Engineering of Sr33 and Sr50 plant immune receptors to alter recognition specificity and autoactivity

Janina Tamborski¹, Kyungyong Seong¹, Furong Liu^{1,2}, Brian Staskawicz^{1,2}, Ksenia V Krasileva^{1,2}

¹Department of Plant and Microbial Biology, University of California Berkeley, USA; ²Innovative Genomics Institute, UC Berkeley, USA

Abstract

Plants possess cytoplasmic immune receptors called nucleotide-binding leucine-rich repeat receptors (NLRs) that recognize the presence of a pathogen through a range of mechanisms: direct binding of effectors or indirect recognition of effector actions. The direct binding of effectors has been shown to be mediated through the NLR's leucine-rich repeat (LRR) domain. Accurate prediction of amino acids involved in these direct interactions can greatly enhance understanding of effector recognition and inform efforts to engineer new resistance. In this study, we utilized two homologous NLR resistance genes from wheat, Sr33 and Sr50, that recognize distinct effectors by directly binding to them through their LRR domain. While the effector recognized by Sr50 is known and described as AvrSr50, the effector recognized by Sr33 remains unknown. Through a combination of phylogenetics, allele diversity analysis in the LRR and structural modeling, we identified the amino acids in Sr50 likely to physically interact with its effector. Mutation of these sites helped identify 12 amino acids we hypothesized to be sufficient to mediate effector binding in Sr50. Changing these 12 corresponding amino acids in Sr33, we showed AvrSr50-dependent initiation of cell death in wheat protoplasts and *Nicotiana benthamiana*. Furthermore, we were able to pinpoint and change amino acid residues that govern autoactivity of Sr50 in the wheat protoplast cell death assay. These findings are a major advance towards the successful engineering of new effector recognition specificities in direct binder NLRs.

Introduction

Nucleotide-binding leucine-rich repeat receptors (NLRs) are intracellular immune receptors that recognize pathogen-derived molecules, called effectors, that are secreted into the plant cell by a pathogen with the goal of subverting plant immune responses. NLRs deploy different means of monitoring the cell by either directly binding to effectors or recognizing the action of an effector on a plant component. Plant components whose modification is recognized by NLRs are referred to as guardees or decoys. The recognition of an effector leads to the activation of downstream responses which culminate in localized cell death, referred to as hypersensitive response (HR) (Dodds and Rathjen, 2010). Although plant NLRs have been shown to evolve rapidly and recognize a wide range of structurally diverse effectors, robust approaches for rational modifications of NLRs and their recognition specificities remain limited (Tamborski and Krasileva, 2020).

Most NLRs share a common, multi-domain structure including a variable N-terminal domain, an NB-ARC domain, and a leucine-rich repeat (LRR) domain. There are three types of N-terminal domains that enable us to separate NLRs into three categories: CNLs that carry a coiled-coil (CC) domain, TNLs that carry a Toll/interleukin-1 receptor (TIR) domain and RNLs that have an RPW8-like domain (Shao et al., 2016). Furthermore, a subset of NLRs have one or more atypical plant protein domains integrated into their canonical NLR structure and are referred to as NLR-IDs (for integrated domain) (Grund et al., 2019). NLR-IDs usually function in pairs and are often genetically linked in a head-to-head orientation and are thought to exemplify the specialization of NLRs into a sensor and signaler function (Feehan et al., 2020). The integrated domains act as effector binding platforms and have been shown to directly interact with effectors (Cesari et al., 2013; Maqbool et al., 2015).

Several studies have already successfully engineered recognition specificities of NLRs. In direct binder NLRs, recognition was transferred from one NLR to another through chimeras (Ellis et al., 1999; Shen et al., 2003; Sliotweg et al., 2017) and effector recognition was expanded through the means of random and targeted mutagenesis (Racman et al., 2001; Segretin et al., 2014; Chapman et al., 2014; Giannakopoulou et al., 2015). Also the modification of a guardee has proven successful through the introduction of effector substrate sites (Pottinger et al., 2020; DeYoung et al., 2012; Kim et al., 2016). Recently there have been some major advances in engineering the integrated domains of NLR-IDs to broaden or change their recognition specificity (Cesari et al., 2021; De la Concepcion et al., 2019; Liu et al., 2021). A major step forward was the identification and subsequent alteration of specific residues that mediate effector recognition. These successful engineering efforts exemplify that information from structural and effector binding studies help us not only understand but also engineer binding specificities of NLRs. While these most recent studies focused on modifying the integrated plant protein domain that binds to the effector, we have yet to make significant progress in engineering NLRs that bind to effectors through their LRR domain.

The recent cryo-EM structures of plant NLR genes have greatly enhanced our understanding of activated immune receptor function (Martin et al., 2020; Wang et al., 2019a, 2019b; Ma et al., 2020). The CNL ZAR1 and the kinase RKS1 guard the PBL2 kinase and initiate immune signaling upon its modification (Lewis et al., 2010; Wang et al., 2015). ZAR1 and RKS1 exist in pre-formed complexes that assemble into pentamers upon effector-mediated modification of PBL2 (Wang et al., 2019a, 2019b). The CC domains in this pentamer form a funnel-like structure that was shown to be a calcium-permeable channel (J. Wang, Hu, et al. 2019; Bi et al. 2021). Two groups independently defined the structures of the two direct binder TNLs RPP1 and ROQ1 and found that they assemble into tetramers upon activation through their cognate effectors. These studies confirmed the importance of intramolecular interactions to maintain the inactive state of NLRs and intermolecular interactions for resistosome assembly and immune signaling.

NLRs have provided a valuable source of disease resistance in many crop species. Wheat stem rust disease is caused by the fungal agent *Puccinia graminis* f.sp. *tritici* (*Pgt*) and has been a major threat to wheat production, especially since the emergence of the highly virulent *Pgt* race Ug99. A large number of resistance genes that protect against Ug99 have been cloned and deployed in recent years (Zhang et al., 2021; Saintenac et al., 2013; Chen et al., 2017; Steuernagel et al., 2016; Arora et al., 2019). Among them are Sr50 and Sr33: two CNL resistance genes that are highly similar with an amino acid identity of ~80% and provide resistance against wheat stem rust (Mago et al., 2015; Periyannan et al., 2013). They have been introgressed into wheat from *Secale cereale* and *Aegilops tauschii*, respectively, and are closely related homologs of the barley *MLA* (RGH1) family of NLR genes (Mago et al., 2015; Periyannan et al., 2013; Halterman et al., 2001; Wei et al., 2002). This family has been associated with high allelic and functional diversification, exemplified by the more than 30 barley *MLA* alleles that confer resistance to powdery mildew (Seeholzer et al., 2010).

Sr50 has been shown to directly bind to its cognate effector AvrSr50 (Chen et al., 2017). The effector recognized by Sr33 remains unknown to date, but is assumed to be distinct from AvrSr50 based on resistance to different *Pgt* races. (Mago et al., 2015). Sr50 and Sr33 are interesting candidates to investigate as their autoactivity seems directly linked to the level of expression as observed by (Cesari et al., 2016) and this study, which suggests that they might be in a sensitive on-off-state. Moreover, with 80% sequence identity the recognition specificities of these NLRs must be defined by a small number of amino acids in their LRR regions.

Here we report the engineering of Sr33 to recognize the effector AvrSr50. We used a combination of phylogenetics and structural modeling to predict the region within Sr50's LRR that mediates binding to AvrSr50 and to identify individual residues involved in the interaction. We confirmed that the predicted region was sufficient to enable effector binding through the generation of a chimeric NLR that we named Sr33/50. This chimeric NLR was able to recognize and initiate cell death to AvrSr50 in wheat protoplasts and *Nicotiana benthamiana*. We subsequently generated the synthetic NLR Sr33_{syn} in which groups of residues within Sr33's LRR were changed to the corresponding sites of Sr50. The exchange of 12 amino acids was sufficient for Sr33 to initiate cell death in wheat

protoplasts and *N. benthamiana* in the presence of AvrSr50. We also observed that Sr50 could exhibit autoactivity in the wheat protoplast cell death assay, whereas Sr33 did not. We therefore used structural modeling to identify potential sites that could mediate this phenotype. We identified two residues that appear to be in close contact with the NB-ARC domain and mutation of these sites led to a stark reduction in autoactivity of Sr50. This study therefore demonstrates that the combination of sequence diversity analyses and structural modeling can be sufficient to identify and subsequently alter sites mediating effector recognition and autoactivity in direct binder NLRs.

Results

We have previously identified highly variable NLRs (hvNLRs) by analyzing natural diversity of pangenomes of two model species, *Arabidopsis thaliana* and *Brachypodium distachyon* (Prigozhin and Krasileva, 2021). The highly variable amino acid residues within hvNLR homology groups together with structural modeling were predictive of the cognate effector-binding sites in *Arabidopsis thaliana* (Prigozhin and Krasileva, 2021). We decided to translate this knowledge to crop species and test if the same approach can accurately predict residues mediating NLR-effector interaction and subsequently guide rational engineering of NLRs. For this work we chose the wheat hvNLRs Sr50 and Sr33, that have been independently introgressed into bread wheat from *A. tauschii* and rye, respectively (Fig. 1B). We initially assessed sequence diversity of the Sr33 and Sr50 homology groups with homologs obtained with BLAST search against the non-redundant database in The National Center for Biotechnology Information (NCBI). The multiple sequence alignments (MSA) of each homologous groups indicated there was sufficient sequence diversity to predict binding residues in Sr50 (D. Prigozhin, personal communication; Krasileva, Prigozhin, Staskawicz, Liu (2021). Patent No. 113469, WO: U.S.). This approach proves to be an easily accessible way to obtain sequences for Shannon entropy calculation. However, as Sr33 and Sr50 displayed ~80% sequence identity, many collected homologs were shared by the two groups. We therefore needed to refine the initial analysis through phylogenetic analysis of the MLA family to curate Sr33 and Sr50 homologous groups with non-overlapping sets of sequences. We inferred the phylogenetic tree with MLA family members collected from Pooideae species, 14 wheat genomes (Table S1) and selected experimentally validated NLRs (Kourelis et al., 2021). Despite complex evolution within Triticeae species (Fig. 1B), we were able to resolve the NLRs into four distinct phylogenetic sub-families representing Sr50, Sr33, MLA1 and HcMLA1 grouped by similar phylogenetic distances. This allowed us to calculate Shannon entropy for each phylogenetic group with at least 13 unique NLR sequences (Fig. 1C; Table S2). The entropy plots of these four phylogenetic groups appear almost identical across the whole NLR length, suggesting that the same positions of variable or invariable sites are conserved across the MLA gene family.

To incorporate the location of highly variable sites within the NLR structure, we created a plot displaying Shannon entropy measures of the residues within each leucine-rich repeat unit that form the beta sheets that form the inner concave surface of the LRR. We noticed that LRRs 8-14 contained a cluster of sites with high Shannon entropy that we hypothesized to be a potential effector binding pocket (Fig. 1D). Since high entropy plots

of the four phylogenetic groups were remarkably similar, we created the same plot displaying high Shannon entropy measures in key residues of the LRR units (Fig. S1). High Shannon entropy scores occur similarly towards the end of the LRRs in units 8-16 in all four phylogenetic groups. This suggests more specifically, that the effector binding pockets are in the LRR same region across homologous NLRs, despite their ability to recognize structurally distinct effector. These results indicate that highly variable sites within the NLRs of the MLA family could be where diversification of paralogs occurs.

To test our prediction, we created a chimeric protein consisting of Sr33 and the LRRs 8-14 of Sr50 that we will refer to as Sr33/50 throughout. A schematic illustrating the location of the transferred region within the chimera can be found in Figure S2. To test whether the chimera Sr33/50 is capable of recognizing AvrSr50 and subsequently inducing cell death, we utilized a wheat protoplast cell death assay, as previously described (Saur et al., 2019; Yoshida et al., 2009). In this assay, wheat protoplasts are transfected with luciferase and a combination of NLRs and effectors. A positive interaction between NLR and effector leads to cell death as exemplified by a lack of luciferase activity 2 days after transfection. Co-transfection of Sr50 with AvrSr50 led to cell death and strongly reduced luminescence, whereas co-transfection of Sr33 with AvrSr50 did not lead to a reduction as compared to NLRs alone (Fig. 2A). The chimeric protein Sr33/50 was able to recognize AvrSr50 as demonstrated by the strong reduction of luminescence signifying cell death (Fig. 2A). This result confirmed the prediction that LRRs 8-14 mediate effector recognition in Sr50 because the transfer of these LRRs to Sr33 was sufficient to transfer effector specificity.

We then aimed to identify the individual amino acids involved in the recognition of AvrSr50. In total, there are 26 amino acid polymorphisms in the LRRs 8-14 between Sr50 and Sr33. We mapped the variable residues onto the surface of Sr50 and Sr33 and selected 13 amino acids that fulfilled most of these criteria: high Shannon entropy signature, differing amino acid side chain property between Sr33 and Sr50 or exposure at the surface of the beta-sheets that face the NB-ARC domain (Fig. 2D and E). We synthesized several iterations of Sr33 carrying mutations in these predicted amino acid residues (with 10 to 13 amino acid substitutions) that correspond to the hypothesized binding sites in LRRs 8-13 of Sr50 and tested these synthetic NLRs in wheat protoplasts. Among the constructs we tested, the synthetic NLR Sr33_{syn} carries 12 selected amino acid polymorphisms (Fig. 2D) and initiated cell death in wheat protoplasts in an AvrSr50-dependent manner while Sr33_{syn-} did not (Fig. 2B). These two synthetic NLRs differ only in one amino acid residue that seems to affect recognition. While Sr33_{syn} carries a methionine in LRR13 (M840), the NLR Sr33_{syn-} carries a valine residue in the respective position (V840) (Fig. 2D, Fig. S3D). Our observation is therefore that 12 amino acid residues were the minimal amount needed to transfer recognition of AvrSr50 to Sr33. To substantiate this observation, we infiltrated these mutants in *N. benthamiana* with AvrSr50 to test if these mutations enable Sr33 to recognize AvrSr50 and induce cell death. Our results demonstrate that the transfer of 12 amino acids is sufficient to alter recognition specificity of Sr33 as it shows a clear cell death response when co-infiltrated with AvrSr50 (Fig. 2C). In contrast, the effector variant AvrSr50_QCMJC that evades recognition by Sr50 is also not recognized by Sr33_{syn}, demonstrating that the gained effector recognition

is AvrSr50-specific. These results demonstrate that we successfully predicted amino acids involved in the recognition of AvrSr50 through Sr50 and that effector recognition can be transferred to a closely related NLR.

In our protoplast cell death assays, we noticed that Sr50 could exhibit autoactivity whereas Sr33 and Sr33/50 did not. Furthermore, it has been published that Sr50 and Sr33 can exhibit autoactivity when over-expressed in *N. benthamiana*, but not when expressed under the weaker, native Sr50 promoter (Cesari et al., 2016). We observed the same in *N. benthamiana* where expression with *p35S* caused autoactivity whereas expression driven by a previously described, weaker *pRPP13* (Rentel et al., 2008) greatly reduced cell death in the absence of effector (Fig. S3). However, in wheat protoplasts only Sr50 exhibited varying levels of autoactivity while Sr33 consistently did not (Fig. 3A). Since the two NLRs are highly similar, we reasoned that we could reduce the Sr50 autoactive phenotype by changing certain amino acids to the corresponding sites in Sr33. In the iterations of synthetic Sr33 NLRs we observed that many of them exhibited autoactivity. Together with our data from Sr50, Sr33 and Sr33_{syn} we selected two residues to be mutated in Sr50 that we hypothesized to affect the autoactivity phenotype in wheat protoplasts. We tested the autoactivity of the Sr50 mutants in the wheat protoplast cell death assay (Fig. 3A) and observed reduced autoactivity of NLR alone while preserving effector recognition (Fig. 3A). We also tested whether effector recognition is retained and remains specific in *N. benthamiana*. We infiltrated Sr50 and the single and double mutants with both AvrSr50 and AvrSr50_QCMJC. All Sr50 variants induce cell death in an AvrSr50-dependent manner, while no cell death was observed with the AvrSr50_QCMJC effector variant (Fig. 3B). We therefore show that closely related NLRs can also provide information about activation mechanisms and be utilized to reduce autoactivity without affecting effector recognition specificity.

To understand why certain amino acid sites had such a drastic effect on Sr50's autoactivity, we modeled the structure of Sr50 with AlphaFold. The model showed that LRR residues M844 and V868 are in close contact with the NB-ARC domain, and we hypothesize that these residues are important in maintaining the inactive state of the NLR through the interaction with the NB-ARC domain (Fig. 3C). Upon calculation of the pairwise residual distances in the predicted structure, we found that two residues in the NB-ARC domain were in proximity with terminal LRR residues: G223 and L495 (Fig. S6). Among the two residues in the NB-ARC, only L495, that appears in a short stretch of random coils, was in contact with LRRs 8-13 (< 7.5 Å) (Fig. 3C; Fig. S6). We named the residues that define this particular inter-domain interaction the NB-ARC/LRR latch (Fig. 3C). Interestingly, of the specified seven residues involved in the NB-ARC/LRR latch only three were polymorphic between Sr50 and Sr33 (indicated as variable in Fig. 3C) and we identified two of these in our functional study of Sr50 mutants. Moreover, the residue M844 was required for effector recognition in Sr33_{syn}, as the variant Sr33_{syn}- lacks this methionine residue and did not recognize AvrSr50 (Fig. 2B, 2C; Fig. 2D). These results suggest that LRR residues within Sr50 and Sr33_{syn} that mediate effector binding also affect receptor autoactivation.

Since the change of a single amino acid resulted in such a large change of the autoactive phenotype, we were curious if there might be any co-evolutionary signatures between residues in the NB-ARC/LRR latch. Co-evolutionary inference relies on a deep MSA composed of diverse homologs. We therefore constructed an MSA with 21,518 putative CNLs from 330 Viridiplantae species and tested co-evolution with direct coupling analysis (DCA) (Zerihun et al., 2020) and Gremlin (Ovchinnikov et al., 2014). When we mapped the inference to the Sr50 sequence we found that residues located close to each other generally displayed co-evolutionary signature (Fig. 3D; Table S5). However, the NB-ARC domain and LRRs 8-13 (red dotted areas in Fig. 3D) lacked signal of coevolution, although the residues in the NB-ARC/LRR latch are physically close ($< 7.5 \text{ \AA}$, Fig S6). We also investigated whether the sites in the NB-ARC that are in proximity to the LRR latch correspond to any known, highly conserved major or minor motifs in the NB-ARC domain and found that they did not (Fig. 3E). For the MLA family member, although sequence conservation of the NB-ARC latch was not as high as that of the major motifs, glycine residue (G496) was highly conserved (Fig. 3C; Fig. 3E). We postulate that this glycine is important to make a sharp turn in the protein's tertiary structure and give rise to the structural characteristic of the NB-ARC latch. Together, our functional and structural analyses highlight a single LRR region, where residues which are in close physical proximity to the NB-ARC have the dual function in constraining autoactivity and enabling effector specificity.

Discussion

A large number of resistance genes deployed are NLRs, which makes them an excellent target for improvement of plant health. A bottleneck in the rational design of new NLR alleles has been the lack of understanding of effector binding and subsequent receptor activation. Though it has been known for a long time that the LRR domains play a crucial role in effector specificity (Ellis et al., 1999; Shen et al., 2003; Krasileva et al., 2010), difficulties in structural biology have limited our ability to close in on the specific sites mediating effector binding. To successfully generate new effector recognition specificities, we need to decipher sites within NLRs that mediate effector binding and those required to maintain stability and the inactive state. Recent advances both in structural biology (Martin et al., 2020; Wang et al., 2019a, 2019b) and NLR engineering (Cesari et al., 2021; De la Concepcion et al., 2019; Pottinger et al., 2020) have exemplified success in the engineering of NLR proteins.

With the increasing amount of genomic data of many different species and populations to analyze NLR evolution, it became possible to predict and engineer effector binding specificity. It was previously shown that the sequence diversity data and Shannon entropy can be used to predict if an NLR is a direct or indirect binder and where within the NLR effector binding occurs (Prigozhin and Krasileva, 2021). Here, we combined Shannon entropy with structural modeling to predict a potential effector binding site and guide rational engineering to transfer effector recognition specificity between distantly related homologs. In our structural analyses, we initially utilized PHYRE2 (Kelley et al., 2015) to model the Sr50 structure but the homologous template-based model was incomplete as the template structure is not fully resolved and the resulting model was therefore missing

several leucine-rich repeats. Since it was the only model available at the time, we used the incomplete model of the LRR repeats to the extent possible to identify surface-exposed sites for potential engineering. After the release of AlphaFold (Evans et al., 2021; Jumper et al., 2021), we recreated the Sr50 protein model and found that the LRRs of both models were consistent. By mapping Shannon entropy values and amino acid polymorphisms between Sr50 and Sr33 on the exposed beta-sheets of the LRR domain, we were able to use additional criteria to select residues for engineering as opposed to using just Shannon entropy scores alone. This approach greatly enhanced our ability to successfully engineer Sr33 as exemplified by the identification of the residue V840. Its corresponding site M844 in Sr50 has a low Shannon entropy score but was required for the transfer of effector recognition in Sr33_{syn}. We therefore show that the application of several analyses can be complementary and help successfully engineer NLR recognition specificities even in the absence of experimentally determined structural data.

The ‘on’ and ‘off’ states of NLRs are tightly controlled and recent cryo-electron microscopy structures of an indirect binder NLR, *Arabidopsis thaliana* HOPZ- ACTIVATED RESISTANCE 1 (ZAR1), confirmed what has previously been inferred through domain interaction and swap data: the inactive and ADP-bound state was monomeric and had multiple intramolecular contact points between the LRR and NB-ARC (Wang et al., 2019a, 2019b). Moreover, previous studies showed that the correct interactions of NB-ARC and LRR are required to avoid autoactivity of the NLR (Qi et al., 2012; Rairdan and Moffett, 2006; Sliotweg et al., 2013) and even effector responsiveness (Ma et al., 2018). The differing autoactivity phenotypes of Sr33, Sr50 and our Sr33_{syn} variants allowed us to investigate this as the differences among all three proteins was contained to their LRR domains. The data from this study supports the notion that interactions between NB-ARC and LRR are crucial in maintaining the inactive state of the direct binder NLR Sr50. A recent paper suggested that the interaction between the NB-ARC and LRR domains serves the purpose of preventing NLR oligomerization. (Zhao et al., 2021). It would be interesting to see if residues in the NB-ARC/LRR latch also play a role in NLR oligomerization or whether the interaction is merely to prevent access to the sites mediating nucleotide exchange and oligomerization.

Through our introduction of amino acid changes in Sr50 and structural modeling, we identified a direct interaction point of NB-ARC and LRR that we termed NB-ARC/LRR latch. This interaction most likely results in a conformation of the NLR that hinders the exchange of ADP to ATP in the NB-ARC domain or oligomerization into resistosomes. Strikingly, our data suggests that control of autoactivity and effector specificity are located within the same region of the NLR and mediated by the same residues. We hypothesize that the effector competes with and disrupts the interaction of the NB-ARC/LRR latch leading to the activation of the NLR. Our data suggests that this interaction can also be disrupted through the exchange of amino acid residues in the NB-ARC/LRR latch. We therefore hypothesize that the disruption of NB-ARC/LRR latch is a crucial step in NLR activation and can be caused by effector binding or unstable intramolecular interactions. We therefore show that the strength of the NB-ARC/LRR latch interaction is an important criterion to consider when undertaking NLR engineering approaches.

It has been proposed that NLRs exist in fluent states of active and inactive and effector binding stabilizes the active state thereby shifting the equilibrium in its favor – this is referred to as the equilibrium-based switch model (Bernoux et al., 2016). We propose that the activation of hvNLRs, such as Sr50, follows the equilibrium-based switch model (Bernoux et al., 2016) based on our observations of expression-dependent autoactivity and our ability to change this phenotype with a few amino acid substitutions. According to this model there is an equal pool of activated and inactive receptors in each cell in the absence of a trigger. Effector binding stabilizes the activated conformation and leads to the activation of downstream signaling pathways. This implies that the cell can cope with a certain amount of activated receptors and overexpression would lead to an excess of activated receptors. This could push the resulting cellular changes over the threshold required to initiate the cell death response, for instance through increased calcium influx caused by the insertion of CC domains of activated, oligomerized receptors into the membrane. Collectively, our data, which showed that Sr50 exhibits varying degrees of autoactivity depending on its expression level (Fig. 3A, Fig. S3), suggests that this NLR follows the equilibrium switch model of activation. It will therefore be crucial to understand the limitations of deploying NLRs that follow this model of activation and ensure the use of appropriate promoters in the generation of new resistant plants with hvNLRs. We hypothesize that having multiple factors that can drastically change autoactivity phenotypes such as expression levels and intramolecular interaction strengths is what allows for the rapid generation of diversity in hvNLRs.

It is tempting to hypothesize that the amino acid sites we identified in the latch would evolve together to maintain the rate at which the NLR can be activated. However, neither of the analyses we performed detected any co-evolutionary signature and even suggested that co-evolution between NB-ARC and terminal LRRs is lacking. Since there are at least seven residues involved in the NB-ARC/LRR latch, it is conceivable that any co-evolutionary signature would be too diluted among all involved residues to be detected or that the activation mechanism of the NLR is too complicated to be summarized by just three residues. The fluidity of Sr50's activation state is another possible reason for the lack of measurable co-evolution, as the 'on' and 'off' states of this NLR are affected by many different factors. Moreover, the co-evolutionary analyses that are currently at our disposal require rather large datasets with an evolutionarily broad range of NLRs. If the co-evolution of the NB-ARC/LRR latch is lineage-specific, and multiple modes of interactions exist for these inter-domains, the datasets may obscure any co-evolutionary signatures. Elucidating the co-evolution of the NB-ARC/LRR latch will require better understanding of how NLRs are cycled between active and inactive states (there is no experimental structure of an inactive direct binder NLR), how many different modes there may be, and which NLRs follow each mode of activation.

This study shows a major advance towards the successful engineering of new effector recognition specificities in direct binder NLRs. We demonstrate that effector binding sites predicted through diversity analysis can be successfully used to engineer or transfer recognition specificities. Our study also demonstrates that for the successful engineering of an immune receptor we also need to identify sites that are important for maintaining the inactive state of the NLR. In addition to the generation of novel NLRs, we show that

existing NLRs can be engineered to recognize a new effector and more importantly recognition can be transferred between closely related NLRs. This means we might be able to generate resistant plants through precise genome editing of NLR homologs. Through this approach we can use existing NLRs that are available for deployment through map-based cloning approaches and edit their homologs in elite crop germplasms. The region required for the transfer of effector recognition involved as little as 12 amino acid substitutions within 450 bp. When this region were to be introduced into a closely related homolog through homologous recombination the engineered plants would follow the regulatory exemptions described in the SECURE rule from the USDA, given this variation is already present in the edited plant's gene pool (Hoffman, 2021). The ability to transfer effector recognition between close homologs also gives us the ability to avoid interspecies transfer of genes, which is controversial for some opponents of transgenic plants. The information provided in this study may therefore help to drive forward the generation of resistant elite crop germplasm.

Materials and Methods

Plant materials and growth conditions

Nicotiana benthamiana plants were grown in a Conviron growth chamber with 16 hours of 150 μ mol/m²/s light at 24°C. *Triticum aestivum* cultivar Fielder seeds were sterilized in 10% bleach solution for 20 minutes. They were placed on water and filter paper in a plastic cup and the cup was sealed by placing another cup on top and adhering them together with Millipore tape. The cups were kept on the west-facing windowsill for 7-14 days, depending on the season.

Cloning and vectors

All vectors cloned for this study were cloned following the Golden Gate cloning technique (Weber et al., 2011; Werner et al., 2012) by utilizing the vectors deposited as MoClo Plant Parts (#1000000047) and Tool kits (#1000000044) at Addgene. For expression in wheat protoplasts, all genes were driven by the *pZmUBQ* promoter (pICSL12009) and the *tNos* terminator (pICH41421), whereas for expression in *N. benthamiana* the *pRPP13* promoter (Rentel et al., 2008) and *tNos* terminator, as well as *p35S* (pICH51277) and *t35S* (pICH41414) were used. The following genes were domesticated for Golden Gate by removing internal restriction sites for *BsaI* and *BpiI* and were cloned into the corresponding level 0 acceptor: Sr33, Sr50, AvrSr50, AvrSr50_QCMJC, and pRPP13. All coding sequences were cloned both with and without stop codon. Mutants of Sr33 were created by synthesizing the LRRs 8-13 with 10 amino acid substitutions and ligating it together with PCR fragments of the start and end of Sr33. Additional mutations were added using site-directed mutagenesis PCR, by designing two overlapping primers with the desired mutation in the middle, amplifying the vector by Phusion PCR and digesting the template with *DpnI*. Mutants of Sr50 were created by site-directed mutagenesis PCR. All vectors and primers used in this study are described in Table S5 and S6.

Wheat protoplast isolation, transfection, and cell death assay

Wheat protoplast isolation and transfection was adapted from (Saur et al., 2019; Yoshida et al., 2009). The epidermis of 7-14 days old seedlings was peeled off by making a shallow

cut in the adaxial epidermis with a sharp razor blade, folding the leaf over at the cutting edge and pulling back the leaf tip with the abaxial epidermis. The peeled leaf sections were placed next to each other on masking tape with the abaxial side facing upwards and placed into enzyme solution facing down. Washing and transfection were performed as previously described. Protoplast number was counted with a hemocytometer and adjusted to 300,000 protoplasts/mL. All plasmids were transfected at a concentration of 15 µg and an empty vector was added to ensure all protoplasts were transfected with the same final DNA concentration. After transfection, protoplasts were incubated in a BSA-coated 12-well plate at room temperature in the dark for 48 hours. Subsequently, protoplasts were collected by centrifugation at 100g for 3 minutes. Supernatant was discarded and protoplasts resuspended in 2x Cell Culture Lysis Reagent (Promega, Cat. #E1531) and incubated on ice for 20 minutes in the dark. The lysed cell contents were centrifuged at 1,000g for 3 minutes and divided into a white 96-well plate (Greiner Bio-One, Cat. # 655075) in triplicate. An equal amount of Luciferase Assay Substrate (Promega, Cat. #E4550) was added and luminescence was measured in a TECAN Infinite F Plex plate reader (20 minutes, settle time 100 ms, no attenuation).

Agrobacterium-mediated transformation of *N. benthamiana*

Agrobacterium tumefaciens GV3101:pMP90 transformed with the corresponding construct were grown in LB medium (supplemented with 50 µg/ml Rifampicin, 25 µg/ml Gentamycin and 100 µg/ml Carbenicillin) at 28°C. The cultures were centrifuged at 3,000 g for 15 minutes and the pellet resuspended in infiltration medium (10 mM MES pH=5.6, 10 mM MgCl₂). Optical density was adjusted to a final OD₆₀₀=0.3. If several constructs were co-infiltrated the suspensions of individual constructs were mixed in a 1:1 ratio while ensuring that all constructs reached a final density of 0.3. Leaves of 4-5 weeks old *N. benthamiana* plants were inoculated with the suspension using a blunt syringe. Pictures of leaves were taken 2-3 days after inoculation.

Manual curation of NLRs

To ensure sequence diversity needed to calculate entropy, we manually curated a subset of NLRs from 14 wheat genomes (Walkowiak et al., 2020) (Table S1). From the wheat reference genome, we first identified the NLRs that belong to the MLA family, guided by the previous study (Steuernagel et al., 2020). Their NB-ARC domain was searched with TBLASTN (-evalue 1E-10) (Camacho et al., 2009) against the 14 wheat genomes. We identified matches that showed 85% or more sequence identity against the queries and extracted the putative NLR loci with 10,000 flanking regions on both ends. We used MAKER v3.01.04 (Cantarel et al., 2008) to obtain gene models in these regions, relying on two ab initio predictors, Augustus v3.4.0 (--species wheat) (Stanke et al., 2006) and GeneMark v4.68_lic (Brůna et al., 2020), as well as CDS and protein sequences of the wheat reference genome annotation. The gene models were loaded into Apollo genome browser v2.0.6 (Dunn et al., 2019) and manually curated. The manual curation generated many redundant NLRs from the 14 wheat genomes, the sequences of which are nearly identical. The redundant sequences can inflate their weighting and copy numbers in entropy calculations and the phylogenetic analyses, which can mislead evolutionary conclusions. We therefore reduced redundancy of the wheat NLRs based on their sequence identity (95%) with CD-hit (-c 0.95) v4.8.1 (Li and Godzik, 2006).

Phylogenetic analyses

To infer a species tree of the selected Pooideae members, we first identified complete BUSCO genes present in all nine species with BUSCO v5.2.2 (Seppey et al., 2019) and the poales_odb10 database and randomly selected 500 (Table S1). Each set of orthologous protein sequences was aligned with MAFFT v7.487 (--maxiterate 1000 --globalpair) (Kato and Standley, 2013). All MSAs were concatenated, trimmed with TrimAl v1.4.rev15 (-gt 0.3) (Capella-Gutiérrez et al., 2009) and used to infer a phylogenetic tree with FastTree v2.1.10 (Price et al., 2010).

To infer the MLA family tree, we collected the MLA family members from the Pooideae species, filtered wheat NLR annotation sets and experimentally validated NLRs (Kourelis et al., 2021), guided by the classification of the previous study (Steuernagel et al., 2020). Their full-length sequences were aligned with MAFFT, and we trimmed out all columns where both Sr33 and Sr50 contained gaps. The resulting MSA was used to infer a phylogenetic tree with FastTree (-slow). Sr35 was used as an outgroup to root the tree.

Protein structure prediction and structural analysis

We used AlphaFold v2.0.0 to predict the structure of Sr50 (Jumper et al., 2021; Evans et al., 2021). The homologs were collected from the full databases, and homologous templates downloadable by July 20th, 2020, were included in the template database. Models 1, 3, 4 and 5 as well as ptm model 2 were used for structural inference. The best model was selected as a final structure based on the average predicted IDDT scores. We calculated the distances between residues in the predicted structure with Biopython (Cock et al., 2009). The structure was visualized through PyMOL v2.5.

Co-evolutionary analyses

To collect CNLs for co-evolutionary analyses, we extracted the protein sequences from the Viridiplantae species from the uniref100 database (2022-0222). The gene models containing the NB-ARC domain (PF00931) were identified as putative NLRs with hmmsearch 3.1b2 (--domE 1e-4 -E 1e-4) and annotated with InterProScan v5.52-86.0 (-appl Pfam-33.1) (Eddy, 2011; Jones et al., 2014). Any sequences with other PFAM entries than NB-ARC domain, Rx N-terminal domain (PF18052) and Leucine Rich Repeat (PF00560, PF07725 and PF13855) were excluded. The NB-ARC domain of the remaining NLRs were aligned with MAFFT (--auto --global), trimmed with TrimAl (-gt 0.3) and used to infer a phylogenetic tree with FastTree. We identified remaining TNL and RNL clades and removed the clade members.

We searched this database with Sr50 as a query with jackhmmer (--domE 1e-10 -E 1e-10 -N 5). The MSA was trimmed with respect to Sr50, and homologs without $\geq 75\%$ coverage were removed. We used this filtered MSA for DCA (Zerihun et al., 2020) and Gremlin (Ovchinnikov et al., 2014). The pseudolikelihood maximization algorithm of the pydca package was run for DCA (--max_ iterations 1000 --apc). The filtered MSA was also submitted to the online Gremlin server (HHblits iterations = 0, Coverage = 50, Remove gaps = 50).

Shannon entropy calculation

We computed normalized Shannon entropy ($-\sum_{i=1}^{20} \frac{p_i \log_2 p_i}{\log_2 20}$) where p_i is the probability of observing an amino acid, i , out of the twenty in the given position of the input MSA. Gaps were ignored in entropy calculation. However, if the position of the MSA was composed of 50% or more gap characters, the entropy was not calculated. The normalized Shannon entropy ranges from 0 (no variability) to 1 (maximum variability). The normalized entropy score of 0.347 was used to define highly variable residues, following the previous study (Jumper et al., 2021; Prigozhin and Krasileva, 2021).

Data Availability

All data and scripts used in this study are available in Github: https://github.com/krasileva-group/Sr33-Sr50_analysis.

Acknowledgements

We thank Dr. Isabel Saur and Salome Wilson for helpful practical advice on the protoplast cell death assay. We thank Dr. Daniil Prigozhin for initial Shannon entropy analyses and effector binding site prediction on Sr50, and helpful comments on the manuscript. JAT is grateful for many personal discussions with Dr. Erin Baggs about this project and thoughtful suggestions on the bioinformatics analyses. We are grateful to all members of the Krasileva lab for thoughtful comments and discussions of the presented material. The project has been funded by Gordon and Betty Moore Foundation, the Foundation for Food and Agriculture and 2Blades Foundation. KS has been supported by the Berkeley Graduate Fellowship.

Competing Interests

KVK, BJS and FL have previously filed a patent application related to this work (Patent publication number: WO/2021/113569, Publication date: 10.06.2021. International Application No. PCT/US2020/063203. International Filing Date: 04.12.2020).

References

- Arora, S. et al.** (2019). Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nat. Biotechnol.* **37**: 139–143.
- Bernoux, M., Burdett, H., Williams, S.J., Zhang, X., Chen, C., Newell, K., Lawrence, G.J., Kobe, B., Ellis, J.G., Anderson, P.A., and Dodds, P.N.** (2016). Comparative Analysis of the Flax Immune Receptors L6 and L7 Suggests an Equilibrium-Based Switch Activation Model. *Plant Cell* **28**: 146–159.
- Brúna, T., Lomsadze, A., and Borodovsky, M.** (2020). GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom Bioinform* **2**: lqaa026.

- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L.** (2009). BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.
- Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., and Yandell, M.** (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**: 188–196.
- Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T.** (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.
- Cesari, S. et al.** (2013). The rice resistance protein pair RGA4/RGA5 recognizes the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by direct binding. *Plant Cell* **25**: 1463–1481.
- Cesari, S., Moore, J., Chen, C., Webb, D., Periyannan, S., Mago, R., Bernoux, M., Lagudah, E.S., and Dodds, P.N.** (2016). Cytosolic activation of cell death and stem rust resistance by cereal MLA-family CC-NLR proteins. *Proc. Natl. Acad. Sci. U. S. A.* **113**: 10204–10209.
- Cesari, S., Xi, Y., Declerck, N., Chalvon, V., Mammri, L., Pugnière, M., Henriquet, C., de Guillen, K., Padilla, A., and Kroj, T.** (2021). Design of a new effector recognition specificity in a plant NLR immune receptor by molecular engineering of its integrated decoy domain.
- Chapman, S., Stevens, L.J., Boevink, P.C., Engelhardt, S., Alexander, C.J., Harrower, B., Champouret, N., McGeachy, K., Van Weymers, P.S.M., Chen, X., Birch, P.R.J., and Hein, I.** (2014). Detection of the virulent form of AVR3a from *Phytophthora infestans* following artificial evolution of potato resistance gene R3a. *PLoS One* **9**: e110158.
- Chen, J. et al.** (2017). Loss of AvrSr50 by somatic exchange in stem rust leads to virulence for Sr50 resistance in wheat. *Science* **358**: 1607–1610.
- Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and de Hoon, M.J.L.** (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**: 1422–1423.
- De la Concepcion, J.C., Franceschetti, M., MacLean, D., Terauchi, R., Kamoun, S., and Banfield, M.J.** (2019). Protein engineering expands the effector recognition profile of a rice NLR immune receptor. *Elife* **8**.
- DeYoung, B.J., Qi, D., Kim, S.-H., Burke, T.P., and Innes, R.W.** (2012). Activation of a plant nucleotide binding-leucine rich repeat disease resistance protein by a modified self protein. *Cellular Microbiology* **14**: 1071–1084.

- Dodds, P.N. and Rathjen, J.P.** (2010). Plant immunity: towards an integrated view of plant–pathogen interactions. *Nat. Rev. Genet.* **11**: 539.
- Dunn, N.A., Unni, D.R., Diesh, C., Munoz-Torres, M., Harris, N.L., Yao, E., Rasche, H., Holmes, I.H., Elsik, C.G., and Lewis, S.E.** (2019). Apollo: Democratizing genome annotation. *PLoS Comput. Biol.* **15**: e1006790.
- Eddy, S.R.** (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**: e1002195.
- Ellis, J.G., Lawrence, G.J., Luck, J.E., and Dodds, P.N.** (1999). Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. *Plant Cell* **11**: 495–506.
- Evans, R. et al.** (2021). Protein complex prediction with AlphaFold-Multimer. *bioRxiv*.
- Feehan, J.M., Castel, B., Bentham, A.R., and Jones, J.D.** (2020). Plant NLRs get by with a little help from their friends. *Curr. Opin. Plant Biol.* **56**: 99–108.
- Giannakopoulou, A., Steele, J.F.C., Segretin, M.E., Bozkurt, T.O., Zhou, J., Robatzek, S., Banfield, M.J., Pais, M., and Kamoun, S.** (2015). Tomato I2 Immune Receptor Can Be Engineered to Confer Partial Resistance to the Oomycete *Phytophthora infestans* in Addition to the Fungus *Fusarium oxysporum*. *Mol. Plant. Microbe. Interact.* **28**: 1316–1329.
- Grund, E., Tremousaygue, D., and Deslandes, L.** (2019). Plant NLRs with integrated domains: Unity makes strength. *Plant Physiol.* **179**: 1227–1235.
- Halterman, D., Zhou, F., Wei, F., Wise, R.P., and Schulze-Lefert, P.** (2001). The MLA6 coiled-coil, NBS-LRR protein confers AvrMLA6-dependent resistance specificity to *Blumeria graminis* f. sp. *hordei* in barley and wheat. *Plant J.* **25**: 335–348.
- Hoffman, N.E.** (2021). Revisions to USDA biotechnology regulations: The SECURE rule. *Proc. Natl. Acad. Sci. U. S. A.* **118**: e2004841118.
- Jones, P. et al.** (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240.
- Jumper, J. et al.** (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**: 583–589.
- Katoh, K. and Standley, D.M.** (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**: 772–780.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E.** (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**: 845–858.

- Kim, S.H., Qi, D., Ashfield, T., Helm, M., and Innes, R.W.** (2016). Using decoys to expand the recognition specificity of a plant disease resistance protein. *Science* **351**: 684–687.
- Kourelis, J., Sakai, T., Adachi, H., and Kamoun, S.** (2021). RefPlantNLR is a comprehensive collection of experimentally validated plant disease resistance proteins from the NLR family. *PLoS Biol.* **19**: e3001124.
- Krasileva, K.V., Dahlbeck, D., and Staskawicz, B.J.** (2010). Activation of an Arabidopsis resistance protein is specified by the in planta association of its leucine-rich repeat domain with the cognate oomycete effector. *Plant Cell* **22**: 2444–2458.
- Lewis, J.D., Wu, R., Guttman, D.S., and Desveaux, D.** (2010). Allele-specific virulence attenuation of the *Pseudomonas syringae* HopZ1a type III effector via the Arabidopsis ZAR1 resistance protein. *PLoS Genet.* **6**: e1000894.
- Li, W. and Godzik, A.** (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.
- Liu, Y., Zhang, X., Yuan, G., Wang, D., Zheng, Y., Ma, M., Guo, L., Bhadauria, V., Peng, Y.-L., and Liu, J.** (2021). A designer rice NLR immune receptor confers resistance to the rice blast fungus carrying noncorresponding avirulence effectors. *Proc. Natl. Acad. Sci. U. S. A.* **118**: e2110751118.
- Ma, S. et al.** (2020). Direct pathogen-induced assembly of an NLR immune receptor complex to form a holoenzyme. *Science* **370**: eabe3069.
- Ma, Y., Guo, H., Hu, L., Martinez, P.P., Moschou, P.N., Cevik, V., Ding, P., Duxbury, Z., Sarris, P.F., and Jones, J.D.G.** (2018). Distinct modes of derepression of an Arabidopsis immune receptor complex by two different bacterial effectors. *Proc. Natl. Acad. Sci. U. S. A.* **115**: 10218–10227.
- Mago, R. et al.** (2015). The wheat Sr50 gene reveals rich diversity at a cereal disease resistance locus. *Nat. Plants* **1**: 15186.
- Maqbool, A., Saitoh, H., Franceschetti, M., Stevenson, C.E.M., Uemura, A., Kanzaki, H., Kamoun, S., Terauchi, R., and Banfield, M.J.** (2015). Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *Elife* **4**.
- Martin, R., Qi, T., Zhang, H., Liu, F., King, M., Toth, C., Nogales, E., and Staskawicz, B.J.** (2020). Structure of the activated ROQ1 resistosome directly recognizing the pathogen effector XopQ. *Science* **370**: eabd9993.
- Ovchinnikov, S., Kamisetty, H., and Baker, D.** (2014). Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife* **3**: e02030.

- Periyannan, S. et al.** (2013). The gene Sr33, an ortholog of barley MLA genes, encodes resistance to wheat stem rust race Ug99. *Science* **341**: 786–788.
- Pottinger, S.E., Bak, A., Margets, A., Helm, M., Tang, L., Casteel, C., and Innes, R.W.** (2020). Optimizing the PBS1 decoy system to confer resistance to Potyvirus infection in Arabidopsis and soybean. *Mol. Plant. Microbe. Interact.* **33**: 932–944.
- Price, M.N., Dehal, P.S., and Arkin, A.P.** (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.
- Prigozhin, D.M. and Krasileva, K.V.** (2021). Analysis of intraspecies diversity reveals a subset of highly variable plant immune receptors and predicts their binding sites. *Plant Cell* **33**: 998–1015.
- Qi, D., DeYoung, B.J., and Innes, R.W.** (2012). Structure-function analysis of the coiled-coil and leucine-rich repeat domains of the RPS5 disease resistance protein. *Plant Physiol.* **158**: 1819–1832.
- Racman, D.S., Mcgeachy, K., Reavy, B., Strukelj, B., Zel, J., and Barker, H.** (2001). Strong resistance to potato tuber necrotic ringspot disease in potato induced by transformation with coat protein gene sequences from an NTN isolate of Potato virus Y. *Ann. Appl. Biol.* **139**: 269–275.
- Rairdan, G.J. and Moffett, P.** (2006). Distinct domains in the ARC region of the potato resistance protein Rx mediate LRR binding and inhibition of activation. *Plant Cell* **18**: 2082–2093.
- Rentel, M.C., Leonelli, L., Dahlbeck, D., Zhao, B., and Staskawicz, B.J.** (2008). Recognition of the *Hyaloperonospora parasitica* effector ATR13 triggers resistance against oomycete, bacterial, and viral pathogens. *Proc. Natl. Acad. Sci. U. S. A.* **105**: 1091–1096.
- Saintenac, C., Zhang, W., Salcedo, A., Rouse, M.N., Trick, H.N., Akhunov, E., and Dubcovsky, J.** (2013). Identification of wheat gene Sr35 that confers resistance to Ug99 stem rust race group. *Science* **341**: 783–786.
- Saur, I.M.L., Bauer, S., Lu, X., and Schulze-Lefert, P.** (2019). A cell death assay in barley and wheat protoplasts for identification and validation of matching pathogen AVR effector and plant NLR immune receptors. *Plant Methods* **15**: 118.
- Seeholzer, S., Tsuchimatsu, T., Jordan, T., Bieri, S., Pajonk, S., Yang, W., Jahoor, A., Shimizu, K.K., Keller, B., and Schulze-Lefert, P.** (2010). Diversity at the MLA powdery mildew resistance locus from cultivated barley reveals sites of positive selection. *Mol. Plant. Microbe. Interact.* **23**: 497–509.
- Segretin, M.E., Pais, M., Franceschetti, M., Chaparro-Garcia, A., Bos, J.I.B., Banfield, M.J., and Kamoun, S.** (2014). Single amino acid mutations in the potato immune receptor R3a expand response to *Phytophthora* effectors. *Mol. Plant. Microbe. Interact.* **27**: 624–637.

- Seppey, M., Manni, M., and Zdobnov, E.M.** (2019). BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **1962**: 227–245.
- Shao, Z.-Q., Xue, J.-Y., Wu, P., Zhang, Y.-M., Wu, Y., Hang, Y.-Y., Wang, B., and Chen, J.-Q.** (2016). Large-scale analyses of angiosperm nucleotide-binding site-leucine-rich repeat genes reveal three anciently diverged classes with distinct evolutionary patterns. *Plant Physiol.* **170**: 2095–2109.
- Shen, Q.-H., Zhou, F., Bieri, S., Haizel, T., Shirasu, K., and Schulze-Lefert, P.** (2003). Recognition specificity and RAR1/SGT1 dependence in barley MLA disease resistance genes to the powdery mildew fungus. *Plant Cell* **15**: 732–744.
- Slootweg, E. et al.** (2017). Sequence Exchange between Homologous NB-LRR Genes Converts Virus Resistance into Nematode Resistance, and Vice Versa. *Plant Physiol.* **175**: 498–510.
- Slootweg, E.J., Spiridon, L.N., Roosien, J., Butterbach, P., Pomp, R., Westerhof, L., Wilbers, R., Bakker, E., Bakker, J., Petrescu, A.-J., Smant, G., and Goverse, A.** (2013). Structural determinants at the interface of the ARC2 and leucine-rich repeat domains control the activation of the plant immune receptors Rx1 and Gpa2. *Plant Physiol.* **162**: 1510–1528.
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B.** (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**: W435-9.
- Steuernagel, B. et al.** (2020). The NLR-Annotator tool enables annotation of the intracellular immune receptor repertoire. *Plant Physiol.* **183**: 468–482.
- Steuernagel, B., Periyannan, S.K., Hernández-Pinzón, I., Witek, K., Rouse, M.N., Yu, G., Hatta, A., Ayliffe, M., Bariana, H., Jones, J.D.G., Lagudah, E.S., and Wulff, B.B.H.** (2016). Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat. Biotechnol.* **34**: 652–655.
- Tamborski, J. and Krasileva, K.V.** (2020). Evolution of plant NLRs: From natural history to precise modifications. *Annu. Rev. Plant Biol.* **71**: 355–378.
- Walkowiak, S. et al.** (2020). Multiple wheat genomes reveal global variation in modern breeding. *Nature* **588**: 277–283.
- Wang, G. et al.** (2015). The decoy substrate of a pathogen effector and a pseudokinase specify pathogen-induced modified-self recognition and immunity in plants. *Cell Host Microbe* **18**: 285–295.
- Wang, J., Hu, M., Wang, J., Qi, J., Han, Z., Wang, G., Qi, Y., Wang, H.-W., Zhou, J.-M., and Chai, J.** (2019a). Reconstitution and structure of a plant NLR resistosome conferring immunity. *Science* **364**: eaav5870.

- Wang, J., Wang, J., Hu, M., Wu, S., Qi, J., Wang, G., Han, Z., Qi, Y., Gao, N., Wang, H.-W., Zhou, J.-M., and Chai, J.** (2019b). Ligand-triggered allosteric ADP release primes a plant NLR complex. *Science* **364**: eaav5868.
- Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S.** (2011). A modular cloning system for standardized assembly of multigene constructs. *PLoS One* **6**: e16765.
- Wei, F., Wing, R.A., and Wise, R.P.** (2002). Genome dynamics and evolution of the MLA (powdery mildew) resistance locus in barley. *Plant Cell* **14**: 1903–1917.
- Werner, S., Engler, C., Weber, E., Gruetzner, R., and Marillonnet, S.** (2012). Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. *Bioeng. Bugs* **3**: 38–43.
- Yoshida, K., Saitoh, H., Fujisawa, S., Kanzaki, H., Matsumura, H., Yoshida, K., Tosa, Y., Chuma, I., Takano, Y., Win, J., Kamoun, S., and Terauchi, R.** (2009). Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell* **21**: 1573–1591.
- Zerihun, M.B., Pucci, F., Peter, E.K., and Schug, A.** (2020). pydca v1.0: a comprehensive software for direct coupling analysis of RNA and protein sequences. *Bioinformatics* **36**: 2264–2265.
- Zhang, J. et al.** (2021). A recombined Sr26 and Sr61 disease resistance gene stack in wheat encodes unrelated NLR genes. *Nat. Commun.* **12**: 3378.
- Zhao, X. et al.** (2021). The Sw-5b NLR nucleotide-binding domain plays a role in oligomerization, and its self-association is important for activation of cell death signaling. *J. Exp. Bot.* **72**: 6581–6595.

A

B

D

C

Fig 1. Evolutionary analyses and Shannon entropy scores of Sr33 and Sr50 related MLAs reveal variable hotspots in the NLR structure.

A. The phylogenetic tree of selected MLA family members. The bootstraps are provided only for nodes used to distinguish the four phylogenetic groups or higher. The four phylogenetic clades were named based on the representative, experimentally validated NLRs that belong to the groups: Sr50, HcMLA1, Sr33 and MLA. All phylogenetic groups are supported with bootstrapping values ≥ 98 , except for the HcMLA1 family that has a bootstrapping value of 63. The tree was rooted with the outgroup Sr35. **B.** The species tree of Pooideae members. The source of Sr33 and Sr50 introgression is indicated, and branches are colored based on tribe membership. No MLA family members were identified from the Poeae species, and these species are indicated with empty circles. **C.** The normalized Shannon entropy of the four phylogenetic group members in A. The normalized Shannon entropy was calculated with the indicated number of members (N) in each phylogenetic group. The normalized Shannon entropy ranges from 0 (no variability) to 1 (complete variability), and the red line indicates the previously used cut-off (Prigozhin and Krasileva 2021) to determine highly variable residues (Entropy = 1.5 or normalized entropy = 0.347). **D.** The distribution of the normalized Shannon entropy across the leucine-rich repeat (LRR) domain of the Sr50 group. The LRR region of the normalized Shannon entropy for the Sr50 group provided in C is depicted as a heatmap to provide a closer view of sequence variations. Each LRR unit and each position within the LRR unit was defined according to the general motif LXXLXLXX(C/N).

Fig 2: Transfer of domains and amino acid groups from Sr50 to Sr33 enables the recognition of the previously unrecognized effector AvrSr50.

A. The chimeric protein Sr33/50 recognizes AvrSr50 and induces cell death in the wheat protoplast cell death assay. Graph shows data from four independent biological replicates and asterisk indicates significant difference between EV and AvrSr50 treatment (paired t-test, $n=4$, $*p < 0.05$). Seedlings of the wheat cultivar Fielder were peeled and digested in enzyme solution for protoplast generation. After successful protoplast isolation they were transfected with constructs for the wheat protoplast assay as described previously (Saur et al., 2019). Protoplasts were harvested 2 days later and luminescence was measured in a TECAN Infinite F Plex plate reader utilizing the Promega Luciferase Assay System (Promega, Cat. #E4550). **B.** Twelve amino acid changes on Sr33 are sufficient to enable recognition of AvrSr50 as exemplified by the synthetic NLR Sr33_{syn}. We introduced twelve amino acid changes into the NLR Sr33 that is unable to induce cell death in wheat protoplasts in response to AvrSr50 and named the new NLR Sr33_{syn}. In contrast to Sr33, Sr33_{syn} is able to recognize and execute cell death in an AvrSr50-dependent manner. The synthetic NLR Sr33_{syn} that only contains 11 amino acid substitutions does not recognize AvrSr50. Graph shows data from three independent biological replicates and asterisk indicates significant difference between EV and AvrSr50 treatment (paired t-test, $n=3$, $**p < 0.01$). **C.** Sr33_{syn} induces cell death in response to AvrSr50 in *N. benthamiana*. None of the NLRs initiated cell death in response to the effector variant AvrSr50_QCMJC. Leaves of 4-5 weeks old *N. benthamiana* plants were infiltrated with *A. tumefaciens* carrying NLRs and effectors at OD₆₀₀=0.3 and 0.4, respectively. UV images were taken on a BioRad Gel Imager 3-5 days after infiltration. Numbers on images correspond to the numbers of leaves that showed the depicted phenotype. **D.** Amino acid residues in Sr33 and Sr50 targeted for swapping. The amino acid sequences and positions in LRRs 8-13 are depicted if the sequences are characterized by substitutions between Sr33 and Sr50. Low and high entropy were determined based on the cut-off (0.347) and are indicated in different colors. The 12 amino acid sequences used to generate Sr33_{syn} are indicated as 'swapped'. **E.** Variable amino acids mapped to the Sr50 structure. The structure of Sr50 was predicted with AlphaFold (Jumper et al. 2021), and only the NB-ARC domain and LRRs are shown. The residues targeted for swapping are annotated.

A

B

C

E

D

Fig 3: Sr50 autoactivity is decreased through mutations in its LRR.

A. Single and double mutants of Sr50 show decreased autoactivity in wheat protoplasts compared to wild type Sr50. Graph shows data from three independent biological replicates and asterisk indicates significant difference between EV and AvrSr50 treatment (paired t-test, $n=3$, $**p < 0.01$). We introduced single and double amino acid changes into the NLR Sr50 to decrease its autoactivity. The sites changed were changed to the corresponding amino acids in that position within Sr33 as this NLR does not exhibit the same autoactivity phenotype in wheat protoplasts. **B.** Single and double mutants of Sr50 induce cell death in *N. benthamiana* in response to AvrSr50. None of the NLRs initiated cell death in response to the effector variant AvrSr50_QCMJC. Leaves of 4-5 weeks old *N. benthamiana* plants were infiltrated with *A. tumefaciens* carrying NLRs and effectors at $OD_{600}=0.3$ and 0.4 , respectively. UV images were taken on a BioRad Gel Imager 3-5 days after infiltration. Numbers on images correspond to the numbers of leaves that showed the depicted phenotype. **C.** The Sr50 structure and insert to the NB-ARC latch. Structure was generated using AlphaFold. NB-ARC, LRR and CC domains are shown in green, red and orange, respectively. Blue box enlarges contact point between NB-ARC and LRR domains and the involved amino acids; we refer to this interaction as NB-ARC LRR latch. Some of the residues in this NB-ARC LRR latch are depicted. If a homologous position of the residue is substituted in Sr33, the residue is indicated as 'variable'. Otherwise, it is indicated as 'invariable'. **D.** The signature of inter-residual co-evolution captured by Gremlin. The filtered multiple sequence alignment (MSA) collected during structure prediction with AlphaFold for Sr50 was used to infer inter-residual co-evolution with Gremlin. The data provided in Table S5 is visualized. Only the pairs with probability ≥ 0.40 were indicated. If inter-residual co-evolution was detected, a colored dot is indicated in the corresponding position of the pair. The x- and y-axis indicate amino acid position in Sr50. Multi-domain architecture (Coiled coil (CC), NB-ARC, and leucine-rich repeats (LRRs)) is provided around and across the MLA. The red dashed boxes indicate where co-evolution between the NB-ARC domain and terminal LRRs would appear. **E.** The sequence feature of Sr50's NB-ARC domain. The normalized entropy was calculated with all MLA members used in the study. Major and minor motifs within the NB-ARC are shown in relation to normalized entropy values. Gray arrow indicates amino acid involved in NB-ARC/LRR latch interaction.

Supplement

Sr50

Sr33

MLA

HcMLA

Figure S1. The distribution of the normalized Shannon entropy across the leucine-rich repeat (LRR) domain of the four phylogenetic groups from Fig. 1A. The normalized Shannon entropy across LRRs 1-16 for each phylogenetic group is depicted as a heatmap to provide a closer view of sequence variations. Each LRR unit and each position within the LRR unit was defined according to the general motif LXXLXLXX(C/N).

A

B

Figure S2. Schematic representation of swapped the region in the chimeric NLR Sr33/50. **A.** The chimeric protein Sr33/50 was created by replacing amino acids in the region of LRRs 8-14. The NLR consists of Sr33 ranging from AA 1-702 and 887-961. In place the region 703-886 are the corresponding sites in Sr50. **B.** Normalized Shannon entropy of NLRs in the Sr50 clade. The asterisks indicate amino acids that were utilized in the generation of Sr33_{syn}. Red asterisks indicate positions in which either Sr33 or Sr50 has show high entropy peaks determined by the cut-off of 0.347.; The blue asterisk is the position in which both Sr33 and Sr50 had low entropy values and corresponds to the low entropy peak (M844) that governs autoactivity within Sr50.

D

Position (Sr33)	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744
Sr33	Q	E	L	G	K	L	A	R	L	R	E	L	K	I	C	F	K	D	G	S	L	D	L	Y	E	G	F	V	K	S	L	C	N	L	H	H	I	E	S	L	I	V	S	C	N	S	G	E	T	S
Sr33syn	Q	E	L	G	K	L	A	R	L	R	E	L	K	I	C	F	K	D	G	S	L	D	L	Y	E	G	F	V	K	S	L	C	N	L	H	H	I	E	S	L	S	V	S	Y	N	S	G	E	T	S
Sr50	Q	E	L	G	N	L	T	R	L	R	E	L	K	I	C	F	K	D	G	N	L	D	S	Y	K	L	F	V	K	S	L	G	N	L	H	H	I	E	S	L	S	I	S	Y	N	S	K	E	T	S

Position (Sr50)	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748
Sr50	Q	E	L	G	N	L	T	R	L	R	E	L	K	I	C	F	K	D	G	N	L	D	S	Y	K	L	F	V	K	S	L	G	N	L	H	H	I	E	S	L	S	I	S	Y	N	S	K	E	T	S

Position (Sr33)	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794
Sr33	F	E	L	M	D	L	L	G	E	Q	W	V	P	P	V	H	L	R	E	F	V	S	E	M	P	S	Q	L	S	A	L	R	G	W	I	K	R	D	P	S	H	L	S	N	L	S	E	L	I	L
Sr33syn	F	E	L	M	D	L	L	G	E	Q	W	V	P	P	V	H	L	R	E	F	V	S	W	M	P	S	Q	L	S	A	L	R	G	W	I	K	R	D	P	S	H	L	S	N	L	S	E	L	I	L
Sr50	F	E	L	M	D	L	L	G	E	R	W	V	P	P	V	H	L	R	E	F	V	S	W	M	P	S	Q	L	S	A	L	R	G	W	I	K	R	D	P	S	H	L	S	N	L	S	E	L	I	L

Position (Sr50)	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798
Sr50	Q	E	L	G	N	L	T	R	L	R	E	L	K	I	C	F	K	D	G	N	L	D	S	Y	K	L	F	V	K	S	L	G	N	L	H	H	I	E	S	L	S	I	S	Y	N	S	K	E	T	S

Position (Sr33)	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844
Sr33	P	T	V	K	E	V	Q	Q	E	D	V	E	I	I	G	G	L	L	S	L	R	R	L	L	I	E	S	T	H	Q	T	Q	R	L	L	V	I	R	A	D	G	F	R	C	M	V	D	F	Y	L
Sr33syn	W	P	V	K	E	V	Q	Q	E	D	V	E	I	I	G	G	L	L	S	L	R	R	L	W	I	K	S	T	H	Q	T	Q	R	L	L	V	I	R	A	D	G	F	R	C	M	V	D	F	E	L
Sr50	W	P	V	K	E	V	Q	Q	E	D	V	E	I	I	G	G	L	L	S	L	R	R	L	W	I	K	S	T	H	Q	T	Q	R	L	L	V	I	R	A	D	G	F	R	C	M	V	D	F	E	L

Position (Sr50)	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848
Sr50	Q	E	L	G	N	L	T	R	L	R	E	L	K	I	C	F	K	D	G	N	L	D	S	Y	K	L	F	V	K	S	L	G	N	L	H	H	I	E	S	L	S	I	S	Y	N	S	K	E	T	S

Position (Sr33)	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894
Sr33	N	C	G	S	A	T	Q	I	M	F	E	S	G	A	L	P	R	A	E	E	V	C	F	S	L	G	V	R	V	A	K	E	D	G	N	R	G	F	D	L	G	L	Q	G	N	L	L	S	L	R
Sr33syn	N	C	G	S	A	T	Q	I	M	F	E	S	G	A	L	P	R	A	E	E	V	V	F	S	L	G	V	R	V	A	K	E	D	G	N	R	G	F	D	L	G	L	Q	G	N	L	L	S	L	R
Sr50	N	C	G	S	A	T	Q	I	M	F	E	S	G	A	L	P	R	A	E	E	V	V	F	S	L	G	V	R	V	A	K	E	D	G	N	R	G	F	D	L	G	L	Q	G	N	L	L	S	L	R

Position (Sr50)	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898
Sr50	Q	E	L	G	N	L	T	R	L	R	E	L	K	I	C	F	K	D	G	N	L	D	S	Y	K	L	F	V	K	S	L	G	N	L	H	H	I	E	S	L	S	I	S	Y	N	S	K	E	T	S

Figure S3. Pictures of *N. benthamiana* leaves of infiltrations of Sr33_{syn} and controls used to generate Figure 2C. Leaves of 4-5 weeks old *N. benthamiana* plants were infiltrated with *A. tumefaciens* carrying NLRs and effectors at OD₆₀₀=0.3 and 0.4, respectively. UV images were taken on a BioRad Gel Imager 3-5 days after infiltration. **A.** Sr50, Sr33/50 and Sr33_{syn} induce cell death in response to AvrSr50 in *N. benthamiana*. **B.** None of the NLRs initiated cell death in response to the effector variant AvrSr50_QCMJC. **C.** Over-expression of Sr33 and Sr50 with *p35S* leads to cell death in absence of effector in *N. benthamiana*.

Figure S4. Pictures of *N. benthamiana* leaves infiltrated with Sr50 single and double mutants, and controls that were used to generate Figure 3B. Leaves of 4-5 weeks old *N. benthamiana* plants were infiltrated with *A. tumefaciens* carrying NLRs and effectors at OD₆₀₀=0.3 and 0.4, respectively. UV images were taken on a BioRad Gel Imager 3-5 days after infiltration. **A.** Left panel. Sr50 wild type, single and double mutants induce cell death when co-infiltrated with AvrSr50. **B.** Right panel. Sr50 wild type, single and double mutants do not induce cell death when co-infiltrated with the effector variant AvrSr50_QCMJC.

Swapped Low entropy High entropy

Figure S5. Variable LRR residues mapped into the Sr50 structure.

The residues in the leucine-rich repeat (LRR) units 8-13 of Sr50 were compared to the LRR of Sr33. The sites that display polymorphism were mapped to the predicted protein structure of Sr50. The residues that display low and high entropy based on the entropy calculation of the Sr50 group were indicated in ivory and pink. The 12 residues used to synthesize Sr33_{syn} were colored in red and indicated as “swapped”.

Figure S6. The inter-residual distances of Sr50.

The pairwise distances between alpha carbon atoms in the predicted structure of Sr50 are depicted. The minimum observed distance was around 3Å. Only the pairs with the pairwise distance < 7.5Å were indicated. The x- and y-axis indicate amino acid position in Sr50. Multi-domain architecture (Coiled coil (CC), NB-ARC, and leucine-rich repeats (LRRs)) is provided around and across the axes. If a pair of residues are in proximity in the protein structure, a dot is indicated in their corresponding position. The insert captures pairwise distances between the residues in the NB-ARC domain and terminal LRRs.