ritter, Witek Kwiatkowski, Senyon Choe, Marie-Lise Maddelein, Frederique Ness, Sandra Cescau, Alice Soragni, Dominik Leitz, Sven J. Saupe, and Roland Riek. The mechanism of prion inhibition by HET-S. *Molecular Cell*, 38(6):889–899, June 2010. ISSN 10972765. doi: 10.1016/j.molcel.2010.05.019.

Benguo Gu, Toby Parkes, Fernando Rabanal, Caroline Smith, Fu-Hao Lu, Neil McKenzie, Hui Dong, Detlef Weigel, Jonathan D.G. Jones, Volkan Cevik, and Michael W. Bevan. The integrated LIM-peptidase domain of the CSA1-CHS3/DAR4 paired immune receptor detects changes in DA1 peptidase inhibitors in arabidopsis. *Cell Host & Microbe*, 31(6):949–961.e5, June 2023. ISSN 19313128. doi: 10.1016/j.chom.2023.04.009.

Guanghao Guo, Kaihong Bai, Yikun Hou, Zhen Gong, Huaizhi Zhang, Qiuhong Wu, Ping Lu, Miaomiao Li, Lingli Dong, Jingzhong Xie, Yongxing Chen, Panpan Zhang, Keyu Zhu, Beibei Li, Wenling Li, Lei Dong, Yijun Yang, Dan Qiu, Gaojie Wang, Hee-Kyung Ahn, He Zhao, Chengguo Yuan, Wenqi Shi, Minfeng Xue, Lijun Yang, Dazao Yu, Yusheng Zhao, Yuhang Chen, Hongjie Li, Tiezhu Hu, Guan-Zhu Han, Jonathan D G Jones, and Zhiyong Liu. The wheat NLR pair RXL/Pm5e confers resistance to powdery mildew. *Plant Biotechnology Journal*, 23(4):1260–1276, 2025. ISSN 1467-7652. doi: 10.1111/pbi.14584.

Guanghao Guo, He Zhao, Kaihong Bai, Jian Lu, Qiuhong Wu, Lei Lu, Yue Zhang, Lingli Dong, Guangwei Li, Yongxing Chen, Yikun Hou, Ping Lu, Miaomiao Li, Huaizhi Zhang, Gaojie Wang, Keyu Zhu, Baoge Huang, Xuejia Cui, Hongkui Fu, Chenchen Hu, Zhiying

Chu, Xue Lyu, Sophien Kamoun, Chao Wang, Zhiyong Liu, Muniyandi Selvaraj, and Jonathan D.G. Jones. An activated wheat CCG10-NLR immune receptor forms an octameric resistosome. *Cell*, 189(10):2955–2970.e16, May 2026. ISSN 00928674. doi: 10.1016/j.cell.2026.02.024.

Jean Hausser and Korbinian Strimmer. Entropy inference and the james-stein estimator, with application to nonlinear gene association networks. *Journal of Machine Learning Research*, 10(50):1469–1484, 2009. ISSN 1533-7928. doi: 10.5555/1577069.1755833.

Magnus Haraldson Høie, Erik Nicolas Kiehl, Bent Petersen, Morten Nielsen, Ole Winther, Henrik Nielsen, Jeppe Hallgren, and Paolo Marcatili. NetSurfP-3.0: Accurate and fast prediction of protein structural features by protein language models and deep learning. *Nucleic Acids Research*, 50(W1):W510–W515, July 2022. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkac439.

Felix Holst, Anthony Bolger, Christopher Günther, Janina Maß, Sebastian Triesch, Felicitas Kindel, Niklas Kiel, Nima Saadat, Oliver Ebenhöh, Björn Usadel, Rainer Schwacke, Marie Bolger, Andreas P.M. Weber, and Alisandra K. Denton. Helixer– *de novo* prediction of primary eukaryotic gene models combining deep learning and a hidden markov model, February 2023.

Shane Horsefield, Hayden Burdett, Xiaoxiao Zhang, Mohammad K. Manik, Yun Shi, Jian Chen, Tiancong Qi, Jonathan Gilley, Jhih-Siang Lai, Maxwell X. Rank, Lachlan W. Casey, Weixi Gu, Daniel J. Ericsson, Gabriel Foley, Robert O. Hughes, Todd Bosanac, Mark von Itzstein, John P. Rathjen, Jeffrey D. Nanson, Mikael Boden, Ian B. Dry, Simon J. Williams, Brian J. Staskawicz, Michael P. Coleman, Thomas Ve, Peter N. Dodds, and Bostjan Kobe. NAD + cleavage activity by animal and plant TIR domains in cell death pathways. *Science*, 365(6455):793–799, August 2019. ISSN 1095-9203. doi: 10.1126/science.aax1911.

Zehan Hu and Jijie Chai. Assembly and architecture of NLR resistosomes and inflammasomes. *Annual Review of Biophysics*, 52(1):207–228, May 2023. ISSN 1936-122X, 1936-1238. doi: 10.1146/annurev-biophys-092922-073050.

Zehan Hu, Qiang Zhou, Chenlu Zhang, Shilong Fan, Wei Cheng, Yue Zhao, Feng Shao, Hong-Wei Wang, Sen-Fang Sui, and Jijie Chai. Structural and biochemical basis for induced self-propagation of NLRC4. *Science*, 350(6259):399–404, October 2015. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aac5489.

Shijia Huang, Aolin Jia, Wen Song, Giuliana Hessler, Yonggang Meng, Yue Sun, Lina Xu, Henriette Laessle, Jan Jirschitzka, Shoucai Ma, Yu Xiao, Dongli Yu, Jiao Hou, Ruiqi Liu, Huanhuan Sun, Xiaohui Liu, Zhifu Han, Junbiao Chang, Jane E. Parker, and Jijie

Chai. Identification and receptor mechanism of TIR-catalyzed small molecules in plant immunity. *Science*, 377(6605):eabq3297, July 2022. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abq3297.

Shijia Huang, Ertong Li, Fangshuai Jia, Zhifu Han, and Jijie Chai. Assembly and functional mechanisms of plant NLR resistosomes. *Current Opinion in Structural Biology*, 90: 102977, February 2025. ISSN 0959-440X. doi: 10.1016/j.sbi.2024.102977.

Daniel H. Huson and Celine Scornavacca. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Systematic Biology*, 61(6):1061–1067, December 2012. ISSN 1076-836X, 1063-5157. doi: 10.1093/sysbio/sys062.

Tarhan Ibrahim, Virendrasinh Khandare, Federico Gabriel Mirkin, Yasin Tumtas, Doryen Bubeck, and Tolga O. Bozkurt. AlphaFold2-multimer guided high-accuracy prediction of typical and atypical ATG8-binding motifs. *PLOS Biology*, 21(2):e3001962, February 2023. ISSN 1545-7885. doi: 10.1371/journal.pbio.3001962.

Tarhan Ibrahim, Freddie J. King, AmirAli Toghani, Luyao Wang, Saskia Jenkins, Enoch Lok Him Yuen, Hung-Yu Wang, Cristina Vuolo, Nick Eilmann, Vanda Adamkova, Khong-Sam Chia, Baptiste Castel, Jonathan D. G. Jones, Philip Carella, Chih-Hang Wu, Jiorgos Kourelis, Sophien Kamoun, and Tolga O. Bozkurt. A helper NLR channels organellar calcium to trigger plant immunity. *Science*, 392(6797):499–505, April 2026. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aeb6690.

Pierre Jacob, Nak Hyun Kim, Feihua Wu, Farid El-Kasmi, Yuan Chi, William G. Walton, Oliver J. Furzer, Adam D. Lietzan, Sruthi Sunil, Korina Kempthorn, Matthew R. Redinbo, Zhen-Ming Pei, Li Wan, and Jeffery L. Dangl. Plant "helper" immune receptors are Ca^{2+} -permeable nonselective cation channels. *Science*, 373(6553):420–425, July 2021. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abg7917.

Aolin Jia, Shijia Huang, Wen Song, Junli Wang, Yonggang Meng, Yue Sun, Lina Xu, Henriette Laessle, Jan Jirschitzka, Jiao Hou, Tiantian Zhang, Wenquan Yu, Giuliana Hessler, Ertong Li, Shoucai Ma, Dongli Yu, Jan Gebauer, Ulrich Baumann, Xiaohui Liu, Zhifu Han, Junbiao Chang, Jane E. Parker, and Jijie Chai. TIR-catalyzed ADP-ribosylation reactions produce signaling molecules for plant immunity. *Science*, 377(6605):eabq8180, July 2022. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abq8180.

Aolin Jia, Shijia Huang, Shoucai Ma, Xiaoyu Chang, Zhifu Han, and Jijie Chai. TIR-catalyzed nucleotide signaling molecules in plant defense. *Current Opinion in Plant Biology*, 73:102334, June 2023. ISSN 13695266. doi: 10.1016/j.pbi.2022.102334.

Alex G. Johnson and Philip J. Kranzusch. What bacterial cell death teaches us about life. *PLOS Pathogens*, 18(10):e1010879, October 2022. ISSN 1553-7374. doi: 10.1371/journal.ppat.1010879.

Alex G. Johnson, Megan L. Mayer, Stefan L. Schaefer, Nora K. McNamara-Bordewick, Gerhard Hummer, and Philip J. Kranzusch. Structure and assembly of a bacterial gasdermin pore. *Nature*, 628(8008):657–663, April 2024. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-024-07216-3.

Jonathan D. G. Jones and Jeffery L. Dangl. The plant immune system. *Nature*, 444(7117):323–329, November 2006. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature05286.

Jonathan D. G. Jones, Russell E. Vance, and Jeffery L. Dangl. Intracellular innate immune surveillance devices in plants and animals. *Science*, 354(6316):aaf6395, December 2016. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aaf6395.

Lance M Jubic, Svenja Saile, Oliver J Furzer, Farid El Kasmi, and Jeffery L Dangl. Help wanted: Helper NLRs and plant immune responses. *Current Opinion in Plant Biology*, 50(Id):82–94, August 2019. ISSN 13695266. doi: 10.1016/j.pbi.2019.03.013.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-021-03819-2.

Florian Jupe, Leighton Pritchard, Graham J Etherington, Katrin MacKenzie, Peter Ja Cock, Frank Wright, Sanjeev Kumar Sharma, Dan Bolser, Glenn J Bryan, Jonathan Dg Jones, and Ingo Hein. Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genomics*, 13(1):75–88, December 2012. ISSN 1471-2164. doi: 10.1186/1471-2164-13-75.

K. Katoh and D. M. Standley. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4):772–780, April 2013. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/mst010.

Yousuf A. Khan, K. Ian White, and Axel T. Brunger. The AAA+ superfamily: A review of the structural and mechanistic principles of these molecular machines. *Critical Reviews in*

Biochemistry and Molecular Biology, 57(2):156–187, March 2022. ISSN 1040-9238, 1549-7798. doi: 10.1080/10409238.2021.1979460.

Emily M. Kibby, Amy N. Conte, A. Maxwell Burroughs, Toni A. Nagy, Jose A. Vargas, Lindsay A. Whalen, L. Aravind, and Aaron T. Whiteley. Bacterial NLR-related proteins protect against phage. *Cell*, 186(11):2410–2424.e18, May 2023. ISSN 00928674. doi: 10.1016/j.cell.2023.04.015.

Saet-Byul Kim, Hye-Young Lee, Seungyeon Seo, Joo Hyun Lee, and Doil Choi. RNA-dependent RNA polymerase (NIb) of the potyviruses is an avirulence factor for the broad-spectrum resistance gene Pvr4 in capsicum annuum cv. CM334. *PLOS One*, 10 (3):e0119639, March 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0119639.

Seungill Kim, Jieun Park, Seon-In Yeom, Yong-Min Kim, Eunyoung Seo, Ki-Tae Kim, Myung-Shin Kim, Je Min Lee, Kyeongchae Cheong, Ho-Sub Shin, Saet-Byul Kim, Koeun Han, Jundae Lee, Minkyu Park, Hyun-Ah Lee, Hye-Young Lee, Youngsill Lee, Soohyun Oh, Joo Hyun Lee, Eunhye Choi, Eunbi Choi, So Eui Lee, Jongbum Jeon, Hyunbin Kim, Gobong Choi, Hyeunjeong Song, JunKi Lee, Sang-Choon Lee, Jin-Kyung Kwon, Hea-Young Lee, Namjin Koo, Yunji Hong, Ryan W. Kim, Won-Hee Kang, Jin Hoe Huh, Byoung-Cheorl Kang, Tae-Jin Yang, Yong-Hwan Lee, Jeffrey L. Bennetzen, and Doil Choi. New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biology*, 18(1):210–220, December 2017. ISSN 1474-760X. doi: 10.1186/s13059-017-1341-9.

Jiorgos Kourelis and Renier A.L. Van Der Hoorn. Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. *Plant Cell*, 30(2):285–299, February 2018. ISSN 1040-4651, 1532-298X. doi: 10.1105/tpc.17.00579.

Jiorgos Kourelis, Toshiyuki Sakai, Hiroaki Adachi, and Sophien Kamoun. RefPlantNLR is a comprehensive collection of experimentally validated plant disease resistance proteins from the NLR family. *PLOS Biology*, 19(10):e3001124, October 2021. ISSN 1545-7885. doi: 10.1371/journal.pbio.3001124.

Jiorgos Kourelis, Mauricio P. Contreras, Adeline Harant, Hsuan Pai, Daniel Lüdke, Hiroaki Adachi, Lida Derevnina, Chih-Hang Wu, and Sophien Kamoun. The helper NLR immune protein NRC3 mediates the hypersensitive cell death caused by the cell-surface receptor Cf-4. *PLOS Genetics*, 18(9):e1010414, September 2022. ISSN 1553-7404. doi: 10.1371/journal.pgen.1010414.

Jiorgos Kourelis, Clemence Marchal, Andres Posbeyikian, Adeline Harant, and Sophien Kamoun. NLR immune receptor–nanobody fusions confer plant disease resistance. *Science*, 379(6635):934–939, March 2023. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abn4116.

Thomas Kroj, Emilie Chanclud, Corinne Michel-Romiti, Xavier Grand, and Jean-Benoit Morel. Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytologist*, 210(2):618–626, April 2016. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.13869.

Hanhui Kuang, Sung-Sick Woo, Blake C. Meyers, Eviatar Nevo, and Richard W. Michelmore. Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce[W]. *Plant Cell*, 16(11):2870–2894, November 2004. ISSN 1532-298X. doi: 10.1105/tpc.104.025502.

Sudhir Kumar, Michael Suleski, Jack M Craig, Adrienne E Kasprowicz, Maxwell Sanderford, Michael Li, Glen Stecher, and S Blair Hedges. TimeTree 5: An expanded resource for species divergence times. *Molecular Biology and Evolution*, 39(8):msac174, August 2022. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/msac174.

Dmitry Lapin, Viera Kovacova, Xinhua Sun, Joram A. Dongus, Deepak Bhandari, Patrick Von Born, Jaqueline Bautor, Nina Guarneri, Jakub Rzemieniewski, Johannes Stuttmann, Andreas Beyer, and Jane E. Parker. A coevolved EDS1-SAG101-NRG1 module mediates cell death signaling by TIR-domain immune receptors. *Plant Cell*, 31(10):2430–2455, October 2019. ISSN 1040-4651, 1532-298X. doi: 10.1105/tpc.19.00118.

Dmitry Lapin, Deepak D. Bhandari, and Jane E. Parker. Origins and immunity networking functions of EDS1 family proteins. *Annual Review of Phytopathology*, 58(1):253–276, August 2020. ISSN 1545-2107. doi: 10.1146/annurev-phyto-010820-012840.

Michael Lawrence, Robert Gentleman, and Vincent Carey. Rtracklayer: An R package for interfacing with genome browsers. *Bioinformatics*, 25(14):1841–1842, July 2009. ISSN 1367-4811, 1367-4803. doi: 10.1093/bioinformatics/btp328.

Pedro Leão, Mary E. Little, Kathryn E. Appler, Daphne Sahaya, Emily Aguilar-Pine, Kathryn Currie, Ilya J. Finkelstein, Valerie De Anda, and Brett J. Baker. Asgard archaea defense systems and their roles in the origin of eukaryotic immunity. *Nature Communications*, 15(1):6386, July 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-50195-2.

Hye-Young Lee, Hyunggon Mang, Eunhye Choi, Ye-Eun Seo, Myung-Shin Kim, Soohyun Oh, Saet-Byul Kim, and Doil Choi. Genome-wide functional analysis of hot pepper immune receptors reveals an autonomous NLR clade in seed plants. *New Phytologist*, 229(1):532–547, January 2021. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.16878.

Hye-Young Lee, Jeen Choi, Minji Kang, Joo Hyun Lee, Myung-Shin Kim, and Doil Choi. Protein stability governed by α1–2 helices in Pvr4 is essential for localization and cell

death. *Plant Journal*, 116(2):510–523, October 2023. ISSN 0960-7412, 1365-313X. doi: 10.1111/tpj.16388.

Rachelle R.Q. Lee and Eunyoung Chae. Variation patterns of NLR clusters in arabidopsis thaliana genomes. *Plant Communications*, 1(4):100089, July 2020. ISSN 25903462. doi: 10.1016/j.xplc.2020.100089.

Detlef D. Leipe, Eugene V. Koonin, and L. Aravind. STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: Multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *Journal of Molecular Biology*, 343(1):1–28, October 2004. ISSN 0022-2836. doi: 10.1016/j.jmb.2004.08.023.

Ning Li, Qiang He, Juan Wang, Baike Wang, Jiantao Zhao, Shaoyong Huang, Tao Yang, Yaping Tang, Shengbao Yang, Patiguli Aisimutuola, Ruiqiang Xu, Jiahui Hu, Chunping Jia, Kai Ma, Zhiqiang Li, Fangling Jiang, Jie Gao, Haiyan Lan, Yongfeng Zhou, Xinyan Zhang, Sanwen Huang, Zhangjun Fei, Huan Wang, Hongbo Li, and Qinghui Yu. Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nature Genetics*, 55(5):852–860, May 2023. ISSN 1061-4036, 1546-1718. doi: 10.1038/s41588-023-01340-y.

Pingchuan Li, Xiande Quan, Gaofeng Jia, Jin Xiao, Sylvie Cloutier, and Frank M. You. RGAugury: A pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. *BMC Genomics*, 17(1):852–861, December 2016. ISSN 1471-2164. doi: 10.1186/s12864-016-3197-x.

Sai-Xi Li, Yang Liu, Yan-Mei Zhang, Jian-Qun Chen, and Zhu-Qing Shao. Convergent reduction of immune receptor repertoires during plant adaptation to diverse special lifestyles and habitats. *Nature Plants*, 11(2):248–262, January 2025. ISSN 2055-0278. doi: 10.1038/s41477-024-01901-x.

Pei-Yu Lin, Shiang-Chin Huang, Kuan-Lin Chen, Yu-Chun Huang, Chia-Yu Liao, Guan-Jun Lin, HueyTyng Lee, and Pao-Yang Chen. Analysing protein complexes in plant science: Insights and limitation with AlphaFold 3. *Botanical Studies*, 66(1):14–25, May 2025. ISSN 1999-3110. doi: 10.1186/s40529-025-00462-2.

Xiao Lin, Andrea Olave-Achury, Robert Heal, Kamil Witek, Hari S. Karki, Tianqiao Song, Chih-hang Wu, Hiroaki Adachi, Sophien Kamoun, Vivianne G. A. A. Vleeshouwers, and Jonathan D. G. Jones. *Rpi-Amr3* confers resistance to multiple *Phytophthora* species by recognizing a conserved RXLR effector, June 2021.

Furong Liu, Zhenlin Yang, Chao Wang, Raoul Martin, Wenjie Qiao, Jan E. Carette, Sheng

Luan, Eva Nogales, and Brian Staskawicz. The activated plant NRC4 immune receptor forms a hexameric resistosome, December 2023a.

Yang Liu, Zhen Zeng, Yan-Mei Zhang, Qian Li, Xing-Mei Jiang, Zhen Jiang, Ji-Hong Tang, Dijun Chen, Qiang Wang, Jian-Qun Chen, and Zhu-Qing Shao. An angiosperm NLR atlas reveals that NLR gene reduction is associated with ecological specialization and signal transduction component deletion. *Molecular Plant*, 14(12):2015–2031, December 2021. ISSN 16742052. doi: 10.1016/j.molp.2021.08.001.

Yang Liu, Yan-Mei Zhang, Yao Tang, Jian-Qun Chen, and Zhu-Qing Shao. The evolution of plant NLR immune receptors and downstream signal components. *Current Opinion in Plant Biology*, 73:102363, June 2023b. ISSN 13695266. doi: 10.1016/j.pbi.2023.102363.

Zhenyu Liu, Jorunn I. B. Bos, Miles Armstrong, Stephen C. Whisson, Luis Da Cunha, Trudy Torto-Alalibo, Joe Win, Anna O. Avrova, Frank Wright, Paul R. J. Birch, and Sophien Kamoun. Patterns of diversifying selection in the phytotoxin-like scr74 gene family of phytophthora infestans. *Molecular Biology and Evolution*, 22(3):659–672, March 2005. ISSN 1537-1719, 0737-4038. doi: 10.1093/molbev/msi049.

Zong-Yan Liu, Ana Berthel, Eric Czech, Edgar Marroquin, Michelle Stitzer, Sheng-Kai Hsu, Matt Pennell, Edward S. Buckler, and Jingjing Zhai. GeneCAD: Plant genome annotation with a DNA foundation model, November 2025. ISSN 2692-8205.

Daniel Lüdke, Toshiyuki Sakai, Jiorgos Kourelis, AmirAli Toghani, Hiroaki Adachi, Andrés Posbeyikian, Raoul Frijters, Hsuan Pai, Adeline Harant, Karin Ernst, Martin Ganal, Adriaan Verhage, Chih-Hang Wu, and Sophien Kamoun. A root-specific NLR network confers resistance to plant parasitic nematodes, December 2023.

Daniel Lüdke, Toshiyuki Sakai, Jiorgos Kourelis, AmirAli Toghani, Hiroaki Adachi, Andrés Posbeyikian, Raoul Frijters, Hsuan Pai, Adeline Harant, Juan Carlos Lopez-Agudelo, Bozeng Tang, Karin Ernst, Martin Ganal, Adriaan Verhage, Chih-Hang Wu, and Sophien Kamoun. A root-specific NLR network mediates immune signaling of resistance genes against plant parasitic nematodes. *Plant Cell*, 37(7):koaf145, July 2025. ISSN 1040-4651, 1532-298X. doi: 10.1093/plcell/koaf145.

Shoucai Ma, Dmitry Lapin, Li Liu, Yue Sun, Wen Song, Xiaoxiao Zhang, Elke Logemann, Dongli Yu, Jia Wang, Jan Jirschitzka, Zhifu Han, Paul Schulze-Lefert, Jane E. Parker, and Jijie Chai. Direct pathogen-induced assembly of an NLR immune receptor complex to form a holoenzyme. *Science*, 370(6521):eabe3069, December 2020. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abe3069.

Shoucai Ma, Chunpeng An, Aaron W. Lawson, Yu Cao, Yue Sun, Eddie Yong Jun Tan, Jinheng Pan, Jan Jirschitzka, Florian Kümmel, Nitika Mukhi, Zhifu Han, Shan Feng,

Bin Wu, Paul Schulze-Lefert, and Jijie Chai. Oligomerization-mediated autoinhibition and cofactor binding of a plant NLR. *Nature*, 632(8026):869–876, August 2024. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-024-07668-7.

Yan Ma, Hailong Guo, Lanxi Hu, Paula Pons Martinez, Panagiotis N. Moschou, Volkan Cevik, Pingtao Ding, Zane Duxbury, Panagiotis F. Sarris, and Jonathan D. G. Jones. Distinct modes of derepression of an *Arabidopsis* immune receptor complex by two different bacterial effectors. *Proceedings of the National Academy of Sciences*, 115(41): 10218–10227, October 2018. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1811858115.

Jogi Madhuprakash, AmirAli Toghani, Mauricio P. Contreras, Andres Posbeyikian, Jake Richardson, Jiorgos Kourelis, Tolga O. Bozkurt, Michael W. Webster, and Sophien Kamoun. A disease resistance protein triggers oligomerization of its NLR helper into a hexameric resistosome to mediate innate immunity. *Science Advances*, 10(45): eadr2594, November 2024. ISSN 2375-2548. doi: 10.1126/sciadv.adr2594.

Takaki Maekawa, Barbara Kracher, Isabel M. L. Saur, Makoto Yoshikawa-Maekawa, Ronny Kellner, Artem Pankin, Maria Von Korff, and Paul Schulze-Lefert. Subfamily-specific specialization of RGH1/MLA immune receptors in wild barley. *Molecular Plant-Microbe Interactions®*, 32(1):107–119, January 2019. ISSN 0894-0282, 1943-7706. doi: 10.1094/MPMI-07-18-0186-FI.

Takaki Maekawa, Hamid Kashkar, and Núria S. Coll. Dying in self-defence: A comparative overview of immunogenic cell death signalling in animals and plants. *Cell death and differentiation*, 30(2):258–268, February 2023. ISSN 1350-9047, 1476-5403. doi: 10.1038/s41418-022-01060-6.

Lisa K. Mahdi, Menghang Huang, Xiaoxiao Zhang, Ryohei Thomas Nakano, Leïla Brulé Kopp, Isabel M.L. Saur, Florence Jacob, Viera Kovacova, Dmitry Lapin, Jane E. Parker, James M. Murphy, Kay Hofmann, Paul Schulze-Lefert, Jijie Chai, and Takaki Maekawa. Discovery of a family of mixed lineage kinase domain-like proteins in plants and their role in innate immune signaling. *Cell Host & Microbe*, 28(6):813–824.e6, December 2020. ISSN 19313128. doi: 10.1016/j.chom.2020.08.012.

Mosè Manni, Matthew R Berkeley, Mathieu Seppey, Felipe A Simão, and Evgeny M Zdobnov. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution*, 38(10):4647–4654, September 2021. ISSN 1537-1719. doi: 10.1093/molbev/msab199.

A Maqbool, H Saitoh, M Franceschetti, Cem Stevenson, A Uemura, H Kanzaki, S Kamoun, R Terauchi, and Mj Banfield. Structural basis of pathogen recognition by an integrated

HMA domain in a plant NLR immune receptor. *eLife*, 4:e08709, August 2015. ISSN 2050-084X. doi: 10.7554/eLife.08709.

Clemence Marchal, Jianping Zhang, Peng Zhang, Paul Fenwick, Burkhard Steuernagel, Nikolai M. Adamski, Lesley Boyd, Robert McIntosh, Brande B. H. Wulff, Simon Berry, Evans Lagudah, and Cristobal Uauy. BED-domain-containing immune receptors confer diverse resistance spectra to yellow rust. *Nature Plants*, 4(9):662–668, August 2018. ISSN 2055-0278. doi: 10.1038/s41477-018-0236-4.

Rose A. Marks, Scott Hotaling, Paul B. Frandsen, and Robert VanBuren. Representation and participation across 20 years of plant genome sequencing. *Nature Plants*, 7(12): 1571–1578, November 2021. ISSN 2055-0278. doi: 10.1038/s41477-021-01031-8.

Raoul Martin, Tiancong Qi, Haibo Zhang, Furong Liu, Miles King, Claire Toth, Eva Nogales, and Brian J. Staskawicz. Structure of the activated ROQ1 resistosome directly recognizing the pathogen effector XopQ. *Science*, 370(6521):eabd9993, December 2020. ISSN 1095-9203. doi: 10.1126/science.abd9993.

Fabio Martinon, Kimberly Burns, and Jürg Tschopp. The inflammasome: A molecular platform triggering activation of inflammatory caspases and processing of proIL-β. *Molecular Cell*, 10(2):417–426, August 2002. ISSN 1097-2765. doi: 10.1016/S1097-2765(02)00599-3.

Natsumi Maruta, Mitchell Sorbello, and Bostjan Kobe. TIR-mediated immune signaling through the EDS1:PAD4:ADR1 node is conserved in monocots and dicots. *Molecular Plant*, 18(2):192–194, February 2025. ISSN 1674-2052. doi: 10.1016/j.molp.2024.12.017.

Natsumi Maruta, Weixi Gu, Bryan Ying Jie Lim, Mitchell Sorbello, Dalton Heng Yong Ngu, Chacko Jobichen, Jeffrey D. Nanson, Yan Li, Jian Chen, Megan A. Outram, Maud Bernoux, Motiur Rahman, Thi Danh Vu, Hongyi Xu, Lei Wang, Kayden Kwah, Hayden Burdett, Mehdi Mobli, Thomas Ve, Jeffrey G. Ellis, Peter A. Anderson, Simon J. Williams, Peter N. Dodds, and Bostjan Kobe. Structural basis of canonical TIR-NLR activation in plant innate immunity, February 2026. ISSN 2692-8205.

Conor J. McClune and Michael T. Laub. Constraints on the expansion of paralogous protein families. *Current Biology*, 30(10):R460–R464, May 2020. ISSN 09609822. doi: 10.1016/j.cub.2020.02.075.

Leah K. McHale, Maria José Truco, Alexander Kozik, Tadeusz Wroblewski, Oswaldo E. Ochoa, Kirsten A. Lahre, Steven J. Knapp, and Richard W. Michelmore. The genomic architecture of disease resistance in lettuce. *Theoretical and Applied Genetics*, 118(3):565–580, February 2009. ISSN 0040-5752, 1432-2242. doi: 10.1007/s00122-008-0921-1.

Anthony E Melton, Andrew W Child, Richard S Beard, Carlos Dave C Dumaguit, Jennifer S Forbey, Matthew Germino, Marie-Anne De Graaff, Andrew Kliskey, Ilia J Leitch, Peggy Martinez, Stephen J Novak, Jaume Pellicer, Bryce A Richardson, Desiree Self, Marcelo Serpe, and Sven Buerki. A haploid pseudo-chromosome genome assembly for a keystone sagebrush species of western north American rangelands. *G3 Genes|Genomes|Genetics*, 12(7):jkac122, July 2022. ISSN 2160-1836. doi: 10.1093/g3journal/jkac122.

Haydee Mesa-Galloso, Lohans Pedrera, and Uris Ros. Pore-forming proteins: From defense factors to endogenous executors of cell death. *Chemistry and Physics of Lipids*, 234:105026, January 2021. ISSN 00093084. doi: 10.1016/j.chemphyslip.2020.105026.

Blake C. Meyers, Doris B. Chin, Katherine A. Shen, Subramoniam Sivaramakrishnan, Dean O. Lavelle, Zhen Zhang, and Richard W. Michelmore. The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. *Plant Cell*, 10(11):1817–1832, November 1998. ISSN 1040-4651, 1532-298X. doi: 10.1105/tpc.10.11.1817.

Richard W. Michelmore and Blake C. Meyers. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research*, 8(11):1113–1130, November 1998. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.8.11.1113.

Byoungnam Min, Igor V Grigoriev, and In-Geol Choi. FunGAP: Fungal genome annotation pipeline using evidence-based gene model evaluation. *Bioinformatics*, 33(18):2936–2937, September 2017. ISSN 1367-4803, 1367-4811. doi: 10.1093/bioinformatics/btx353.

Michael Mindrinos, Fumiaki Katagiri, Guo-Liang Yu, and Frederick M. Ausubel. The a. thaliana disease resistance gene RPS2 encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell*, 78(6):1089–1099, September 1994. ISSN 00928674. doi: 10.1016/0092-8674(94)90282-8.

Bui Quang Minh, Heiko A Schmidt, Olga Chernomor, Dominik Schrempf, Michael D Woodhams, Arndt Von Haeseler, and Robert Lanfear. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5):1530–1534, May 2020. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/msaa015.

Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. ColabFold: Making protein folding accessible to all. *Nature Methods*, 19(6):679–682, June 2022. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-022-01488-1.

James M. Murphy, Peter E. Czabotar, Joanne M. Hildebrand, Isabelle S. Lucet, Jian-Guo Zhang, Silvia Alvarez-Diaz, Rowena Lewis, Najoua Lalaoui, Donald Metcalf, Andrew I. Webb, Samuel N. Young, Leila N. Varghese, Gillian M. Tannahill, Esme C. Hatchell, Ian J. Majewski, Toru Okamoto, Renwick C.J. Dobson, Douglas J. Hilton, Jeffrey J. Babon, Nicos A. Nicola, Andreas Strasser, John Silke, and Warren S. Alexander. The pseudokinase MLKL mediates necroptosis via a molecular switch mechanism. *Immunity*, 39(3):443–453, September 2013. ISSN 10747613. doi: 10.1016/j.immuni.2013.06.018.

M Nei and T Gojobori. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, 3(5):418–426, September 1986. ISSN 0737-4038. doi: 10.1093/oxfordjournals.molbev.a040410.

Kim Newton, Andreas Strasser, Nobuhiko Kayagaki, and Vishva M. Dixit. Cell death. *Cell*, 187(2):235–256, January 2024. ISSN 00928674. doi: 10.1016/j.cell.2023.11.044.

Bruno Pok Man Ngou, Hee-Kyung Ahn, Pingtao Ding, and Jonathan D. G. Jones. Mutual potentiation of plant immunity by cell-surface and intracellular receptors. *Nature*, 592(7852):110–115, April 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-021-03315-7.

Bruno Pok Man Ngou, Pingtao Ding, and Jonathan D G Jones. Thirty years of resistance: Zig-zag through the plant immune system. *Plant Cell*, 34(5):1447–1478, April 2022. ISSN 1040-4651, 1532-298X. doi: 10.1093/plcell/koac041.

Bruno Pok Man Ngou, Michele Wyler, Marc W. Schmid, Yasuhiro Kadota, and Ken Shirasu. Evolutionary trajectory of pattern recognition receptors in plants. *Nature Communications*, 15(1):308–329, February 2024. ISSN 2041-1723. doi: 10.1038/s41467-023-44408-3.

Gal Ofir, Ehud Herbst, Maya Baroz, Daniel Cohen, Adi Millman, Shany Doron, Nitzan Tal, Daniel B. A. Malheiro, Sergey Malitsky, Gil Amitai, and Rotem Sorek. Antiviral activity of bacterial TIR domains via immune signalling molecules. *Nature*, 600(7887):116–120, December 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-021-04098-7.

Soohyun Oh, Sejun Kim, Hyo-Jeong Park, Myung-Shin Kim, Min-Ki Seo, Chih-Hang Wu, Hyun-Ah Lee, Hyun-Soon Kim, Sophien Kamoun, and Doil Choi. Nucleotide-binding leucine-rich repeat network underlies nonhost resistance of pepper against the irish potato famine pathogen *Phytophthora infestans*. *Plant Biotechnology Journal*, 21(7):1361–1372, July 2023. ISSN 1467-7644, 1467-7652. doi: 10.1111/pbi.14039.

Diana Ortiz and Peter N. Dodds. Plant NLR origins traced back to green algae. *Trends in*

Plant Science, 23(8):651–654, August 2018. ISSN 13601385. doi: 10.1016/j.tplants.2018.05.009.

Hsuan Pai, Mauricio P. Contreras, Jose Salguero Linares, Daniel Lüdke, Andres Posbeyikian, Jiorgos Kourelis, Sophien Kamoun, and Clemence Marchal. The pik NLR pair accumulates at the plasma membrane as a hetero-oligomeric sensor-helper immune protein complex prior to activation, December 2025a. ISSN 2692-8205.

Hsuan Pai, Mauricio P. Contreras, Jose Salguero Linares, Daniel Lüdke, Andres Posbeyikian, Jiorgos Kourelis, Sophien Kamoun, and Clemence Marchal. The pik NLR pair accumulates at the plasma membrane as a hetero-oligomeric sensor-helper immune protein complex prior to activation, December 2025b. ISSN 2692-8205.

Hsuan Pai, Toshiyuki Sakai, Andres Posbeyikian, Raoul Frijters, Yu Sugihara, Mauricio P. Contreras, Jiorgos Kourelis, Hiroaki Adachi, Sophien Kamoun, and AmirAli Toghani. Contrasting evolutionary patterns of helper and sensor NRC NLRs in lettuce reflect functional divergence following subfunctionalization. *PLOS Genetics*, 22(7):e1012245, July 2026. ISSN 1553-7404. doi: 10.1371/journal.pgen.1012245.

Yuxuan Pang, Xiao-chen Bai, Chuangye Yan, Qi Hao, Zheqin Chen, Jia-Wei Wang, Sjors H.W. Scheres, and Yigong Shi. Structure of the apoptosome: Mechanistic insights into activation of an initiator caspase from *Drosophila*. *Genes & Development*, 29(3):277–287, February 2015. ISSN 0890-9369, 1549-5477. doi: 10.1101/gad.255877.114.

Geo Pertea and Mihaela Pertea. GFF utilities: GffRead and GffCompare. *F1000Research*, 9:304–323, April 2020. ISSN 2046-1402. doi: 10.12688/f1000research.23297.1.

Emma J. Petrie, Peter E. Czabotar, and James M. Murphy. The structural basis of necroptotic cell death signaling. *Trends in Biochemical Sciences*, 44(1):53–63, January 2019. ISSN 09680004. doi: 10.1016/j.tibs.2018.11.002.

Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Elaine C. Meng, Gregory S. Couch, Tristan I. Croll, John H. Morris, and Thomas E. Ferrin. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, 30(1):70–82, January 2021. ISSN 0961-8368, 1469-896X. doi: 10.1002/pro.3943.

Eugenia Pitsili, Ujjal J. Phukan, and Nuria S. Coll. Cell death in plant immunity. *Cold Spring Harbor Perspectives in Biology*, 12(6):a036483, October 2019. ISSN 1943-0264. doi: 10.1101/cshperspect.a036483.

Andrés Posbeyikian, AmirAli Toghani, Alexander J McClelland, Sophien Kamoun, Yu Sugihara, and Mauricio Pablo Contreras. Fold on tight: Ten things to know about AI structure prediction in plant immunity. August 2026. doi: 10.5281/zenodo.22100279.

Sarah E. Pottinger and Roger W. Innes. RPS5-mediated disease resistance: Fundamental insights and translational applications. *Annual Review of Phytopathology*, 58(1):139–160, August 2020. ISSN 0066-4286, 1545-2107. doi: 10.1146/annurev-phyto-010820-012733.

Morgan N. Price, Paramvir S. Dehal, and Adam P. Arkin. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLOS One*, 5(3):e9490, March 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0009490.

Daniil M Prigozhin and Ksenia V Krasileva. Analysis of intraspecies diversity reveals a subset of highly variable plant immune receptors and predicts their binding sites. *Plant Cell*, 33(4):998–1015, May 2021. ISSN 1040-4651, 1532-298X. doi: 10.1093/plcell/koab013.

Daniil M. Prigozhin, Chandler A. Sutherland, Sanjay Rangavajjhala, and Ksenia V. Krasileva. Majority of the highly variable NLRs in maize share genomic location and contain additional target-binding domains. *Molecular Plant-Microbe Interactions®*, 38(2):275–284, January 2025. ISSN 0894-0282, 1943-7706. doi: 10.1094/MPMI-05-24-0047-FI.

K. D. Pruitt, T. Tatusova, and D. R. Maglott. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 35(Database):D61–D65, January 2007. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkl842.

Rory N. Pruitt, Federica Locci, Friederike Wanke, Lisha Zhang, Svenja C. Saile, Anna Joe, Darya Karelina, Chenlei Hua, Katja Fröhlich, Wei-Lin Wan, Meijuan Hu, Shaofei Rao, Sara C. Stolze, Anne Harzen, Andrea A. Gust, Klaus Harter, Matthieu H. A. J. Joosten, Bart P. H. J. Thomma, Jian-Min Zhou, Jeffery L. Dangl, Detlef Weigel, Hirofumi Nakagami, Claudia Oecking, Farid El Kasmi, Jane E. Parker, and Thorsten Nürnberger. The EDS1–PAD4–ADR1 node mediates arabidopsis pattern-triggered immunity. *Nature*, 598(7881):495–499, October 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-021-03829-0.

Dong Qi, Brody J. DeYoung, and Roger W. Innes. Structure-function analysis of the coiled-coil and leucine-rich repeat domains of the RPS5 disease resistance protein. *Plant Physiology*, 158(4):1819–1832, April 2012. ISSN 1532-2548. doi: 10.1104/pp.112.194035.

Shiqian Qi, Yuxuan Pang, Qi Hu, Qun Liu, Hua Li, Yulian Zhou, Tianxi He, Qionglin Liang, Yexing Liu, Xiaoqiu Yuan, Guoan Luo, Huilin Li, Jiawei Wang, Nieng Yan, and Yigong Shi. Crystal structure of the caenorhabditis elegans apoptosome reveals an octameric

assembly of CED-4. *Cell*, 141(3):446–457, April 2010. ISSN 00928674. doi: 10.1016/j.cell.2010.03.017.

Tiancong Qi, Kyungyong Seong, Daniela P. T. Thomazella, Joonyoung Ryan Kim, Julie Pham, Eunyoung Seo, Myeong-Je Cho, Alex Schultink, and Brian J. Staskawicz. NRG1 functions downstream of EDS1 to regulate TIR-NLR-mediated plant immunity in *Nicotiana benthamiana*. *Proceedings of the National Academy of Sciences*, 115(46): E10979–E10987, November 2018. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1814856115.

Vincent Ranwez, Emmanuel J P Douzery, Cédric Cambon, Nathalie Chantret, and Frédéric Delsuc. MACSE v2: Toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Molecular Biology and Evolution*, 35(10):2582–2584, October 2018. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/msy159.

Jianbin Ruan, Shiyu Xia, Xing Liu, Judy Lieberman, and Hao Wu. Cryo-EM structure of the gasdermin A3 membrane pore. *Nature*, 557(7703):62–67, May 2018. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-018-0058-6.

Dziugas Sabonis, Carmel Avraham, Renee B. Chang, Allen Lu, Ehud Herbst, Arunas Silanskas, Deividas Vilutis, Azita Leavitt, Erez Yirmiya, Hunter C. Toyoda, Audrone Ruksenaite, Mindaugas Zaremba, Ilya Osterman, Gil Amitai, Philip J. Kranzusch, Rotem Sorek, and Giedre Tamulaitiene. TIR domains produce histidine-ADPR as an immune signal in bacteria. *Nature*, 642(8067):467–473, June 2025. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-025-08930-2.

Svenja C. Saile, Pierre Jacob, Baptiste Castel, Lance M. Jubic, Isai Salas-Gonzáles, Marcel Bäcker, Jonathan D. G. Jones, Jeffery L. Dangl, and Farid El Kasmi. Two unequally redundant "helper" immune receptor families mediate arabidopsis thaliana intracellular "sensor" immune receptor functions. *PLOS Biology*, 18(9):e3000783, September 2020. ISSN 1545-7885. doi: 10.1371/journal.pbio.3000783.

Cyrille Saintenac, Wenjun Zhang, Andres Salcedo, Matthew N. Rouse, Harold N. Trick, Eduard Akhunov, and Jorge Dubcovsky. Identification of wheat gene *Sr35* that confers resistance to Ug99 stem rust race group. *Science*, 341(6147):783–786, August 2013. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1239022.

Toshiyuki Sakai, Mauricio P Contreras, Claudia Martinez-Anaya, Daniel Lüdke, Sophien Kamoun, Chih-Hang Wu, and Hiroaki Adachi. The *NRC0* gene cluster of sensor and helper NLR immune receptors is functionally conserved across asterid plants. *Plant Cell*, 36(9):3344–3361, September 2024. ISSN 1040-4651, 1532-298X. doi: 10.1093/plcell/koae154.

Steven L. Salzberg. Next-generation genome annotation: We still struggle to get it right. *Genome Biology*, 20(1):92, s13059–019–1715–2, December 2019. ISSN 1474-760X. doi: 10.1186/s13059-019-1715-2.

Andre L. Samson, Ying Zhang, Niall D. Geoghegan, Xavier J. Gavin, Katherine A. Davies, Michael J. Mlodzianoski, Lachlan W. Whitehead, Daniel Frank, Sarah E. Garnish, Cheree Fitzgibbon, Anne Hempel, Samuel N. Young, Annette V. Jacobsen, Wayne Cawthorne, Emma J. Petrie, Maree C. Faux, Kristy Shield-Artin, Najoua Lalaoui, Joanne M. Hildebrand, John Silke, Kelly L. Rogers, Guillaume Lessene, Edwin D. Hawkins, and James M. Murphy. MLKL trafficking and accumulation at the plasma membrane control the kinetics and threshold for necroptosis. *Nature Communications*, 11(1):3151, June 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-16887-1.

Panagiotis F. Sarris, Zane Duxbury, Sung Un Huh, Yan Ma, Cécile Segonzac, Jan Sklenar, Paul Derbyshire, Volkan Cevik, Ghanasyam Rallapalli, Simon B. Saucet, Lennart Wirthmueller, Frank L.H. Menke, Kee Hoon Sohn, and Jonathan D.G. Jones. A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell*, 161(5):1089–1100, May 2015. ISSN 00928674. doi: 10.1016/j.cell.2015.04.024.

Panagiotis F. Sarris, Volkan Cevik, Gulay Dagdas, Jonathan D. G. Jones, and Ksenia V. Krasileva. Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biology*, 14(1):8–26, December 2016. ISSN 1741-7007. doi: 10.1186/s12915-016-0228-7.

Simon B. Saucet, Yan Ma, Panagiotis F. Sarris, Oliver J. Furzer, Kee Hoon Sohn, and Jonathan D.G. Jones. Two linked pairs of arabidopsis TNL resistance genes independently confer recognition of bacterial effector AvrRps4. *Nature Communications*, 6(1): 6338, March 2015. ISSN 2041-1723. doi: 10.1038/ncomms7338.

Isabel M. L. Saur, Ralph Panstruga, and Paul Schulze-Lefert. NOD-like receptor-mediated plant immunity: From structure to cell death. *Nature Reviews Immunology*, 21(5):305–318, May 2021. ISSN 1474-1733, 1474-1741. doi: 10.1038/s41577-020-00473-z.

Isabel Marie-Luise Saur, Brendon Francis Conlan, and John Paul Rathjen. The N-terminal domain of the tomato immune protein prf contains multiple homotypic and pto kinase interaction sites. *Journal of Biological Chemistry*, 290(18):11258–11267, May 2015. ISSN 00219258. doi: 10.1074/jbc.M114.616532.

Isabel Ml Saur, Saskia Bauer, Barbara Kracher, Xunli Lu, Lamprinos Franzeskakis, Marion C Müller, Björn Sabelleck, Florian Kümmel, Ralph Panstruga, Takaki Maekawa, and Paul Schulze-Lefert. Multiple pairs of allelic MLA immune receptor-powdery mildew AVRA effectors argue for a direct recognition mechanism. *eLife*, 8:e44471, February 2019. ISSN 2050-084X. doi: 10.7554/eLife.44471.

Benjamin A. Seager, Adeline Harant, Mauricio P. Contreras, Liang-Yu Hou, Chih Hang Wu, Sophien Kamoun, and Jogi Madhuprakash. A plant pathogen effector blocks stepwise assembly of a helper NLR resistosome. *Science Advances*, 12(23):eaeb1931, June 2026. ISSN 2375-2548. doi: 10.1126/sciadv.aeb1931.

Muniyandi Selvaraj, AmirAli Toghani, Hsuan Pai, Yu Sugihara, Jiorgos Kourelis, Enoch Lok Him Yuen, Tarhan Ibrahim, He Zhao, Rongrong Xie, Abbas Maqbool, Juan Carlos De La Concepcion, Mark J. Banfield, Lida Derevnina, Benjamin Petre, David M. Lawson, Tolga O. Bozkurt, Chih-Hang Wu, Sophien Kamoun, and Mauricio P. Contreras. Activation of plant immunity through conversion of a helper NLR homodimer into a resistosome. *PLOS Biology*, 22(10):e3002868, October 2024. ISSN 1545-7885. doi: 10.1371/journal.pbio.3002868.

Andrew W. Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander W. R. Nelson, Alex Bridgland, Hugo Penedones, Stig Petersen, Karen Simonyan, Steve Crossan, Pushmeet Kohli, David T. Jones, David Silver, Koray Kavukcuoglu, and Demis Hassabis. Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792):706–710, January 2020. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-019-1923-7.

Kyungyong Seong and Ksenia V. Krasileva. Computational structural genomics unravels common folds and novel families in the secretome of fungal phytopathogen *Magnaporthe oryzae*. *Molecular Plant-Microbe Interactions®*, 34(11):1267–1280, November 2021. ISSN 0894-0282, 1943-7706. doi: 10.1094/MPMI-03-21-0071-R.

Kyungyong Seong, Eunyoung Seo, Kamil Witek, Meng Li, and Brian Staskawicz. Evolution of NLR resistance genes with noncanonical N-terminal domains in wild tomato species. *New Phytologist*, 227(5):1530–1543, September 2020. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.16628.

Thiago V. Seraphim and Walid A. Houry. AAA+ proteins. *Current Biology*, 30(6):R251–R257, March 2020. ISSN 09609822. doi: 10.1016/j.cub.2020.01.044.

Carolin Seuring, Jason Greenwald, Christian Wasmer, Roger Wepf, Sven J. Saupe, Beat H. Meier, and Roland Riek. The mechanism of toxicity in HET-S/HET-s prion incompatibility. *PLOS Biology*, 10(12):e1001451, December 2012. ISSN 1545-7885. doi: 10.1371/journal.pbio.1001451.

Wei Shao, Gongfu Shi, Han Chu, Wenjia Du, Zikai Zhou, and Hada Wuriyanghan. Development of an NLR-ID toolkit and identification of novel disease-resistance genes in soybean. *Plants*, 13(5):668–689, February 2024. ISSN 2223-7747. doi: 10.3390/plants13050668.

Zhu-Qing Shao, Jia-Yu Xue, Ping Wu, Yan-Mei Zhang, Yue Wu, Yue-Yu Hang, Bin Wang, and Jian-Qun Chen. Large-scale analyses of angiosperm nucleotide-binding site-leucine-rich repeat genes reveal three anciently diverged classes with distinct evolutionary patterns. *Plant Physiology*, 170(4):2095–2109, March 2016. ISSN 1532-2548. doi: 10.1104/pp.15.01487.

Zhu-Qing Shao, Jia-Yu Xue, Qiang Wang, Bin Wang, and Jian-Qun Chen. Revisiting the origin of plant NBS-LRR genes. *Trends in Plant Science*, 24(1):9–12, January 2019. ISSN 13601385. doi: 10.1016/j.tplants.2018.10.015.

Qiaochu Shen, Keiichi Hasegawa, Nicole Oelerich, Anna Prakken, Lea Weiler Tersch, Junli Wang, Frowin Reichhardt, Alexandra Tersch, Je Cuan Choo, Ton Timmers, Kay Hofmann, Jane E. Parker, Jijie Chai, and Takaki Maekawa. Cytoplasmic calcium influx mediated by plant MLKLs confers TNL-triggered immunity. *Cell Host & Microbe*, 32(4): 453–465.e6, April 2024. ISSN 19313128. doi: 10.1016/j.chom.2024.02.016.

Yun Shi, Philip S. Kerry, Jeffrey D. Nanson, Todd Bosanac, Yo Sasaki, Raul Krauss, Forhad K. Saikot, Sarah E. Adams, Tamim Mosaiab, Veronika Masic, Xianrong Mao, Faith Rose, Eduardo Vasquez, Marieke Furrer, Katie Cunnea, Andrew Brearley, Weixi Gu, Zhenyao Luo, Lou Brillault, Michael J. Landsberg, Aaron DiAntonio, Bostjan Kobe, Jeffrey Milbrandt, Robert O. Hughes, and Thomas Ve. Structural basis of SARM1 activation, substrate recognition, and inhibition by small molecules. *Molecular Cell*, 82(9):1643–1659.e10, May 2022. ISSN 1097-2765. doi: 10.1016/j.molcel.2022.03.007.

Annelise G. Snyder and Andrew Oberst. The antisocial network: Cross talk between cell death programs in host defense. *Annual Review of Immunology*, 39(1):77–101, April 2021. ISSN 0732-0582, 1545-3278. doi: 10.1146/annurev-immunol-112019-072301.

Wen Song, Li Liu, Dongli Yu, Hanna Bernardy, Jan Jirschitzka, Shijia Huang, Aolin Jia, Wictoria Jemielniak, Julia Acker, Henriette Laessle, Junli Wang, Qiaochu Shen, Weijie Chen, Pilong Li, Jane E. Parker, Zhifu Han, Paul Schulze-Lefert, and Jijie Chai. Substrate-induced condensation activates plant TIR domain proteins. *Nature*, 627 (8005):847–853, March 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07183-9.

Jacob L. Steenwyk, Thomas J. Buida, Yuanning Li, Xing-Xing Shen, and Antonis Rokas. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLOS Biology*, 18(12):e3001007, December 2020. ISSN 1545-7885. doi: 10.1371/journal.pbio.3001007.

Joshua C. Stein, Yeisoo Yu, Dario Copetti, Derrick J. Zwickl, Li Zhang, Chengjun Zhang, Kapeel Chougule, Dongying Gao, Aiko Iwata, Jose Luis Goicoechea, Sharon Wei, Jun Wang, Yi Liao, Muhua Wang, Julie Jacquemin, Claude Becker, Dave Kudrna, Jianwei

Zhang, Carlos E. M. Londono, Xiang Song, Seunghee Lee, Paul Sanchez, Andrea Zuccolo, Jetty S. S. Ammiraju, Jayson Talag, Ann Danowitz, Luis F. Rivera, Andrea R. Gschwend, Christos Noutsos, Cheng-chieh Wu, Shu-min Kao, Jhih-wun Zeng, Fu-jin Wei, Qiang Zhao, Qi Feng, Moaine El Baidouri, Marie-Christine Carpentier, Eric Lasserre, Richard Cooke, Daniel Da Rosa Farias, Luciano Carlos Da Maia, Railson S. Dos Santos, Kevin G. Nyberg, Kenneth L. McNally, Ramil Mauleon, Nickolai Alexandrov, Jeremy Schmutz, Dave Flowers, Chuanzhu Fan, Detlef Weigel, Kshirod K. Jena, Thomas Wicker, Mingsheng Chen, Bin Han, Robert Henry, Yue-ie C. Hsing, Nori Kurata, Antonio Costa De Oliveira, Olivier Panaud, Scott A. Jackson, Carlos A. Machado, Michael J. Sanderson, Manyuan Long, Doreen Ware, and Rod A. Wing. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus oryza. *Nature Genetics*, 50(2):285–296, February 2018. ISSN 1061-4036, 1546-1718. doi: 10.1038/s41588-018-0040-0.

Burkhard Steuernagel, Florian Jupe, Kamil Witek, Jonathan D.G. Jones, and Brande B.H. Wulff. NLR-parser: Rapid annotation of plant NLR complements. *Bioinformatics*, 31(10): 1665–1667, May 2015. ISSN 1367-4811, 1367-4803. doi: 10.1093/bioinformatics/btv005.

Burkhard Steuernagel, Kamil Witek, Simon G. Krattinger, Ricardo H. Ramirez-Gonzalez, Henk-jan Schoonbeek, Guotai Yu, Erin Baggs, Agnieszka I. Witek, Inderjit Yadav, Ksenia V. Krasileva, Jonathan D.G. Jones, Cristobal Uauy, Beat Keller, Christopher J. Ridout, and Brande B.H. Wulff. The NLR-annotator tool enables annotation of the intracellular immune receptor repertoire. *Plant Physiology*, 183(2):468–482, June 2020. ISSN 0032-0889, 1532-2548. doi: 10.1104/pp.19.01273.

Felix Stiehler, Marvin Steinborn, Stephan Scholz, Daniela Dey, Andreas P M Weber, and Alisandra K Denton. Helixer: Cross-species gene annotation of large eukaryotic genomes using deep learning. *Bioinformatics*, 36(22-23):5291–5298, April 2021. ISSN 1367-4803, 1367-4811. doi: 10.1093/bioinformatics/btaa1044.

Yu Sugihara, AmirAli Toghani, Sophien Kamoun, Jiorgos Kourelis, Yu Sugihara, AmirAli Toghani, Sophien Kamoun, Jiorgos Kourelis, Yu Sugihara, AmirAli Toghani, Sophien Kamoun, and Jiorgos Kourelis. NLRome dataset from 124 genomes of plants in the solanaceae family, December 2023.

Yu Sugihara, Jiorgos Kourelis, Mauricio P. Contreras, Hsuan Pai, Adeline Harant, Muniyandi Selvaraj, AmirAli Toghani, Claudia Martínez-Anaya, and Sophien Kamoun. Helper NLR immune protein NRC3 evolved to evade inhibition by a cyst nematode virulence effector. *PLOS Genetics*, 21(4):e1011653, April 2025. ISSN 1553-7404. doi: 10.1371/journal.pgen.1011653.

Yu Sugihara, AmirAli Toghani, and Sophien Kamoun. No mokusatsu in our genomics. pages 1–8, June 2026. doi: 10.5281/ZENODO.20814410.

Oliver Sulkowski, Anna Ovodova, Andrea Leisse, Keelan Gögelein, and Alexander Förderer. Diversification of the "EDVID" packing motif underpins structural and functional variation in plant NLR coiled-coil domains, June 2025.

Hequan Sun, Sergio Tusso, Craig I. Dent, Manish Goel, Raúl Y. Wijfjes, Lisa C. Baus, Xiao Dong, José A. Campoy, Ana Kurdadze, Birgit Walkemeier, Christine Sänger, Bruno Huettel, Ronald C. B. Hutten, Herman J. Van Eck, Klaus J. Dehmer, and Korbinian Schneeberger. The phased pan-genome of tetraploid european potato. *Nature*, 642(8067):389–397, June 2025. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-025-08843-0.

Xinhua Sun, Dmitry Lapin, Joanna M. Feehan, Sara C. Stolze, Katharina Kramer, Joram A. Dongus, Jakub Rzemieniewski, Servane Blanvillain-Baufumé, Anne Harzen, Jaqueline Bautor, Paul Derbyshire, Frank L. H. Menke, Iris Finkemeier, Hirofumi Nakagami, Jonathan D. G. Jones, and Jane E. Parker. Pathogen effector recognition-dependent association of NRG1 with EDS1 and SAG101 in TNL receptor immunity. *Nature Communications*, 12(1):3335, June 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-23614-x.

Balamurugan Sundaram, Rebecca E. Tweedell, Sivakumar Prasanth Kumar, and Thirumala-Devi Kanneganti. The NLR family of innate immune and cell death sensors. *Immunity*, 57(4):674–699, April 2024. ISSN 10747613. doi: 10.1016/j.immuni.2024.03.012.

Sruthi Sunil, Simon Beeh, Eva Stöbbe, Kathrin Fischer, Franziska Wilhelm, Aron Meral, Celia Paris, Luisa Teasdale, Zhihao Jiang, Lisha Zhang, Moritz Urban, Emmanuel Aguilar Parras, Thorsten Nürnberger, Detlef Weigel, Rosa Lozano-Duran, and Farid El Kasmi. Activation of an atypical plant NLR with an N-terminal deletion initiates cell death at the vacuole. *EMBO Reports*, 25(10):4358–4386, September 2024. ISSN 1469-3178. doi: 10.1038/s44319-024-00240-4.

Frank Lw Takken and Aska Goverse. How to build a pathogen detector: Structural basis of NB-LRR function. *Current Opinion in Plant Biology*, 15(4):375–384, August 2012. ISSN 13695266. doi: 10.1016/j.pbi.2012.05.001.

Frank Lw Takken, Mario Albrecht, and Wladimir Il Tameling. Resistance proteins: Molecular switches of plant defence. *Current Opinion in Plant Biology*, 9(4):383–390, August 2006. ISSN 13695266. doi: 10.1016/j.pbi.2006.05.009.

Janina Tamborski and Ksenia V. Krasileva. Evolution of plant NLRs: From natural history

to precise modifications. *Annual Review of Plant Biology*, 71(1):355–378, April 2020. ISSN 1543-5008, 1545-2123. doi: 10.1146/annurev-arplant-081519-035901.

D Ellen K Tarr and Helen M Alexander. TIR-NBS-LRR genes are rare in monocots: Evidence from diverse monocot orders. *BMC Research Notes*, 2(1):197–206, 2009. ISSN 1756-0500. doi: 10.1186/1756-0500-2-197.

Vardges Ter-Hovhannisyan, Alexandre Lomsadze, Yury O. Chernoff, and Mark Borodovsky. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Research*, 18(12):1979–1990, December 2008. ISSN 1088-9051. doi: 10.1101/gr.081612.108.

The Angiosperm Phylogeny Group. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, 181(1):1–20, May 2016. ISSN 00244074. doi: 10.1111/boj.12385.

Hainan Tian, Zhongshou Wu, Siyu Chen, Kevin Ao, Weijie Huang, Hoda Yaghmaiean, Tongjun Sun, Fang Xu, Yanjun Zhang, Shucai Wang, Xin Li, and Yuelin Zhang. Activation of TIR signalling boosts pattern-triggered immunity. *Nature*, 598(7881):500–503, October 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-021-03987-1.

AmirAli Toghani. AI usage declaration - chat snapshot. September 2026. doi: 10.5281/zenodo.22927141.

AmirAli Toghani and Sophien Kamoun. Functional annotation of 180 RefSeq reference plant proteomes reveals a dataset of 113,684 NLR proteins, September 2024.

AmirAli Toghani, Raoul Frijters, Tolga O. Bozkurt, Ryohei Terauchi, Sophien Kamoun, and Yu Sugihara. Can AI modeling of protein structures distinguish between sensor and helper NLR immune receptors? *New Phytologist*, 248(1):17–23, October 2025a. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.70391.

AmirAli Toghani, Yu Sugihara, and Sophien Kamoun. Deep-learning-based annotation of 230 superasterid genomes reveals a harmonized dataset of 91,366 NLRs (v250214_91366), January 2025b.

AmirAli Toghani, Maián Garro, Raoul Frijters, Sophien Kamoun, and Mauricio P. Contreras. A conserved structural logic underlies sensor–helper NLR communication in the NRC immune receptor network, May 2026a. ISSN 2692-8205.

AmirAli Toghani, Benjamin A. Seager, Yu Sugihara, Lisa-Marie Roijen, Juan M. Azcue, Maián Garro, Maryam Sargolzaei, Ioanna Morianou, Adeline Harant, Sam Gallop, Jiorgos Kourelis, Dan MacLean, Mauricio P. Contreras, Sophien Kamoun, and Daniel

Lüdke. AI-guided discovery of atypical protein assemblies, May 2026b. ISSN 2692-8205.

AmirAli Toghani, Yu Sugihara, Sophien Kamoun, AmirAli Toghani, Yu Sugihara, Sophien Kamoun, AmirAli Toghani, Yu Sugihara, Sophien Kamoun, AmirAli Toghani, Yu Sugihara, Sophien Kamoun, AmirAli Toghani, Yu Sugihara, and Sophien Kamoun. Harmonized genome annotation of 346 solanaceae species reveals 197,834 NLR immune receptors, April 2026c.

Phu-Tri Tran, Kristin Widyasari, Jang-Kyun Seo, and Kook-Hyung Kim. Isolation and validation of a candidate Rsv3 gene from a soybean genotype that confers strain-specific resistance to soybean mosaic virus. *Virology*, 513:153–159, January 2018. ISSN 00426822. doi: 10.1016/j.virol.2017.10.014.

Jessie Uehling, Aurélie Deveau, and Mathieu Paoletti. Do fungi have an innate immune response? An NLR-based comparison to plant and animal immune systems. *PLOS Pathogens*, 13(10):e1006578, October 2017. ISSN 1553-7374. doi: 10.1371/journal.ppat.1006578.

Lisa Vader, Colin J.B. Harvey, Tilmann Weber, and Lawrence S. Hon. geneML: Gene annotation across diverse fungal species using deep learning, May 2026. ISSN 2692-8205.

Anna-Lena Van De Weyer, Freddy Monteiro, Oliver J. Furzer, Marc T. Nishimura, Volkan Cevik, Kamil Witek, Jonathan D.G. Jones, Jeffery L. Dangl, Detlef Weigel, and Felix Bemm. A species-wide inventory of NLR genes and alleles in arabidopsis thaliana. *Cell*, 178(5):1260–1272.e14, August 2019. ISSN 00928674. doi: 10.1016/j.cell.2019.07.038.

Stijn Van Dongen. Graph clustering via a discrete uncoupling process. *SIAM Journal on Matrix Analysis and Applications*, 30(1):121–141, January 2008. ISSN 0895-4798, 1095-7162. doi: 10.1137/040608635.

Stephen Branden Van Oss and Anne-Ruxandra Carvunis. De novo gene birth. *PLOS Genetics*, 15(5):e1008160, May 2019. ISSN 1553-7404. doi: 10.1371/journal.pgen.1008160.

Dirk-Jan M. Van Workum, Sarah L. Mehrem, Basten L. Snoek, Marrit C. Alderkamp, Dmitry Lapin, Flip F.M. Mulder, Guido Van Den Ackerveken, Dick De Ridder, M. Eric Schranz, and Sandra Smit. *Lactuca* super-pangenome reduces bias towards reference genes in lettuce research, June 2024.

Omar Wagih. Ggseqlogo: A versatile R package for drawing sequence logos. *Bioinformatics*, 33(22):3645–3647, November 2017. ISSN 1367-4803, 1367-4811. doi: 10.1093/bioinformatics/btx469.

Li Wan, Kow Essuman, Ryan G. Anderson, Yo Sasaki, Freddy Monteiro, Eui-Hwan Chung, Erin Osborne Nishimura, Aaron DiAntonio, Jeffrey Milbrandt, Jeffery L. Dangl, and Marc T. Nishimura. TIR domains of plant immune receptors are NAD + -cleaving enzymes that promote cell death. *Science*, 365(6455):799–803, August 2019. ISSN 1095-9203. doi: 10.1126/science.aax1771.

Jizong Wang, Meijuan Hu, Jia Wang, Jinfeng Qi, Zhifu Han, Guoxun Wang, Yijun Qi, Hong-Wei Wang, Jian-Min Zhou, and Jijie Chai. Reconstitution and structure of a plant NLR resistosome conferring immunity. *Science*, 364(6435):eaav5870, April 2019a. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aav5870.

Jizong Wang, Jia Wang, Meijuan Hu, Shan Wu, Jinfeng Qi, Guoxun Wang, Zhifu Han, Yijun Qi, Ning Gao, Hong-Wei Wang, Jian-Min Zhou, and Jijie Chai. Ligand-triggered allosteric ADP release primes a plant NLR complex. *Science*, 364(6435):eaav5868, April 2019b. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aav5868.

Luyao Wang, Hongbo Li, Yuhang Ke, Ping Zhu, Yuying Li, Pei Wang, Haohao Liu, Yajie Li, Jingqi Chen, Shaoqun Zhou, Li Wan, Savithramma P. Dinesh-Kumar, Suomeng Dong, and Sanwen Huang. Plug-in strategy for resistance engineering inspired by potato NLRome. *Nature*, 649(8096):396–405, January 2026. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-025-09678-5.

Zaiqing Wang, Xiaoxiao Liu, Jie Yu, Shuining Yin, Wenjuan Cai, Nak Hyun Kim, Farid El Kasmi, Jeffery L. Dangl, and Li Wan. Plasma membrane association and resistosome formation of plant helper immune receptors. *Proceedings of the National Academy of Sciences*, 120(32):e2222036120, August 2023. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2222036120.

Tanita Wein and Rotem Sorek. Bacterial origins of human cell-autonomous innate immune mechanisms. *Nature Reviews Immunology*, 22(10):629–638, October 2022. ISSN 1474-1733, 1474-1741. doi: 10.1038/s41577-022-00705-4.

Caroline M. Weisman, Andrew W. Murray, and Sean R. Eddy. Mixing genome annotation methods in a comparative analysis inflates the apparent number of lineage-specific genes. *Current Biology*, 32(12):2632–2639.e2, June 2022. ISSN 09609822. doi: 10.1016/j.cub.2022.04.085.

Hadley Wickham. *Ggplot2*. Use R! Springer International Publishing, Cham, 2016. ISBN 978-3-319-24275-0 978-3-319-24277-4. doi: 10.1007/978-3-319-24277-4.

Luzie Wingen, Duncan Crosbie, Yiheng Hu, Eric Kemen, Xinyi Liu, Marion Müller, Niklas Schandry, Korbinian Schneeberger, Detlef Weigel, and Aurélien Tellier. Towards a

quantitative view of the NLR gene family 4evolution in the genome space, December 2025.

Chih-Hang Wu, Ahmed Abd-El-Haliem, Tolga O. Bozkurt, Khaoula Belhaj, Ryohei Terauchi, Jack H. Vossen, and Sophien Kamoun. NLR network mediates immunity to diverse plant pathogens. *Proceedings of the National Academy of Sciences*, 114(30): 8113–8118, July 2017. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1702041114.

Chih-Hang Wu, Lida Derevnina, and Sophien Kamoun. Receptor networks underpin plant immunity. *Science*, 360(6395):1300–1301, June 2018. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aat2623.

Zhongshou Wu. TIR innovations in plant immunity. *New Phytologist*, 249(2):758–763, June 2025. ISSN 1469-8137. doi: 10.1111/nph.70314.

Zhongshou Wu, Meng Li, Oliver Xiaoou Dong, Shitou Xia, Wanwan Liang, Yongkang Bao, Geoffrey Wasteneys, and Xin Li. Differential regulation of TNL-mediated immune signaling by redundant helper CNLs. *New Phytologist*, 222(2):938–953, April 2019. ISSN 0028-646X, 1469-8137. doi: 10.1111/nph.15665.

Zhongshou Wu, Lei Tian, Xueru Liu, Weijie Huang, Yuelin Zhang, and Xin Li. The N-terminally truncated helper NLR *NRG1C* antagonizes immunity mediated by its full-length neighbors *NRG1A* and *NRG1B*. *Plant Cell*, 34(5):1621–1640, April 2022. ISSN 1040-4651, 1532-298X. doi: 10.1093/plcell/koab285.

Le Xiao, Venkat Giri Magupalli, and Hao Wu. Cryo-EM structures of the active NLRP3 inflammasome disc. *Nature*, 613(7944):595–600, January 2023. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-022-05570-8.

Lihong Xiao, Ge Yang, Liechi Zhang, Xinhua Yang, Shuang Zhao, Zhongzhong Ji, Qing Zhou, Min Hu, Yu Wang, Ming Chen, Yu Xu, Haijing Jin, Xuan Xiao, Guipeng Hu, Fang Bao, Yong Hu, Ping Wan, Legong Li, Xin Deng, Tingyun Kuang, Chengbin Xiang, Jian-Kang Zhu, Melvin J. Oliver, and Yikun He. The resurrection genome of *Boea hygrometrica*: A blueprint for survival of dehydration. *Proceedings of the National Academy of Sciences*, 112(18):5833–5837, May 2015. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1505811112.

Z. Yang. Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, 22(4):1107–1118, January 2005. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/msi097.

Ziheng Yang. PAML: A program package for phylogenetic analysis by maximum likelihood. *Bioinformatics*, 13(5):555–556, 1997. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/13.5.555.

Dongli Yu, Wen Song, Eddie Yong Jun Tan, Li Liu, Yu Cao, Jan Jirschitzka, Ertong Li, Elke Logemann, Chenrui Xu, Shijia Huang, Aolin Jia, Xiaoyu Chang, Zhifu Han, Bin Wu, Paul Schulze-Lefert, and Jijie Chai. TIR domains of plant immune receptors are 2′,3′-cAMP/cGMP synthetases mediating cell death. *Cell*, 185(13):2370–2386.e18, June 2022. ISSN 00928674. doi: 10.1016/j.cell.2022.04.032.

Hua Yu, Weiying Xu, Sisi Chen, Xiaoxian Wu, Weiwei Rao, Xiaoxiao Liu, Xiaoyan Xu, Jingqi Chen, Marc T. Nishimura, Yu Zhang, and Li Wan. Activation of a helper NLR by plant and bacterial TIR immune signaling. *Science*, 386(6728):1413–1420, December 2024. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.adr3150.

Minhang Yuan, Zeyu Jiang, Guozhi Bi, Kinya Nomura, Menghui Liu, Yiping Wang, Boying Cai, Jian-Min Zhou, Sheng Yang He, and Xiu-Fang Xin. Pattern-recognition receptors are required for NLR-mediated plant immunity. *Nature*, 592(7852):105–109, April 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-021-03316-6.

Shujun Yuan, Maya Topf, Thomas F. Reubold, Susanne Eschenburg, and Christopher W. Akey. Changes in apaf-1 conformation that drive apoptosome assembly. *Biochemistry*, 52(13):2319–2327, April 2013. ISSN 0006-2960, 1520-4995. doi: 10.1021/bi301721g.

Benedict Yuen, Joanne M. Bayes, and Sandie M. Degnan. The characterization of sponge NLRs provides insight into the origin and evolution of this innate immune gene family in animals. *Molecular Biology and Evolution*, 31(1):106–120, January 2014. ISSN 1537-1719, 0737-4038. doi: 10.1093/molbev/mst174.

Rafał Zdrzałek, Sophien Kamoun, Ryohei Terauchi, Hiromasa Saitoh, and Mark J. Banfield. The rice NLR pair pikp-1/pikp-2 initiates cell death through receptor cooperation rather than negative regulation. *PLOS One*, 15(9):e0238616, September 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0238616.

Yan-Bo Zhao, Meng-Xi Liu, Tao-Tao Chen, Xiaomin Ma, Ze-Kai Li, Zichao Zheng, Si-Ru Zheng, Lifei Chen, You-Zhi Li, Li-Rui Tang, Qi Chen, Peiyi Wang, and Songying Ouyang. Pathogen effector AvrSr35 triggers Sr35 resistosome assembly via a direct recognition mechanism. *Science Advances*, 8(36):eabq5108, September 2022. ISSN 2375-2548. doi: 10.1126/sciadv.abq5108.

Yao Zhou, Zhiyang Zhang, Zhigui Bao, Hongbo Li, Yaqing Lyu, Yanjun Zan, Yaoyao Wu, Lin Cheng, Yuhan Fang, Kun Wu, Jinzhe Zhang, Hongjun Lyu, Tao Lin, Qiang Gao, Surya Saha, Lukas Mueller, Zhangjun Fei, Thomas Städler, Shizhong Xu, Zhiwu Zhang, Doug Speed, and Sanwen Huang. Graph pangenome captures missing heritability and empowers tomato breeding. *Nature*, 606(7914):527–534, June 2022. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-022-04808-9.

Appendix I: Publications and datasets

During my PhD, I had the opportunity to participate in many projects and contribute to several publications and datasets. This appendix lists the publications and datasets resulting from this work. Under each entry, I provide a statement detailing my specific contribution to the project. Entries are arranged by year, with the most recent listed first.

(†) indicates joint first authorship. (*) indicates a corresponding author.

Papers and preprints

1. Posbeyikian A.†, **Toghani A.**†, McClelland A.J.†, Kamoun S.*, Sugihara Y.*, Contreras M.P.* (2026) Fold on tight: ten things to know about AI structure prediction in plant immunity. *Zenodo*. doi:10.5281/zenodo.22100278

Together with colleagues at TSL and ZMBP, I co-authored a review synthesizing recent developments in plant immunity into ten practical lessons for applying AI protein structure prediction across the plant sciences.

2. Feng L.†, Hou Y.†, **Toghani A.**, Wang Z., Tang B., Atkinson N., Li H., Qiao Y., Wang Y., Hua J., Zhai J., Ma W.* (2026) A conserved small RNA-generating gene cluster undergoes sequence diversification and contributes to plant immunity. *Nature Communications*. doi:10.1038/s41467-026-77164-1

I conducted a large-scale phylogenetic analysis of PPR genes across an evolutionarily diverse panel of plant species.

3. Pai H., Sakai T., Posbeyikian A., Frijters R., Sugihara Y., Contreras M.P., Kourelis J., Adachi H., Kamoun S.*, **Toghani A.*** (2026) Contrasting evolutionary patterns of helper and sensor NRC NLRs in lettuce reflect functional divergence following subfunctionalization. *PLOS Genetics*. doi:10.1371/journal.pgen.1012245

This work forms the basis of Chapters 4 and 6 of this thesis. I led the project, performed the evolutionary and phylogenetic analyses of NLRs, drafted the manuscript, and prepared the figures.

4. Gao C.[†], Meng X.[†], Chen X., Yang L., Ibrahim T., **Toghani A.**, Yuen E.L.H., Eilmann N., King F., Li K., Wang L., Sun B., Wang Y., Bozkurt T.O.*, Dong S.* (2026) Surface immune signaling unlocks NLR activation through mRNA alternative splicing. *Science*. doi:10.1126/science.adx9929

I carried out large-scale phylogenetic analyses of CC-type NLRs at two taxonomic levels across green plants and the Solanaceae family. I also performed detailed analyses of the MADA motif and identified NLRs with atypical N-terminal extensions preceding the motif.

5. Madhuprakash J., **Toghani A.**, Pai H., Harvey M., Bentham A.R., Seager B.A., Yuen E.L.H., De la Concepcion J.C., Lawson D.M., Stevenson C.E.M., Vergara-Cruces A., Derevnina L., Bozkurt T.O., Banfield M.J., Kamoun S.*, Contreras M.P.* (2026) An effector from the potato late blight pathogen bridges ENTH-domain protein TOL9a to an activated helper NLR to suppress immunity. *Science Advances*. doi:10.1126/sciadv.aea4500

I performed the phylogenetic analyses and structural modeling of the AVRcap1b and PSR2 effector clades from the genus *Phytophthora*. I also contributed to drafting the manuscript and preparing the figures.

6. **Toghani A.**[†], Garro M.[†], Frijters R., Kamoun S.*, Contreras M.P.* (2026) A conserved structural logic underlies sensor–helper NLR communication in the NRC immune receptor network. *bioRxiv*. doi:10.64898/2026.05.18.725810

I conceptualized this study alongside Mauricio Contreras, conducted the structural modeling and phylogenetic analyses of sensor–helper NLRs across the NRC network, and co-wrote the manuscript.

7. **Toghani A.**[†], Seager B.A.[†], Sugihara Y., Roijen L.-M., Azcue J.M., Garro M., Sargolzaei M., Morianou I., Harant A., Gallop S., Kourelis J., MacLean D., Contreras M.P., Kamoun S.*, Lüdke D.* (2026) AI-guided discovery of atypical protein assemblies. ***bioRxiv***. doi:10.64898/2026.05.03.722499

Building on the resistosome modeling framework presented in Chapter 5 of this thesis, I carried out the structural modeling and comparative analysis of the Solanaceae NRC helper family. I also contributed to writing the manuscript and preparing the figures.

8. Ibrahim T.[†], King F.J.[†], **Toghani A.**, Wang L., Jenkins S., Yuen E.L.H., Wang H.-Y., Vuolo C., Eilmann N., Adamkova V., Chia K.-S., Castel B., Jones J.D.G., Carella P., Wu C.-H., Kourelis J.*, Kamoun S.*, Bozkurt T.O.* (2026) A helper NLR channels organellar calcium to trigger plant immunity. ***Science***. doi:10.1126/science.aeb6690

I carried out the phylogenetic analysis of RPW8-type helper NLRs across non-flowering and flowering plants, and performed the N-terminal motif characterization for this protein family.

9. King F.[†], Lopez-Agudelo J.C.[†], Stephens C., Aung M.H., Ibrahim T., Yuen E.L.H., Chen A., Eilmann N.M., Jenkins S., Vuolo C., Swee Y.-N., Liu W.-J., Bruty S., **Toghani A.**, Kuo C.-H., Lai E.-M., Kourelis J., Derevnina L.*, Wu C.-H.*, Bozkurt T.O.* (2026) A broad-host-range *Rhizobium rhizogenes* strain enables transient expression across diverse crops and establishes functional assays in faba bean. ***bioRxiv***. doi:10.64898/2026.04.20.719558

I designed and generated constructs used for the functional transient expression assays in this study.

10. Lüdke D., Pai H., **Toghani A.**, Harant A., Wu C.-H., Kamoun S.* (2025) The autoactivity of tomato helper NLR immune proteins of the NRC clade is unaltered in *prf* mutants of *Nicotiana benthamiana*. ***Plant Physiology***. doi:10.1093/plphys/kiaf506

I performed the phylogenetic analysis of the Solanales NRC helpers and the NRC-S clade containing the tomato *Prf* resistance gene. I also contributed to drafting the manuscript and preparing the figures.

11. **Toghani A.**, Frijters R., Bozkurt T.O., Terauchi R., Kamoun S.*, Sugihara Y.* (2025) Can AI modelling of protein structures distinguish between sensor and helper NLR immune receptors? *New Phytologist*. doi:10.1111/nph.70391

This work is presented in Chapter 6 of this thesis. Together with Yu Sugihara, I conducted the structural modeling and comparative analysis of sensor–helper pairs. I also contributed to writing the manuscript and preparing the figures.

12. Lüdke D., Sakai T., Kourelis J., **Toghani A.**, Adachi H., Posbeyikian A., Frijters R., Pai H., Harant A., Ernst K., Ganal M., Verhage A., Wu C.-H., Kamoun S.* (2025) A root-specific NLR network mediates immune signaling of resistance genes against plant parasitic nematodes. *The Plant Cell*. doi:10.1093/plcell/koaf145

I conducted the phylogenetic analysis of the Solanaceae NRC helper family, focusing on the NRC6 clade and its dependent sensor NLRs including the *Hero* and *Mer1* resistance genes. I also contributed to figure preparation.

13. Sugihara Y., Kourelis J., Contreras M.P., Pai H., Harant A., Selvaraj M., **Toghani A.**, Martinez-Anaya C.*, Kamoun S.* (2025) Helper NLR immune protein NRC3 evolved to evade inhibition by a cyst nematode virulence effector. *PLOS Genetics*. doi:10.1371/journal.pgen.1011653

I assisted with compiling the genomic datasets used in this study and contributed to the initial project conceptualization.

14. Madhuprakash J.*, **Toghani A.**, Contreras M.P., Posbeyikian A., Richardson J., Kourelis J., Bozkurt T.O., Webster M.W.*, Kamoun S.* (2024) A disease resistance protein triggers oligomerization of its NLR helper into a hexameric resistosome to mediate innate immunity. *Science Advances*. doi:10.1126/sciadv.adr2594

This work forms the basis of Chapter 5 of this thesis. I carried out the structural modeling and phylogenetic analyses of NLR proteins across angiosperms, and contributed to drafting the manuscript and preparing the figures.

15. Selvaraj M., **Toghani A.**, Pai H., Sugihara Y., Kourelis J., Yuen E.L.H., Ibrahim T., Zhao H., Xie R., Maqbool A., De la Concepcion J.C., Banfield M.J., Derevnina L., Petre B., Lawson D.M., Bozkurt T.O., Wu C.-H., Kamoun S.*, Contreras M.P.* (2024) Activation of plant immunity through conversion of a helper NLR homodimer into a resistosome. *PLOS Biology*. doi:10.1371/journal.pbio.3002868

I conducted the sequence and structural analyses of the NbNRC2a homodimer and assessed the evolutionary conservation of the dimerization interfaces across the Solanaceae NRC helper family. I also contributed to drafting the manuscript and preparing the figures.

16. Huang C.-Y., Huang Y.-S., Sugihara Y., Wang H.-Y., Huang L.-T., Lopez-Agudelo J.C., Chen Y.-F., Lin K.-Y., Chiang B.-J., **Toghani A.**, Kourelis J., Wang C.-H., Derevnina L., Wu C.-H.* (2024) Subfunctionalization of NRC3 altered the genetic structure of the *Nicotiana* NRC network. *PLOS Genetics*. doi:10.1371/journal.pgen.1011402

I contributed to the compilation and quality control of the NLR dataset utilized in this study.

17. Sakai T., **Toghani A.**, Adachi H.* (2024) Phylogenomics of plant NLR immune receptors to identify functionally conserved sequence motifs. *Bio-protocol*. doi:10.21769/BioProtoc.5023

I co-authored this protocol paper alongside Hiroaki Adachi and Toshiyuki Sakai, contributing to the bioinformatic workflow design and methods text.

18. Contreras M.P.*, Lüdke D., Pai H., **Toghani A.**, Kamoun S.* (2023) NLR receptors in plant immunity: making sense of the alphabet soup. *EMBO Reports*. doi:10.15252/embr.202357495

I generated the comprehensive phylogenetic tree of plant NLRs for this review and contributed to drafting the manuscript.

19. Torres Ascurra Y.C., Zhang L., **Toghani A.**, Hua C., Rangegowda N.J., Posbeyikian A., Pai H., Lin X., Wolters P.J., Wouters D., De Blok R., Steigenga N., Paillart M.J.M., Visser R.G.F., Kamoun S., Nürnberger T.*, Vleeshouwers V.G.A.A.* (2023) Functional diversification of a wild potato immune receptor at its center of origin. *Science*. doi:10.1126/science.adg5261

I conducted the phylogenetic and diversity analyses of the PERU LRR-RLK receptor across the Solanaceae, focusing on wild *Solanum* species from its center of origin. I also contributed to drafting the manuscript and preparing the figures.

Datasets

1. **Toghani A.**, Seager B.A., Lüdke D., Kamoun S. (2026) AlphaFold 3 hexameric resistosome models for 637 NRC-clade NLR proteins from Solanaceae. *Zenodo*. doi:10.5281/zenodo.19860917

Predicted hexameric resistosome structures for 637 Solanaceae NRC-clade NLR proteins modeled using AlphaFold 3. This dataset supports the preprint by Toghani, Seager, et al. (2026).

2. **Toghani A.**, Sugihara Y., Kamoun S. (2026) Harmonized genome annotation of 346 Solanaceae species reveals 197,834 NLR immune receptors. *Zenodo*. doi:10.5281/zenodo.19855162

De novo annotation of 346 Solanaceae genomes and the identification of 197,834 NLR immune receptors. This dataset is presented in Chapter 2 of this thesis.

3. **Toghani A.**, Sugihara Y., Kamoun S. (2025) Deep-learning-based annotation of 230 superasterid genomes reveals a harmonized dataset of 91,366 NLRs. *Dryad*. doi:10.5061/dryad.sxksn03d6

De novo annotation of 230 superasterid genomes and the identification of 91,366 NLR immune receptors. This dataset is presented in Chapter 2 of this thesis.

4. **Toghani A.**, Kamoun S. (2024) Functional annotation of 180 RefSeq reference plant proteomes reveals a dataset of 113,684 NLR proteins. *Zenodo*. zenodo.org/records/13627395

A harmonized dataset of 113,684 NLR proteins identified across 180 RefSeq reference plant proteomes. This dataset is presented in Chapter 2 of this thesis.

5. **Toghani A.**, Bozkurt T.O., Kamoun S. (2024) AlphaFold 3 predicted NLR resistosome structures. *Zenodo*. doi:10.5281/zenodo.11546021

Predicted resistosome structures of representative CC-type NLRs sampled across the angiosperm NLR phylogeny modeled using AlphaFold 3. This dataset supports Madhuprakash et al. (2024) and is presented in Chapter 5 of this thesis.

6. Sugihara Y., **Toghani A.**, Kamoun S., Kourelis J. (2023) NLRome dataset from 124 genomes of plants in the Solanaceae family. *Zenodo*. doi:10.5281/zenodo.10354349

A curated NLRome dataset extracted from 124 publicly available Solanaceae genomes. I compiled and curated this dataset together with Yu Sugihara and Jiorgos Kourelis.