A DANGEROUS MIX

From Natural Variation to Genetic Incompatibilities in *Arabidopsis thaliana* and *Arabidopsis arenosa*

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard-Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Ana Cristina Barragán López

aus Guadalajara, Mexiko

Tübingen

2020

Gedruckt mit Genehmigung der der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard-Karls Universität Tübingen

Tag der mündlichen Qualifikation: 27. November 2020

Stellvertretender Dekan: Prof. Dr. József Fortágh

1. Berichterstatter: Prof. Dr. Detlef Weigel
2. Berichterstatter: Prof. Dr. Marja Timmermans

Acknowledgments

First and foremost I would like to thank Prof. Detlef Weigel and Prof. Eunyoung Chae for their constant support and guidance during my time as a PhD student. Thanks to your positive influence I could advance both scientifically and personally in these past years. Additionally, I would like to thank Prof. Marja Timmermans and Dr. Adrian Streit for forming part of my Thesis Advisory Committee and for their valuable suggestions during our meetings. Also thanks to Georg Felix and Farid El Kasmi for agreeing to take part in my PhD defense.

Thanks to all past and present members of the Weigel lab I had the pleasure of meeting for making the lab such an enjoyable place to work in. I would also like to thank my friends both inside and outside the lab for making the years spent in Tübingen a time I will always remember very fondly. Last, but certainly not least, I would like to thank my family.

Table of Contents

Zusammenfassung

Pflanzen und Krankheitserreger haben sich seit Jahrtausenden gemeinsam entwickelt. Als Teil dieser langfristigen Interaktion ist sowohl die Erhaltung der seit langem bestehenden genetischen Variation als auch die Erzeugung von neuem genetischen Material sowohl von der Pflanzen- als auch von der Erregerseite erforderlich, um im Wettbewerb miteinander konkurrenzfähig zu bleiben. Infolgedessen sind einige Bestandteile des pflanzlichen Immunsystems stark diversifiziert. Dies ist beispielsweise der Fall bei pflanzlichen NLRs, die als intrazelluläre Rezeptoren fungieren, welche eingehende Pathogen-Effektoren erkennen, dadurch eine Signalkaskade in Gang setzen und schließlich den Zelltod auslösen. Die große Variabilität der NLRs ermöglicht die Erkennung eines breiten Spektrums von Pathogen-Effektoren. Manchmal kann diese Variabilität jedoch zum Nachteil werden: Wenn zwei divergierende Elemente des pflanzlichen Immunsystems, oft zwei NLRs oder eine NLR und eine andere Komponente des Immunsystems, in einer Hybridpflanze - der Nachkommenschaft einer Kreuzung zwischen zwei verschiedenen Akzessionen - aufeinander treffen, können sie trotz Abwesenheit eines Pathogens eine Immunantwort auslösen. Dieses Phänomen wird als Hybridunverträglichkeit bezeichnet.

In der hier vorliegenden Arbeit untersuche ich zwei Gruppen von Hybridunverträglichkeitsfällen in *Arabidopsis thaliana* und einen Fall von Inzuchtdepression bei ihrer auskreuzenden verwandten *Arabidopsis arenosa.* Im ersten Projekt untersuche ich eine Reihe von Fällen, die das Ergebnis von inkompatiblen Interaktionen zwischen dem NLR-Cluster *RPP7*, das eine stammspezifische Resistenz gegen den Falschen Mehltau verleiht, und *RPW8*, einem atypischen Nicht-NLR Resistenz Gencluster, das eine Breitspektrumresistenz gegen filamentöse Pathogene verleiht, sind. Ich beschreibe drei unabhängige Fälle, in denen allelspezifische Interaktionen zwischen diesen beiden Loci zu inkompatiblen Hybriden führen. Darüber hinaus identifiziere ich in zwei dieser Fälle die kausalen Gene für die Inkompatibilität von der *RPW8* Seite: *RPW8.1* und *HR4*. Die resultierenden Proteine dieser beiden, für die Inkompatibilität kausalen Gene, zeigen Längenpolymorphismen in verschiedene Akzessionen, die durch 21- oder 14- Aminosäuren-Wiederholungszahlvariationen in ihrem C-Terminus gekennzeichnet sind. Ich zeige, dass diese C-terminalen Repeats den Schweregrad des Hybridphänotyps weitgehend modulieren und dass nur die Akzessionen, die lange RPW8.1- und kurze HR4 Proteinvarianten tragen, inkompatibel sind, wenn sie mit bestimmten RPP7-Proteinen kombiniert werden.

Im zweiten Projekt untersuche ich eine Reihe von Hybrid-Inkompatibilitätsfällen, bei denen der Hybrid stark nekrotisch, sich nicht sich nicht über das Keimblattstadium hinaus entwickelt und früh abstirbt. Ich zeige, dass im Hybrid massive transkriptionelle Veränderungen, einschließlich der Hochregulation der meisten NLR-Gene, stattfinden, die wahrscheinlich zu seinem stark nekrotischen Phänotyp beitragen. Weiterhin identifiziere ich die kausalen Loci für die Unverträglichkeit von *DM10* und *DM11.* Anschließend zeige, dass *DM10* ein Singleton NLR ist, das nach der Speziation von *A. thaliana* aus einem NLR-Gencluster verlagert wurde. Ich zeige auf, dass das Risiko *DM10* Allel ein vorzeitiges Stoppcodon trägt, und das es obwohl es in der globalen *A. thaliana* Population häufig und geographisch weit verbreitet ist, nicht gleichzeitig mit dem Risiko *DM11* Allele vorkommt.

Im dritten Projekt untersuche ich auf das Vorhandensein potentieller Hybrid-Inkompatibilitätsfalle, die in natürlichen *A. arenosa* Populationen auftreten, und zeige, dass schädlicher Phänotypen zwar häufig vorkommt, aber zumindest in einigen Fällen wahrscheinlich das Produkt einer Inzuchtdepression ist.

Zusammenfassend stellt meine hier vorliegende Arbeit einen Plan für die Identifizierung eines Hybrid-Inkompatibilitätsfalls, die Kartierung und experimentelle Bestätigung der für die Inkompatibilität kausalen Loci sowie für die Feststellung der zugrundeliegenden genetischen und evolutionären Prozesse der Hybridinkompatibilität dar.

Summary

Plants and pathogens have co-evolved for millennia. As part of this long-term interaction, both the preservation of long-standing genetic variation, as well as the generation of novel genetic material is required from both the plant and the pathogen side to remain competitive when facing each other. As a consequence, some members of the plant immune system are highly diversified. Such is the case for plant NLRs, which act as intracellular receptors that recognize incoming pathogen effectors, thereby initiating a signalling cascade and ultimately resulting in cell death. The extensive variability of NLRs enables the recognition of a wide spectrum of pathogen effectors. However, sometimes this variability can backfire: When two divergent elements of the plant immune system, often two NLRs, or one NLR and another immune system component, meet in a hybrid plant – the progeny of a cross between two different accessions – they can trigger an immune response in the absence of a pathogen. This phenomenon is called hybrid incompatibility.

Here, I study two sets of hybrid incompatibility cases in *Arabidopsis thaliana* and a case of inbreeding depression in its outcrossing relative *Arabidopsis arenosa*. In the first project, I study a set of *A. thaliana* incompatibility cases which are are the result of incompatible interactions between the NLR cluster *RPP7*, which confers strain-specific resistance to downy mildew, and *RPW8*, an atypical non-NLR resistance (*R*) gene cluster that confers broad-spectrum resistance to filamentous pathogens. I describe three independent cases where allele-specific interactions between these two loci result in incompatible hybrids. In addition, for two of these cases, I identify the causal genes for incompatibility from the *RPW8* side: *RPW8.1* and *HR4.* The resulting proteins of these two causal genes for incompatibility show length polymorphisms across different accessions which are characterized by 21- or 14- amino acid repeat number variations in their C terminal. I show that these C terminal repeats largely modulate the severity of the hybrid phenotype, and that only accessions carrying long RPW8.1 and short HR4 protein variants are incompatible when combined with particular RPP7 proteins.

In the second project, I study a set of *A. thaliana* hybrid incompatibility cases where the hybrid is severely necrotic, does not develop past the cotyledon stage, and dies early on. I show that massive transcriptional changes take place in the hybrid, including the upregulation of most NLR genes, which likely contribute to its severely necrotic phenotype. I then identify the causal loci for incompatibility, *DM10* and *DM11*, and show that *DM10* is a singleton NLR that was relocated from an NLR gene cluster after *A. thaliana* speciation. I establish that the risk *DM10*

allele carries a premature stop codon, and although common and geographically widespread in the global *A. thaliana* population, co-occurrence with the risk *DM11* allele is absent.

In the third project, I screened for the presence of potential hybrid incompatibility cases occurring in natural *A. arenosa* populations, and show that heritable deleterious phenotypes are common, but, at least in some cases, likely the result of inbreeding depression.

In short, my work presents a roadmap starting from identifying potential hybrid incompatibility cases to mapping and experimentally confirming the underlying causal loci, to establishing the underlying genetic and evolutionary processes building up to these incompatibilities.

Publications

Accepted papers

Barragan AC, Wu R, Kim S-T, Xi W, Habring A, Hagmann J, Van de Weyer AL, Zaidem M, Ho WWH, Wang G, Bezrukov I, Weigel D, Chae E (2019) RPW8/HR repeats control NLR activation in *Arabidopsis thaliana*. *PLoS Genetics* 15:e1008313. https://doi.org/10.1371/journal.pgen.1008313

Barragan AC, Collenberg M, Wang J, Lee RRQ, Cher WT, Rabanal FA, Ashkenazy H, Weigel D, Chae E, (2020) A truncated singleton NLR causes hybrid necrosis in *Arabidopsis thaliana. Molecular Biology and Evolution* Accepted for publication, https://doi.org/10.1093/molbev/msaa245

Barragan AC & Weigel D (2020) Plant NLR Diversity: The Known Unknowns of pan-NLRomes. *The Plant Cell,* Accepted for publication

Advanced manuscripts

Barragan AC, Collenberg M, Kerstens M, Schwab R, Bemm F, Bezrukov I, Požárová D, Kolář F, Weigel D. A case of inbreeding depression in a natural *Arabidopsis arenosa* population.

Additional papers not included in this dissertation

Exposito-Alonso, M., Gómez Rodríguez, R., **Barragan AC** (...) Burbano HA, Bossdorf O, Nielsen R, Weigel D. (2019) Natural selection on the *Arabidopsis thaliana* genome in present and future climates. *Nature* 573, 126–129. https://doi.org/10.1038/s41586-019-1520-9

Introduction

Two (Interconnected) Branches of the Plant Immune Response

Plants and their microbes play a key role in shaping each other's genomes. In natural ecosystems, most plants are resistant to most pathogens, and successful pathogen infections are usually the exception rather than the rule. In plants, the immune response is encoded in the germline, so each cell has the capacity to initiate an immune response on its own. This innate immunity is based on detecting foreign molecules through extra- and intracellular receptors, triggering an immune response after recognition (Jones and Dangl 2006). When a pathogen docks in a plant cell it attempts to colonize, it emanates conserved molecules, termed microbe- or pathogen-associated molecular patterns (MAMPs/PAMPs), which stamp it as a potential threat (Jones and Dangl 2006) (**Fig 1**). These PAMPs are detected by the plant's pattern recognition receptors (PRRs), which are localized in the plant membrane and survey extracellular space. PAMP recognition by these PRR receptors leads to several changes in the cell, including increased calcium influx, production of reactive oxygen species, activation of MAPK cascades, as well as the induction of a wide set of defense genes (Boller and Felix 2009). Collectively, this response is termed PAMP or Pattern Triggered Immunity (PTI) (Jones and Dangl 2006). The second layer of defense after a pathogen infection, takes place in the plant's intracellular space. If a pathogen manages to get through the plant cell wall, it releases small non-conserved molecules termed effectors into the cell, which promote the pathogen's colonization of the plant by evading or suppressing PTI (Jones and Dangl 2006) (**Fig 1**). If the plant detects these effectors through its cytoplasmic immune receptors, called Nucleotide-binding leucine-rich repeat proteins (NLRs), these become activated and lead to effector-triggered immunity (ETI) (Chisholm et al. 2006), which leads to similar cellular changes than those seen in PTI. PTI and ETI, the two branches of the plant immune system, have been thought to act largely independent from another, however, recent studies show evidence of interdependence between PTI and ETI components, which enhances the robustness of the defense response (Hatsugai et al. 2017; Ngou et al. 2020; Yuan et al. 2020).

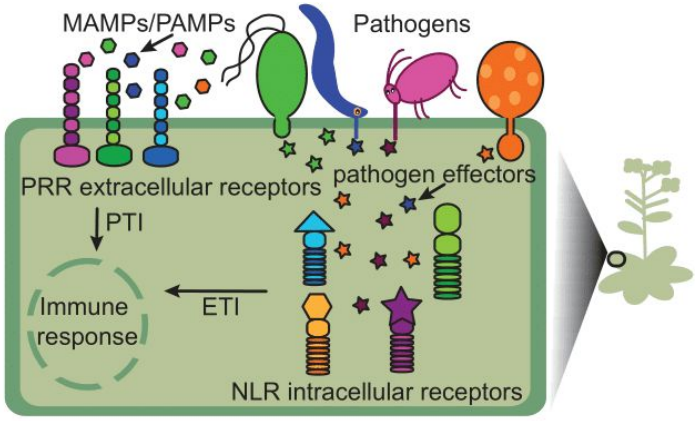


Fig 1. Cartoon of a plant immune response. When a pathogen attempts to infect a plant cell, its conserved MAMPs/PAMPs are recognized by plant extracellular receptors (PRRs), initiating pattern-triggered immunity (PTI). In the meantime, pathogens that manage to get past the plant cell wall release effector molecules into the plant cell cytoplasm to promote its colonization. Pathogen effectors are recognized by plant intracellular Nucleotide-binding leucine rich repeat (NLR) receptors, and an effector triggered immune response (ETI) is initiated. Cartoon is based on (Jeffery L. Dangl, Horvath, and Staskawicz 2013).

Brief History of Plant Resistance Genes

Plant disease resistance has been largely achieved through breeding of disease resistant varieties. Flor's gene-for-gene hypothesis was the first in describing how plant disease resistance is achieved. He stated that for each gene that confers resistance in the host (*R* gene), there is a corresponding gene in the parasite that confers pathogenicity (Flor 1971). Such pathogen genes, which are typically found only in some races of a pathogen species, were originally called avirulence (Avr) genes because their detection by the host makes the pathogen avirulent. The first *R* gene was cloned over 25 years ago: *Hm1* in maize, which encodes a toxin reductase (Johal and Briggs 1992) (**Fig 2**). Shortly afterwards, the first two NLR-effector pairs were cloned: *RPS2* in *A. thaliana* and its cognate effector from the bacteria *Pseudomonas syringa*e AvrRpt2 (Mindrinos et al. 1994; Bent et al. 1994) and *N* from tobacco, which confers resistance against the tobacco mosaic virus (Whitham et al. 1994) (**Fig 2**). Since then, over 300 *R* genes have been cloned, most of which encode for NLR intracellular receptor proteins, followed by RLP and RLKs which act as PRR extracellular receptors (Kourelis and van der Hoorn 2018) (**Fig 2**).

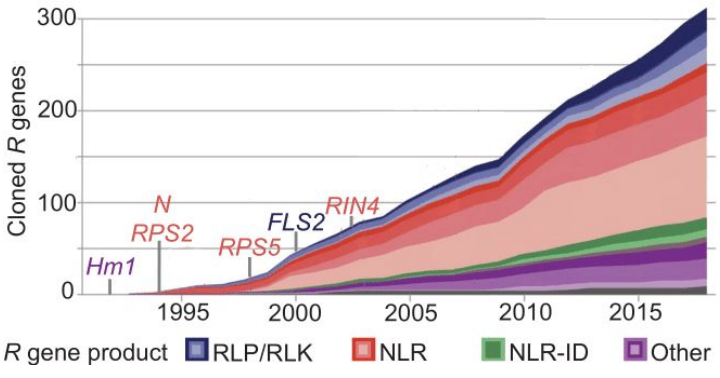


Fig 2. Timeline summarizing the number of *R* genes cloned over time. In the past 25 years there has been a large increase of cloned I genes, especially of those that are encoded by NLR genes, followed by PRRs such as RLP and RLK. The first cloned R gene was *Hm1* a toxin reductase in 1992 (Johal and Briggs 1992), followed by the *N* locus from

tobacco and *RPS2* from *A. thaliana* (Mindrinos et al. 1994; Bent et al. 1994; Whitham et al. 1994). Adapted from (Kourelis and van der Hoorn 2018).

NLRs Genes and NLR Complexes are (Almost) Everywhere

The innate immune system in plants and animals share multiple common features, such as non-self recognition through intra- and extracellular receptors, the induction of localized cell death, as well as systemics signaling (Urbach and Ausubel 2017). NLR genes are found in several eukaryotic kingdoms, including plants, animals, fungi and protists, with similar architectures constituting a finite number of domain combinations; these similarities however, are thought to be the product of convergent evolution (Yue et al. 2012).

Generally, the basic NLR unit follows a modular tri-domain structure where domains work together and become activated after recognizing pathogen effector (A. Bentham et al. 2017). The N terminal domain of NLR proteins can be considered the business end, being the primary structural element of signal transduction (A. Bentham et al. 2017). In animals, examples of common N terminal domains are caspase recruitment domains (CARD), death effector domain (DED), baculovirus inhibitor of apoptosis protein repeat (BIR) domain and pyrin domain (PYD), while in plants, the N-terminal domain is usually either a coiled-coil (CC), a Toll/interleukin-1 receptor/Resistance protein (TIR) or a coiled-coil domain that is reminiscent to the RESISTANCE TO POWDERY MILDEW 8 (CC-R or RPW8) domain (A. Bentham et al. 2017) (**Fig 3**). In fungi, there is high variability in N terminal domains, more so than in animals or plants, and only half of these domains are characterized (Dyrka et al. 2014; Uehling, Deveau, and Paoletti 2017). Known domains include HeLo, HET, which are thought to be related to CC and TIR domains in plants, respectively (Dyrka et al. 2014; Uehling, Deveau, and Paoletti 2017) (**Fig 3**). The central and most conserved NLR domain is a signal-transducing ATPase (STAND) domain, which likely evolved from a common bacterial ancestor and then gave rise to its subclades, NB-ARC and NACHT (Urbach and Ausubel 2017). This domain serves in nucleotide binding (NB) and as a ADP-ATP switch which keeps the NLR in an ON/OFF state. The NB-NACHT subclade is typical for animals and fungi, while NB-ARC is typical for plants. Mutations in the central NB carrying domain of NLR genes can lead to an inflammatory response in animals and autoimmunity in plants, whereas loss of function results in increased plant disease susceptibility (A. Bentham et al. 2017). Lastly, the C terminal domain acts as a ligand binding platform, as well as having an autoinhibitory function; this domain is usually formed of repeats. In plants and some animals, LRR domains are present in the C terminus,

while fungal NLR proteins do not have LRR repeats (Soanes and Talbot 2010; Dyrka et al. 2014) (**Fig 3**).

Plant NLRs can recognize pathogen effectors either by directly binding to them, or by recognizing effector-mediated modifications of another host component. In the latter scenario, the NLR acts as a guard and its host client is its guardee (Cesari 2018). Guardees are typically proteins that are targeted by effectors, with some having lost their original function in plant growth or immunity, and only acting as effector decoys (J. L. Dangl and Jones 2001; van der Hoorn and Kamoun 2008).

Many of the inferences about plant, animal and fungi NLRs came from the study of individual NLR domains and interactions between them and NLR client proteins, e.g. guardees. The similarity in overall function between animal and plant NLRs was subsequently confirmed by structural analyses. In vertebrates, NLR proteins form wheel-shaped oligomers called inflammasomes, which are assembled upon pathogen recognition and activate a signalling cascade that leads to the formation of pores in the plasma membrane, resulting in localized cell death (Broz and Dixit 2016) (**Fig 3**). In plants, a strikingly similar structure to inflammasomes was found in the first identified NLR resistosome (J. Wang et al. 2019; J. Z. Wang et al. 2019). Here, upon the release of the AvrAC effector from *Xanthomonas campestris* pv. *campestris* into the plant cell, it uridylylates PBL2, a decoy kinase, this change is detected by HOPZ-ACTIVATED RESISTANCE 1 (ZAR1), a coiled-coil NLR (CNL) thats acts as its guard (G. Wang et al. 2015). Without the AvrAC effector, ZAR1 and its bound pseudokinase RKS1 are in a heterodimeric complex which is in an OFF state, where ADP is bound to the NB-ARC domain of ZAR1, and RKS1 interacts only with ZAR1's LRR domain. When the ZAR1/RKS1 complex recognizes the uridylylation of PBL2, RKS1 then binds to PBL2, inducing conformational changes; the ADP found in ZAR1's NB-ARC domain is phosphorylated to ATP and activates the ZAR1/RKS1/PBL2 complex (J. Z. Wang et al. 2019). This activated ZAR1/RKS1/PBL2 complex oligomerizes, forming a pentameric wheel-shaped structure that localizes in the plasma membrane (**Fig 3**). Since the CC domain of ZAR1 protrudes from the rest of the complex, it is tempting to speculate that this domain directly inserts itself into the membrane, creating a pore and potentially changing ion flow and triggering cell death (J. Wang et al. 2019). Recently, the N-terminal MADA motif was discovered and found to be present in 20% of all CNLs, this motif corresponds to the α1 helix in ZAR1, the protruding part of the resistosome, indicating that resistome formation and possible membrane insertion might be common among plant CNLs (Adachi et al. 2019).

TIR domains in both plants and animals have been found to cleave NAD+, triggering an immune signalling cascade resulting in cell death, and often requiring oligomerization (Williams et al. 2014; Essuman et al. 2017; Wan et al. 2019; Horsefield et al. 2019). Recently a “plant inflammasome” was generated by fusing the TIR domain from RPS4 and the N terminus of the mammalian NLRC4 (Duxbury et al. 2020). This showed that proximity of plant TIR signalling domains enabled by NLRC4 oligomerization can be enough to activate plant immune signalling. Shortly afterwards, the structure of the first TNL resistosome was presented, where Roq1 from *N. benthamiana* forms a tetrameric clover-shaped complex (F. Liu et al. 2020). This structure resembles that of the ZAR1 resistosome, making the idea that NLRs forming resistosomes that can directly make holes in plasma membranes and thereby initiate further DAMP signaling and/or cell death is common, attractive.

Although animal inflammasomes have not been shown to insert themselves into the plasma membrane, and instead trigger pore formation indirectly, MLKL, an apoptosis protein, has, and MLKL has been homology modelled to the N-terminal RPW8 domain found in plants, and so have the HELo and HELL cell death-inducing domains in fungi (Seuring et al. 2012; H. Wang et al. 2014; Daskalov et al. 2016; A. R. Bentham et al. 2018) (**Fig 3**). This suggests that at least some animal NLR domains also very likely have the potential to insert themselves into the plasma membrane.

If NLRs in plants, animals and fungi are the product of convergent evolution, what is the advantage in having this particular multi-domain structure and mode of action? Immune receptors should ideally act as hair triggers, such that any threat is immediately met, but at the same time, they should also be robust to inadvertent activation, since inappropriate immunity can have devastating consequences (Bomblies et al. 2007). The multi-domain structure allows for self-inhibition through intramolecular interactions, providing a primary safeguard against spurious activation. The formation of higher-order complexes may serve to amplify the triggering signal, but also help to prevent mis-regulation. Lastly, the modular architecture may allow for facile reshuffling of individual domains, endowing them with versatility in recognition specificity, as well as allowing for different selection pressures to act on individual domains.

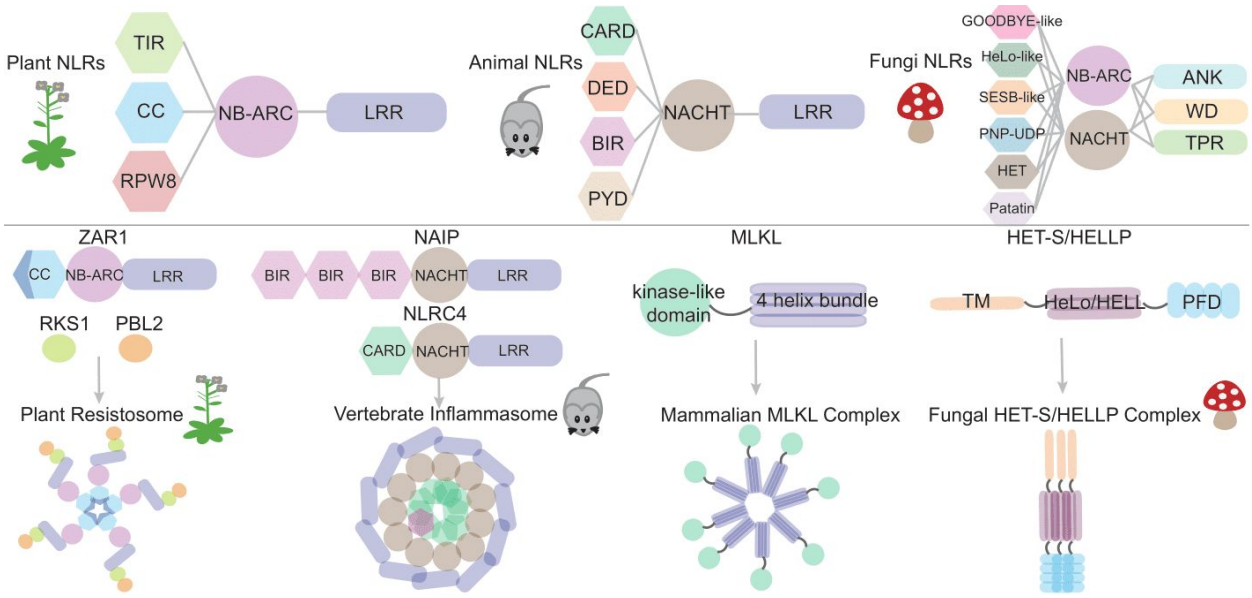


Fig 3. Typical tripartite domain organization in plant, animal and fungi NLRs (top). Similar cell-death inducing complexes are formed in the plant ZAR1 resistosome, the NAIP/NLRC4 vertebrate inflammasome, the MLKL complex from mammals, as well as in HET-S and HELLP in fungi (bottom).

NLR Gene Numbers Vary Greatly Between Plant Species

After colonization of land, plant genomes experienced a massive expansion of NLR genes, going from fewer than a dozen in green algae, where plant NLRs are thought to originate from, to many hundreds in land plants, likely as a consequence of adaptation to new pathogen pressures (Shao et al. 2019) (**Fig 4**). NLR genes appear to turn over rapidly, with frequent births and deaths (Michelmore and Meyers 1998).

NLR numbers across different species are highly variable; among all coding genes, the percentage of NLR genes ranges over many orders of magnitude, from 0.003% in bladderwort (*U. gibba*) to 2% in apple (*M. domestica*) (Jia et al. 2015; Baggs et al. 2020). In addition, many of the NLR expansion and contraction events are lineage-specific; TNLs, although present in some mosses and green algae, are largely absent from monocots for example (Tarr and Alexander 2009; Gao et al. 2018), and in the Solanaceae, CNLs are greatly expanded (Seo et al. 2016), while Rosales and conifers have experienced independent expansions of RNLs (Jia et al. 2015; Van Ghelder et al. 2019) (**Fig 4**).

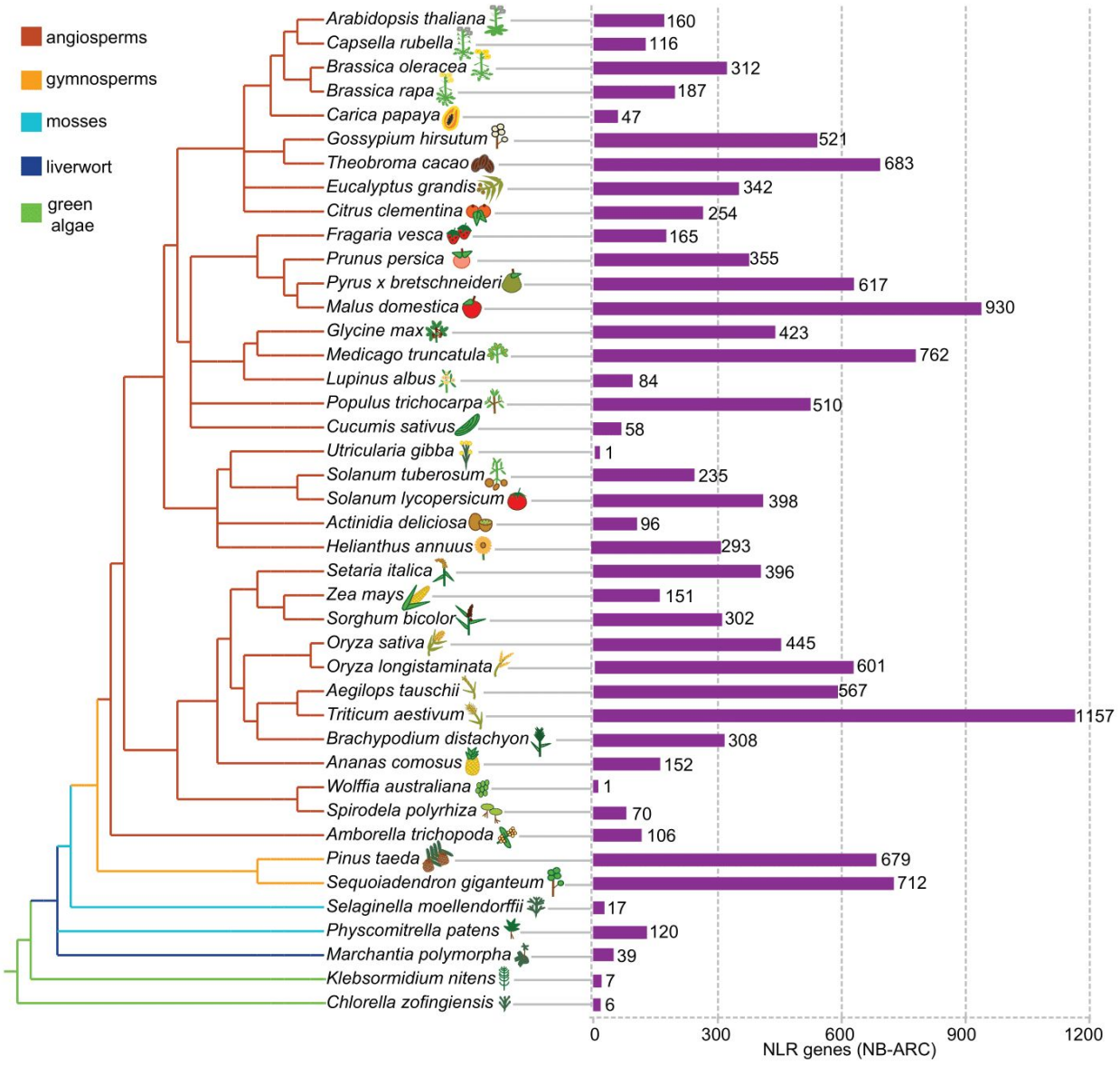


Fig 4. Plant phylogeny and NLR complement sizes. Adapted from (Gao et al. 2018; Munch et al. 2018; Baggs et al. 2020).

NLR numbers are generally low in the Cucurbitaceae, likely as a result of frequent gene losses and few subsequent duplication events (Lin et al. 2013) (**Fig 4**). Low NLR numbers may result from their functional dispensability; for example, *Wolffia australiana,* a duckweed with just over 15,000 genes that potentially represent a minimum set of genes necessary for survival in an aquatic environment, has only one canonical NLR (Michael et al. 2020) (**Fig 4**). Bladderwort, another aquatic plant, has at most one, and perhaps no NLR at all (Baggs et al. 2020). A question is whether there are evolutionary innovations in these plants that compensate for the

loss of NLRs. In support of such compensatory changes, another duckweed with a highly reduced NLR complement, *Spirodela polyrhiza*, appears to have more components for antimicrobial signalling than other plants (An et al. 2019).

Diversity in NLR Genomic Organization

In *A. thaliana*, about half of NLR genes are found as physical singletons, whereas the other half are found in clustered genomic arrangements (Blake C. Meyers et al. 2003; Van de Weyer et al. 2019). NLR clusters often appear to be the products of tandem duplication events, followed by unequal crossing over, as well as intra-cluster rearrangements and gene conversion events (B. C. Meyers et al. 1998; Noël et al. 1999; H. Kuang et al. 2004) (**Fig 5**). Clusters, which are often, but not always, made up of phylogenetically related NLR genes, can range in size from tens of kb, with *RPP5* in *A. thaliana,* which contains five cluster members in Col-0, being an example, to several Mb, with *RGC2* in lettuce being a record holder with ~3.5 Mb and consisting of 24 cluster members (B. C. Meyers et al. 1998; Noël et al. 1999).

What are the evolutionary advantages and disadvantages of having clusters? Recombination during meiosis is reduced by structural differences, which are particularly high in NLR clusters (Jiao and Schneeberger 2020). On the other hand, there is evidence of particularly high historical recombination rates around many NLR genes, as measured by linkage disequilibrium (LD) in natural populations (Horton et al. 2012; K. Choi et al. 2013), and there is only very weak evidence for NLR loci as a group to suppress recombination, even for loci that differ in arrangement between parents (Rowan et al. 2019). These apparently contradictory observations likely reflect simply the high variation in recombination rates across NLR loci, with some acting as recombination coldspots, as expected, but others acting actually as recombination hotspots (Kyuha Choi et al. 2016). This seems to be a function of the extent of structural variation between accessions, and many of the most structurally diverse regions of the *A. thaliana* genome indeed include NLR clusters with severely reduced recombination (Jiao and Schneeberger 2020). As I also discuss below, excessive NLR variation can potentially reduce fitness because of intragenomic immune system conflict, and under reduced recombination, it is more difficult to select for advantageous alleles that are linked to disadvantageous alleles, known as Hill-Robertson effect (Hill and Robertson 1966).

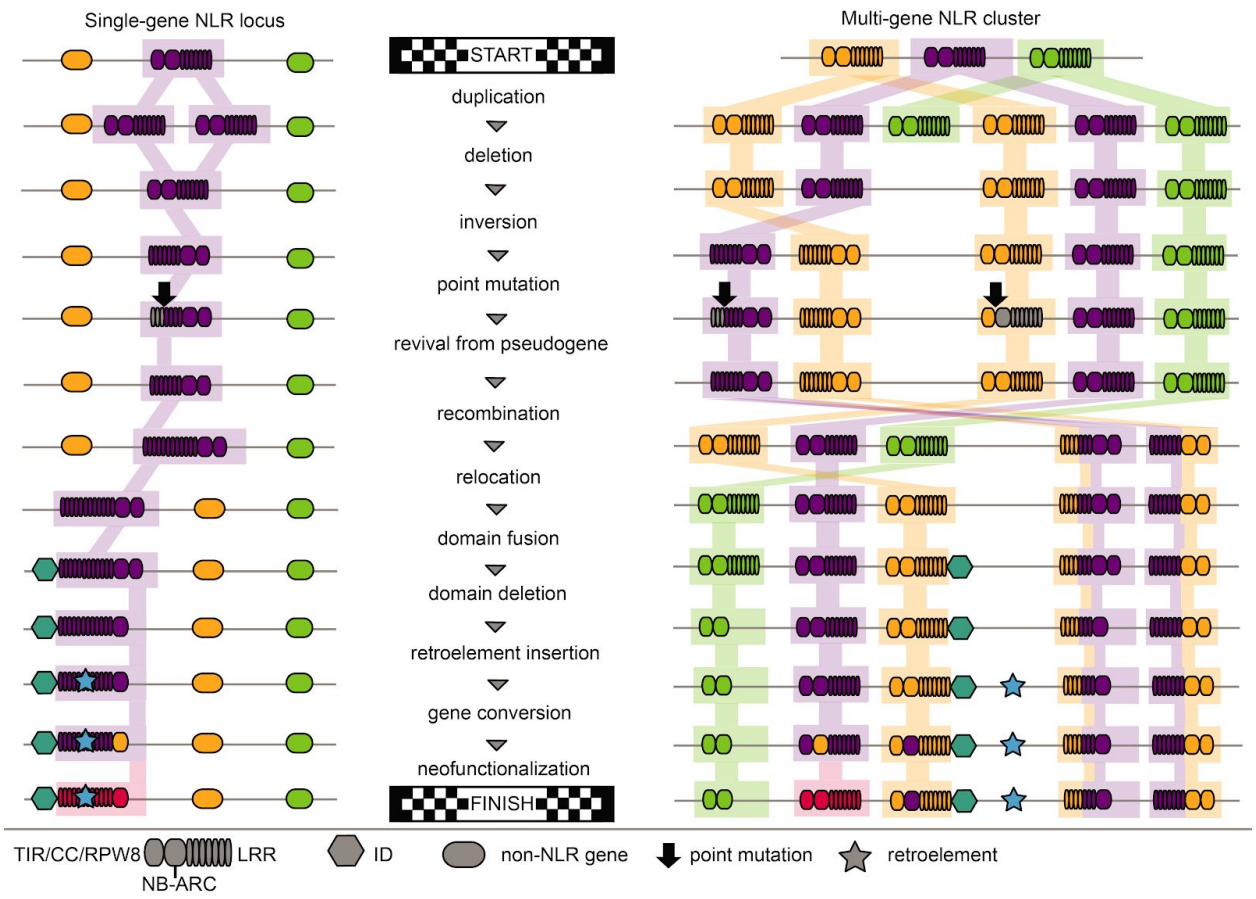


Fig 5. Examples of possible changes in sequence and genomic organization in single-gene and clustered NLRs through evolution. Major changes can occur in an NLR locus, and often the re-construction of the sequence of events will be impossible.

Perhaps a major advantage of clustered gene arrangements comes from several linked, closely related genes providing a means for generating new functional diversity through illegitimate recombination as well as gene conversion involving genes that are not strictly orthologs (H. Kuang et al. 2004; Wicker, Yahiaoui, and Keller 2007) (**Fig 5**). This is further facilitated by the repetitive structure of an important component of NLR exons, the LRR coding sequences.

Importantly, illegitimate recombination can both support expansion and contraction of sequences. The latter is perhaps particularly relevant when considering events at single-gene loci and clusters: At single-gene loci, any deletion will lead to truncation or loss of the gene (**Fig 5**). In contrast, in a cluster, illegitimate recombination between two genes can simultaneously reduce gene numbers and lead to creation of a new full-length gene (**Fig 5**). Although new genes resulting from illegitimate recombination will often be nonfunctional, they can serve as reservoirs for future evolution. In addition, they might combine the activities of the two original genes, or they could have a different activity all together (McDowell et al. 1998). Furthermore,

the more copies of a gene, the higher the chances are that beneficial mutations arise, both because multiple copies constitute a larger mutational target than a single copy, but also because duplicates can undergo relaxed selection due to their functional redundancy (Ohno 1970; Jiang and Assis 2017). Gene clusters thus provide a larger and more flexible genetic basis for evolving new resistance specificities through complete or partial domain swaps between closely related homologs (**Fig 5**).

Origin and Function of RPW8 Proteins

RPW8 proteins are found from early land plants to angiosperms, and like at NLR loci, species-specific expansion and contraction events are common (Yan Zhong and Cheng 2016). Through time, RPW8 proteins have been fused to a variety of other proteins, including NBS-LRRs, giving rise to RPW8-NLRs (RNLs) (Yan Zhong and Cheng 2016) (**Fig 6**). RNLs are a conserved phylogenetic group of so-called helper NLRs, which are genetically required by other NLRs in order to function (Collier, Hamel, and Moffett 2011). Examples of RNLs are *NRG1* from tobacco and *ADR1* from *A. thaliana* (Bonardi et al. 2011; Collier, Hamel, and Moffett 2011; Wu et al. 2017).

Stand-alone RPW8 proteins can act as atypical resistance proteins. Different haplotypes have been described in the *RPW8* locus in *A. thaliana*. One of the main differences between haplotypes is that, while all accessions carry the *HR1-3* genes, some accessions carry *RPW8.1* and *RPW8.2* genes, and others carry *HR4* instead (S. Xiao et al. 2001) (**Fig 6**). Initially, RPW8.1 and RPW8.2 proteins were identified as conferring resistance to powdery mildew isolates (S. Xiao et al. 2001), later on, these were also shown to confer resistance to downy mildew (W. Wang et al. 2007; Ma et al. 2014). Upon pathogen infection, RPW8.2 is targeted to the extrahaustorial membrane, where it activates haustorium-targeted defenses. RPW8.1 on the other hand, accumulates in the mesophyll cells below infected epidermal cells, where RPW8.2 is found (Ma et al. 2014). *HR1-HR3* have been shown to contribute to basal resistance against powdery mildew (Berkey et al. 2017). In the case of *HR4*, while its expression has been shown to increase after interactions with a variety of bacteria and fungi (Sáenz-Mata and Jiménez-Bremont 2012), accessions carrying *HR4* and not *RPW8.1* and *RPW8.2* are typically susceptible to powdery mildew (S. Xiao et al. 2001). Recently, some *HR4* variants have been shown to trigger cell death by disrupting the integrity of the plasma membrane, as well as having the ability to induce oligomerization of a member of the *RPP7* cluster (Li et al. 2020).

Because RPW8 proteins overall contribute to broad-spectrum disease resistance, they are likely targets for pathogen effectors.

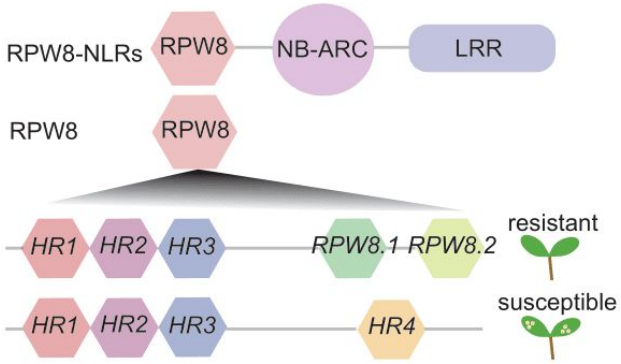


Fig 6. RPW8 proteins can be part of NLRs (RPW8-NLRs or RNLs) or occur as stand-alone proteins. Examples of common haplotypes in the *RPW8* locus, which consists of stand-alone RPW8 proteins. Both haplotypes carry the *HR1-3* genes, but some carry *RPW8.1/RPW8.2* while others carry HR4 instead. Accessions carrying *RPW8.1*/*RPW8.2* are resistant to powdery mildew, while those *HR4* are susceptible.

Collateral Damage of Diversification in Components of the Plant Immune System

Plants and pathogens typically have conflicting interests: while the pathogen wants to avoid recognition by the plant to enable its colonization, the plant wants to recognize as many pathogens as possible to prevent or contain pathogen infection. As a result of this tug-of-war, some members of the plant immune system are highly diversified. Such is the case for NLR intracellular receptors and of some non-NLR resistance genes. This extensive sequence diversity found across some members of the plant immune system, while necessary to fend off pathogens, can backfire and lead to genetic incompatibilities, sometimes resulting in hybrid necrosis. This phenomenon was formally described as early as 1943 where crosses between two different wheat varieties resulted in lethal necrosis in the resulting hybrid progeny (Caldwell and Compton 1943). Hybrid necrosis cases have since been identified in numerous plants, including rice, lettuce, cotton, tomato, *Capsella* spp., *Mimulus tilingii* and *A. thaliana* (Krüger et al. 2002; Bomblies et al. 2007; Alcázar et al. 2009; Jeuken et al. 2009; Yamamoto et al. 2010; Chen et al. 2014; Chae et al. 2014; Todesco et al. 2014; Sicard et al. 2015; Deng et al. 2019; Barragan et al. 2019; Sandstedt, Wu, and Sweigart 2020). The genetic architecture underlying these incompatibilities is usually simple, often involving one locus from each parent that deleteriously interact in the hybrid progeny.

Hybrid Necrosis Can Act as a Postzygotic Reproductive Barrier

While genetic incompatibilities between divergent elements of the plant immune system can be viewed as a carryover from the evolutionary dynamics between plants and their pathogens, these genetic incompatibilities also act as postzygotic barriers, contributing to the process of speciation. Over 100 years ago, a model explaining the evolution of genetic incompatibilities

was first described, this is now called the Bateson-Dobzhansky-Muller (BDM) Model (**Fig 7**) (Orr 1996). Here, subpopulations of a single ancestral population diverge from another, and independent mutations arise at individual loci and can become fixed, either because they provide some adaptive advantage or are neutral in their own context. When divergent loci suddenly co-exist in hybrids – the progeny of individuals from two different subpopulations – genetic incompatibilities can occur due to epistatic, non-additive interaction between these two loci (**Fig 7**). These kinds of genetic interactions are what lead to the *A. thaliana* hybrid incompatibility cases I will present here.

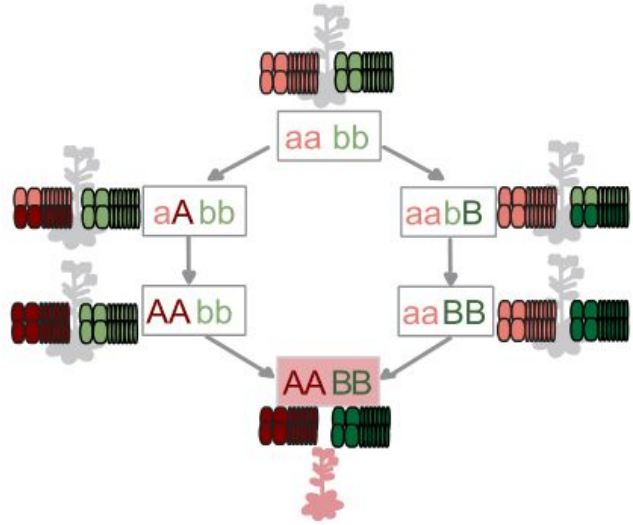


Fig 7. Schematic representation of the Bateson-Dobzhansky-Muller Model. An ancestral population with two genotypes (aabb) at two different loci (aa and bb) becomes separated. The resulting subpopulations diverge from another and new mutations arise at these two loci (aA and bB), resulting in new alleles. These new divergent alleles are not deleterious in their own context and can become fixed in each subpopulation (AA and BB). When these divergent alleles meet in a hybrid plant however (AABB), genetic incompatibilities can arise.

Hybrid Necrosis in *Arabidopsis thaliana*

In *A. thaliana,* hybrid incompatibilities were first identified after two accessions were crossed to each other, and the hybrid progeny was observed to show reduced growth and necrosis (Bomblies et al. 2007). The loci underlying hybrid incompatibility were identified as *DANGEROUS MIX 1* (*DM1*) originating from one parent, and its genetically unlinked interacting incompatibility partner *DM2,* originating from the other parent. Both the *DM1* and *DM2* loci are made up of NLR genes. *DM1* is also known as *SSI4*, whereas *DM2* is located in the highly variable *RPP1* NLR cluster. Later on, and in order to systematically analyze hybrid incompatibilities in *A. thaliana*, a diallel of 80 accessions was created, resulting in over 6,400 crosses (Chae et al. 2014). These 80 accessions were genetically diverse and had been collected from populations across *A. thaliana*'s native range in Eurasia (J. Cao et al. 2011). In addition to *DM1* and *DM2*, the hybrid incompatibility loci *DM3-DM9* were identified in other crosses resulting from this diallel through linkage mapping (Chae et al. 2014) (**Fig 8**). Four key

observations were either newly gained or confirmed from this experiment: (i) Hybrid necrosis is not rare, around 2% of all crosses show incompatibilities; (ii) There are different necrosis severity classes in the hybrid plants, ranging from very mild to lethal in the F_1 hybrid; (iii) There seems to be no correlation between the propensity of two accessions being incompatible with each other and the geographical distance separating the two; (iv) Regions rich in NLR genes are hybrid incompatibility hotspots, and the *DM* loci that are not NLR genes are usually otherwise involved in the plant's immune response (**Fig 8**).

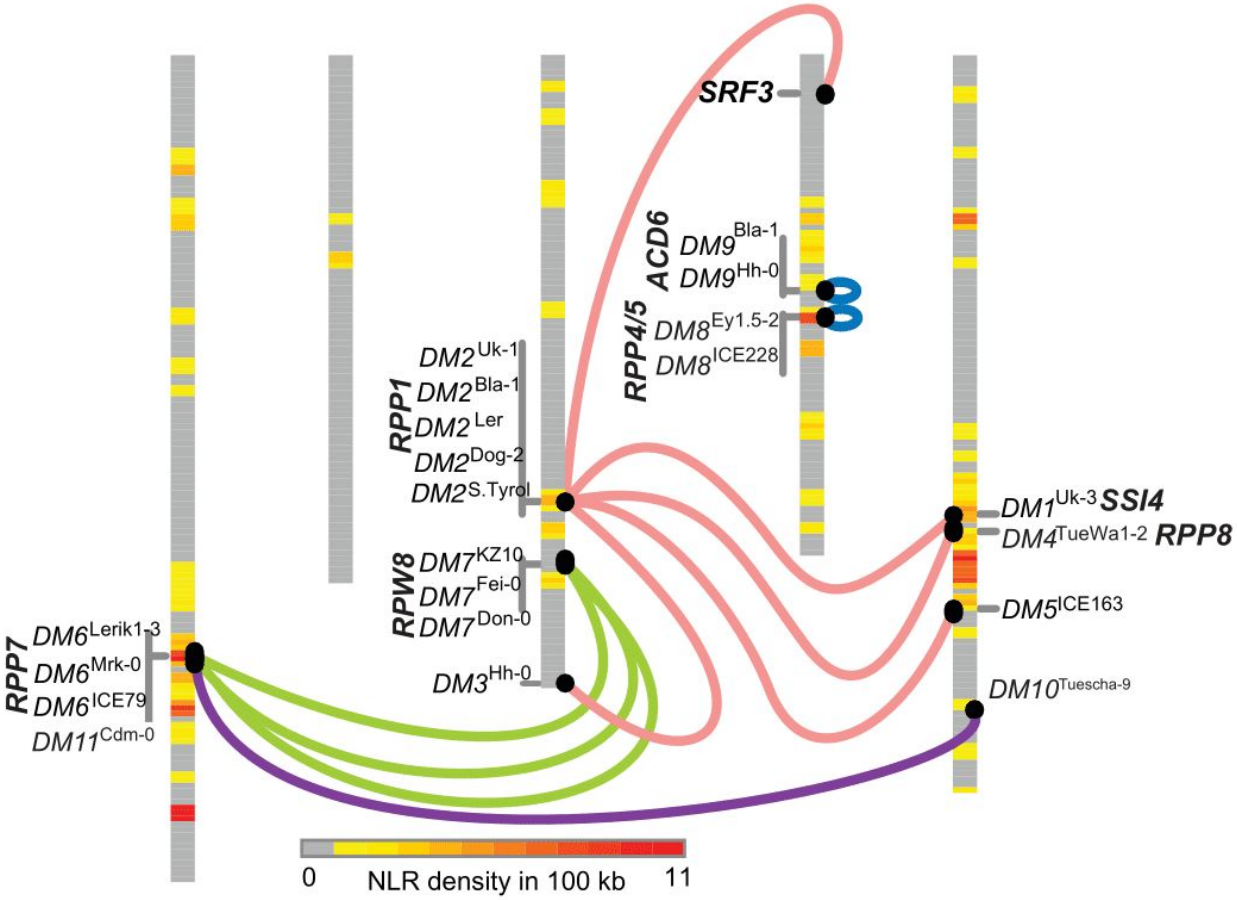


Fig 8. Summary of identified hybrid incompatibility cases in *A. thaliana*. Each line indicates an incompatibility case. Loci containing high density of NLR genes (yellow and red) are hybrid incompatibility hotspots. This present work will focus on hybrid incompatibilities between *RPP7* and *RPW8* (green lines) and *DM10* and *DM11* (purple line) Figure adapted and updated from (Chae et al. 2014).

Plant Mating System and Hybrid Necrosis

So far, plant hybrid incompatibility cases following the previously described BDM model have been mainly described in species or cultivated variants that are predominantly selfing. In addition, the natural co-occurence of hybrid incompatibility alleles have been shown to occur in *A. thaliana* (Todesco et al. 2014). Questions that remain unanswered are whether these kinds of BDM incompatibilities exist in outcrossing plant species, if so how common these are, and if the incompatible loci naturally co-occur. In selfing plants, slightly deleterious mutations tend to accumulate easier than in outcrossing plants, which are typically better at purging deleterious alleles (Deborah Charlesworth and Willis 2009). However, mutations that have strong negative effects on a plant would be under strong purifying selection in selfers, since these cannot be masked by heterozygosity (Arunkumar et al. 2015). Then again, many of the deleterious epistatic interactions between components of the immune system I describe here are dominant or semi-dominant (although there are likely many more that have not yet been described which are recessive), and do not have to occur in a homozygous state to cause incompatibility in the first place (**Fig 7**). Generally, natural selection is expected to eliminate genetic incompatibilities from populations where individuals are interbreeding, unless the advantage incompatible loci confer when present outweigh the disadvantages caused by their potential incompatibility. It is thus not clear how common deleterious epistatic interactions are among freely interbreeding outcrossing plants.

Aims of this PhD Thesis

The aims of the first two projects I present in this thesis, are to first identify further hybrid incompatibility cases in *A. thaliana*, then to characterize the underlying genetic interactions between the incompatible loci in the studied cases, and then to study the evolutionary dynamics that ultimately result in these incompatibilities. I focus on two sets of hybrid incompatibility cases, each described in a separate project. The first set consists of three independent cases of allele-specific deleterious interactions between members of the NLR cluster *RPP7* (*DM6*) and of the non-NLR disease resistance cluster *RPW8* (*DM7*). The second set studies a case of severe incompatibility between the truncated singleton NLR *DM10* and the unlinked locus *DM11* (**Fig 8**). Lastly, in the third project, I study whether these kinds of deleterious epistatic interactions that lead to hybrid incompatibilities occur in natural *A. arenosa* populations, an outcrossing relative of *A. thaliana.*

Some questions I will address are:

- Which loci are causal for the newly identified incompatibility cases?
- What are the differences between alleles and resulting protein variants that cause incompatibility versus those that do not?
- What is the evolutionary history of these incompatible loci?
- Are there any factors that make these loci more prone to cause incompatibilities than others?
- What transcriptional changes does the necrotic hybrid experience when compared to its parents?
- Do these kinds of BDM incompatibilities exist among naturally co-occurring outcrossing plants? If so, how common are these?

By answering these and other questions, insights into the causes and consequences of hybrid incompatibilities between divergent elements of the plant immune system will be gained, as well as into the prevalence of these incompatibilities in outcrossing plants.

Chapter One

RPW8/HR4 repeats control NLR activation in *Arabidopsis thaliana*

Citation

Barragan AC, Wu R, Kim ST, Xi W, Habring A, Hagmann J, Van de Weyer AL, Zaidem M, Ho WWH, Wang G, Bezrukov I, Weigel D, Chae E (2019) RPW8/HR repeats control NLR activation in *Arabidopsis thaliana*. *PLoS Genetics* 15:e1008313
https://doi.org/10.1371/journal.pgen.1008313

Abstract

In many plant species, conflicts between divergent elements of the immune system, especially nucleotide-binding oligomerization domain-like receptors (NLR), can lead to hybrid necrosis. Here, we report deleterious allele-specific interactions between an NLR and a non-NLR gene cluster, resulting in not one, but multiple hybrid necrosis cases in *Arabidopsis thaliana*. The NLR cluster is *RESISTANCE TO PERONOSPORA PARASITICA 7* (*RPP7*), which can confer strain-specific resistance to oomycetes. The non-NLR cluster is *RESISTANCE TO POWDERY MILDEW 8* (*RPW8*) / *HOMOLOG OF RPW8* (*HR*), which can confer broad-spectrum resistance to both fungi and oomycetes. RPW8/HR proteins contain at the N-terminus a potential transmembrane domain, followed by a specific coiled-coil (CC) domain that is similar to a domain found in pore-forming toxins MLKL and HET-S from mammals and fungi. C-terminal to the CC domain is a variable number of 21- or 14-amino acid repeats, reminiscent of regulatory 21-amino acid repeats in fungal HET-S. The number of repeats in different RPW8/HR proteins along with the sequence of a short C-terminal tail predicts their ability to activate immunity in combination with specific RPP7 partners. Whether a larger or smaller number of repeats is more dangerous depends on the specific RPW8/HR autoimmune risk variant.

Author contributions

Author	Author position	Scientific ideas %	Data generation %	Analysis & interpretation %	Paper writing %
A. Cristina Barragan	**1**	**55**	**55**	**60**	**50**
Rui Wu	2	5	10	5	
Sang-Tae Kim	3	5		5	
Wanyan Xi	4	5	5		
Anette Habring	5		5		
Jörg Hagmann	6		2.5		
Anna-Lena Van de Weyer	7		2.5		
Maricris Zaidem	8		2.5		
William Ho	9		2.5		
George Wang	10		2.5		
Ilja Bezrukov	11		2.5		
Detlef Weigel*	12	15		15	25
Eunyoung Chae*	13	15	10	15	25
Title of paper: RPW8/HR4 repeats control NLR activation in *Arabidopsis thaliana*					
Status in publication process: Paper published in *PLoS Genetics* on 25. July 2019					

* Corresponding authors

Chapter Two

A truncated singleton NLR causes hybrid necrosis in *Arabidopsis thaliana*

Citation

Barragan AC, Collenberg M, Wang J, Lee RRQ, Cher WT, Rabanal FA, Ashkenazy H, Weigel D, Chae E, (2020) A truncated singleton NLR causes hybrid necrosis in *Arabidopsis thaliana. Molecular Biology and Evolution* https://doi.org/10.1093/molbev/msaa245

Abstract

Hybrid necrosis in plants arises from conflict between divergent alleles of immunity genes contributed by different parents, resulting in autoimmunity. We investigate a severe hybrid necrosis case in *Arabidopsis thaliana*, where the hybrid does not develop past the cotyledon stage and dies three weeks after sowing. Massive transcriptional changes take place in the hybrid, including the upregulation of most NLR disease resistance genes. This is due to an incompatible interaction between the singleton TIR-NLR gene *DANGEROUS MIX 10* (*DM10)*, which was recently relocated from a larger NLR cluster, and an unlinked locus, *DANGEROUS MIX 11 (DM11)*. There are multiple *DM10* allelic variants in the global *A. thaliana* population, several of which have premature stop codons. One of these, which has a truncated LRR-PL region, corresponds to the *DM10* risk allele. The *DM10* locus and the adjacent genomic region in the risk allele carriers are highly differentiated from those in the non-risk carriers in the global *A. thaliana* population, suggesting that this allele became geographically widespread only relatively recently. The *DM11* risk allele is much rarer and found only in two accessions from southwestern Spain – a region from which the *DM10* risk haplotype is absent – indicating that the ranges of *DM10* and *DM11* risk alleles may be non-overlapping.

Author contributions

Author	Author position	Scientific ideas %	Data generation %	Analysis & interpretation %	Paper writing %
A. Cristina Barragan	**1**	**70**	**65**	**65**	**60**
Maximilian Collenberg	2		17.5	5	
Jinge Wang	3		2.5	2.5	
Rachelle Lee	4		2.5	2.5	
Wei Yuan Cher	5		2.5		
Fernando A. Rabanal	6		2.5		
Haim Ashkenazy	7		2.5		
Detlef Weigel*	8	15			20
Eunyoung Chae*	9	15	5	25	20
Title of paper: A truncated singleton NLR causes hybrid necrosis in *Arabidopsis thaliana*					
Status in publication process: Accepted in *Molecular Biology and Evolution* on 23. September 2020					

* Corresponding authors

Chapter Three

A case of inbreeding depression in a natural *Arabidopsis arenosa* population

Citation

Barragan AC, Collenberg M, Kerstens M, Schwab R, Bemm F, Bezrukov I, Požárová D, Kolář F, Weigel D. A case of inbreeding depression in a natural *Arabidopsis arenosa* population

Abstract

Hybrid incompatibility in plants is usually the result of pairwise deleterious epistatic interactions between one or two loci, which often, but not always, encode for components of the immune system. In *A. thaliana*, the geographical co-occurence of incompatible alleles in natural settings has been shown. What remains elusive, is whether co-occurring incompatible alleles also exist in natural populations of outcrossing plant species, and if so, how common these are. To address this question, we screened over two thousand naturally occuring *A. arenosa* hybrid plants in search for potential incompatibilities. We show that while deleterious phenotypes are common and heritable in these plants, their molecular phenotype differs from that seen in incompatible *A. thaliana* hybrids. In addition, we identified a genomic region associated with one of these abnormal phenotypes through linkage mapping, and show that this region is highly homozygous in affected individuals, indicating that inbreeding depression rather than pairwise genetic incompatibilities may, at least in some cases, give rise to the deleterious phenotypes observed.

Author contributions as of October 2020

Author	Author position	Scientific ideas %	Data generation %	Analysis & interpretation %	Paper writing %
A. Cristina Barragan	**1**	**65**	**65**	**75**	**70**
Maximilian Collenberg	2	10	12.5	20	
Merjin Kerstens	3	5	5	5	
Rebecca Schwab	4		5		
Felix Bemm	5		5		
Ilja Bezrukov	6		2.5		
Doubravka Požárová	7		2.5		
Filip Kolář	8		2.5		
Detlef Weigel*	9	20			30
Title of paper: A case of inbreeding depression in a natural *Arabidopsis arenosa* population					
Status in publication process: Manuscript in preparation					

* Corresponding author

Discussion

In this thesis, I studied two sets of hybrid incompatibility cases in *A. thaliana*: the first set consisted of three independent deleterious interactions between the NLR cluster *RPP7* and the non-NLR cluster *RPW8*, while the second set was a case of severe incompatibility case between the singleton NLR *DM10* and *DM11*, an unlinked locus. In addition, I studied the prevalence of hybrid incompatibilities among individuals from natural *A. arenosa* populations.

In the first project, I identified a third hybrid necrosis case where *RPP7* and *RPW8* were linked to the hybrid phenotype. I then experimentally confirmed the causality of these two loci in the three known incompatibility cases involving *RPW8* and *RPP7*. I then showed that *RPW8*, although not an NLR cluster, is highly variable among *A. thaliana* accessions. Subsequently, I studied the underlying genetics, especially from the *RPW8* side, that lead to incompatibilities in the hybrid plant in two of these three necrosis cases. I identified the genes in the *RPW8* cluster which cause incompatibility when in the presence of certain *RPP7* variants, *HR4* and *RPW8.1*, and found that C terminal repeat variations present in the resulting proteins of these two genes largely modulate the severity of the hybrid phenotype.

In the second project, I identified the most severe hybrid necrosis case described so far as being the result of a deleterious interaction between two loci: *DM10* and *DM11*. I showed that the hybrid shows massive transcriptional changes when compared to either parent, including the upregulation of defense responses and the overexpression of most NLRs. In addition, I determined that the DM10 variant causing incompatibility is a truncated singleton NLR protein which is both common and geographically widespread. *DM10* is the first singleton NLR identified as causal for incompatibility so far, although, as I established, it was actually relocated from an NLR cluster after *A. thaliana* speciation. I then found that the geographical co-occurrence between the mismatching *DM10* and *DM11* alleles seems to be absent in natural *A. thaliana* metapopulations.

In the third project, I collected and then screened over 2,300 *A. arenosa* individuals from natural populations, and observed that deleterious abnormal phenotypes are common and heritable. The transcriptional profile of lines showing an abnormal phenotype differs from that seen in incompatible *A. thaliana* hybrids from the second project. In one particular line a region associated with the abnormal phenotype was identified through linkage mapping, this region was shown to be highly homozygous in abnormal plants, indicating that this phenotype is the result of inbreeding depression.

In short, the results I presented in this thesis show that: (i) Some loci are more prone to be incompatible with each other; (ii) The causal loci for incompatibility are typically either part of gene clusters or derive from them; (iii) Relatively minor sequence changes in a gene and its resulting protein (e.g length variations) is enough to determine whether they will cause incompatibility or not; (iv) Variants that cause incompatibilities can be common and geographically widespread, likely as a result of balancing selection acting on them and (v) deleterious phenotypes are common among wild *A. arenosa* individuals, and in at least some cases, likely the result of inbreeding depression.

Functional Links Between Two Loci and Hybrid Incompatibility

As previously mentioned, a locus which confers resistance to multiple races of the same pathogen, or to different pathogens altogether, is a likely target of pathogen effectors. The integrity of proteins commonly targeted by pathogens (guardees/decoys) is often monitored by other proteins (guards). One factor potentially influencing how likely it is that two loci mismatch and trigger autoimmunity in hybrid plants, may be those that are functionally linked to each other, such as guard and guardee pairs. For example, RIN4, a member of a highly conserved protein family, acts as an immune signalling hub and is targeted by multiple pathogen effectors across various plant species (Toruño et al. 2019). In an interspecific lettuce cross, *RIN4* and what seems to be an NLR that guards *RIN4*, are incompatible with each other, resulting in hybrid autoimmunity (Jeuken et al. 2009). A similar scenario is likely to occur between *RPP7* and *RPW8*, where allele-specific interactions between these two loci represent a third of all hybrid incompatibility cases identified so far in *A. thaliana*. RPW8 proteins are involved in broad-spectrum disease resistance, making them a likely target for pathogen effectors.

The genetic results from Chapter 1 of my thesis provided a solid foundation for a subsequent study that led to the hypothesis that RPP7 likely acts as a guard of HR4, a gene in the RPW8 cluster, by monitoring changes to HR4 (Li et al. 2020) (**Fig 9**). While some HR4 variants have been shown to trigger cell death on their own by disrupting membrane integrity, they can also induce RPP7 oligomerization, thereby initiating an immune response (Li et al. 2020) (**Fig 9**). This higher order complex formation as part of the plant immune response is reminiscent of the previously introduced ZAR1 resistosome. According to this logic, ZAR1 and its guardee PBL2 would also be more likely to be incompatible with each other than expected by chance, and so would RPM1 and RPS2 guards be more likely to be incompatible with their guardee RIN4 in *A. thaliana*. Whether this is actually the case, will be seen as more hybrid incompatibility cases are

identified and the causal loci mapped. One notable difference between *ZAR1*, *RPM1* and *RPS2* with *RPP7* however, is that all three loci are singleton NLRs, while *RPP7* is a clustered NLR locus, and as we have discussed, clustered NLRs seem to be more prone to cause hybrid incompatibilities than singleton NLRs due to their highly dynamic nature.

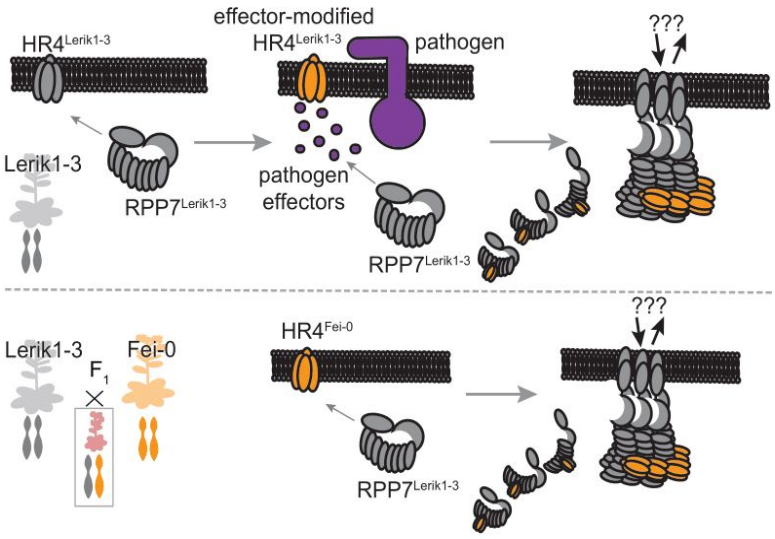


Fig 9. Model of the proposed HR4-RPP7 interaction. In an *A. thaliana* inbred accession like Lerik1-3 (top), RPP7 monitors the integrity of HR4. Upon pathogen infection, HR4 is targeted and modified by pathogen effectors, RPP7 detects this change, leading to RPP7-HR4 oligomerization. This newly formed higher-order complex may have the ability to insert itself into the plasma membrane, causing differential ion fluxes, and ultimately leading to cell death. In the Lerik1-3 x Fei-0 hybrid plant (bottom), there are two sets of NLRs, one from each parent. $RPP7^{Lerik1-3}$ may wrongly recognize $HR4^{Fei-0}$ as being modified by effectors and trigger an immune response in the absence of a pathogen.

Overall, by identifying loci that tend to be incompatible with each other, one may pick up on novel functional links between two proteins, like guard and guardee pairs or, more generally, detect pathogen effector target hubs. Conversely, known functional links between components of the plant immune system may help identify likely candidates for hybrid incompatibility once e.g linkage mapping has been performed and a candidate genomic region identified.

Physical Characteristics of a Locus and Hybrid incompatibility

In singleton NLRs, variation is typically seen as presence/absence polymorphisms (P/A), like in *RPS5* in *A. thaliana* (Henk, Warren, and Innes 1999), or as segregating resistant/susceptible alleles like in *RPS2* (R/S) (Caicedo, Schaal, and Kunkel 1999) (**Fig 10A**). The two allelic variants in these and other cases, are likely maintained in metapopulations though balancing selection (Caicedo, Schaal, and Kunkel 1999; Stahl et al. 1999; Mauricio et al. 2003; Bakker et al. 2006; Koenig et al. 2019; Gos, Slotte, and Wright 2012; Tian et al. 2002). In addition to P/A and R/S, the formation of allelic series is another way variation is observed in singleton NLRs. *RPP13* in *A. thaliana* and the *L* locus in flax are examples of allelic series, where different functional alleles confer disease resistance to different races of the same pathogen (**Fig 10A**).

RLM1A and *RLM1B* are closely related to *DM10* and confer resistance to *L. maculans*, causal agent of blackleg disease (Staal et al. 2006). *RPW8.1* confers resistance to filamentous pathogens, and *HR4* is its close relative (S. Xiao et al. 2004). Although we do not know what the function of *HR4* or *DM10* is, it would seem conceivable that both could confer resistance to an unknown pathogen and form allelic series, since multiple haplotypes are present in the global *A. thaliana* population which are maintained at intermediate frequencies. On the other hand, multiple haplotypes present at a locus may not necessarily mean that they confer resistance to a pathogen, or that they are even functional. This leads us to a big question in the field of NLRs and natural variation: How many of the *R* genes and alleles we observe both in an individual plant and across populations actually confer disease resistance to a pathogen, how many function in basal disease resistance, and how many simply act as sequence reservoirs for generating future NLR or *R* gene diversity?

Until *DM10*, all of the hybrid incompatibility cases identified were caused by NLR genes found in clusters. This highlights the proneness of this genomic arrangement in generating incompatibility alleles. When comparing gene clusters however, are there any specific factors which determine whether a particular cluster will be more likely to generate incompatible alleles? The overall number of genes in the cluster may be one predictive factor, since most *DM* loci are found in relatively large gene clusters like *DM1* (*SSI4*), *DM2* (*RPP1*), *DM4* (*RPP8*), *DM6* (*RPP7*), *DM7* (*RPW8*) and *DM8* (*RPP4/5*). Here, the median number of genes in the clusters ranges from around three to twelve genes, which is above average for *R* gene clusters, which often consist of only 2 or 3 genes (Jiao and Schneeberger 2020) (**Fig 10B**). In addition, it seems that clusters with more genes also tend to have an increased variability in the number of genes across different accessions (Jiao and Schneeberger 2020) (**Fig 10C**). Recently, it was observed that inside NLR clusters, particular “radiating” genes are responsible for most expansion and contraction events found across different *A. thaliana* accessions, whereas other cluster members are more conserved and are “high-fidelity” genes (Lee and Chae 2020). This confirmed what had been previously described in a study comparing the *RGC2* cluster across multiple lettuce genotypes, where it was shown that not all NLR genes evolve at the same speed (H. Kuang et al. 2004). Some genes are rapidly-evolving and undergo frequent sequence exchange between paralogues (type I) while others are more conserved between genotypes (type II) (H. Kuang et al. 2004; Hanhui Kuang et al. 2008). In addition, a study comparing NLR loci across 64 *A. thaliana* accessions found that among the loci consisting of many type I or radiating genes were *DM4* (*RPP8*), *DM6* (*RPP7*) and *DM8* (*RPP4/5*) (Van de Weyer et al.

2019). This indicates that the likeliness of a locus being involved in hybrid incompatibility is linked to the number of genes it has in the cluster which is itself tied to the number of type I, or radiating genes it has. These three loci: *DM4* (*RPP8*), *DM6* (*RPP7*) and *DM8* (*RPP4/5*), carry many type I genes and are known to have paralogs in different accessions which confer resistance to different strains of the same pathogens or to different pathogens altogether (McDowell et al. 1998; Cooley et al. 2000; Takahashi et al. 2002; van der Biezen et al. 2002; J. Liu et al. 2015; Asai et al. 2018). This emphasizes the importance of rapidly evolving type I genes in generating new disease resistance specificities, but also how rapid evolution can sometimes lead to high levels of polymorphisms, which can backfire by generating alleles that result in hybrid incompatibilities.

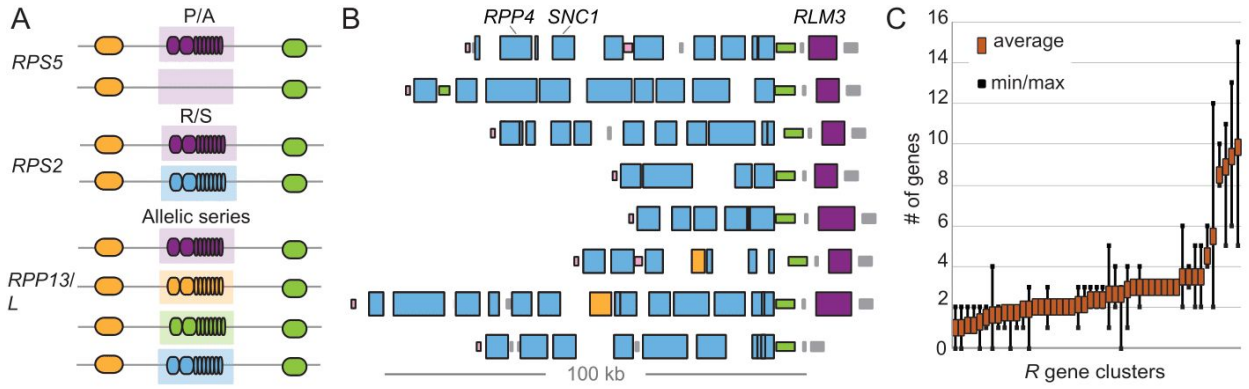


Fig 10. **A**. Examples of allelic variation at individual NLR loci. P/A, presence/absence. R/S, resistant/susceptible. **B.** *RPP4*/*RPP5* NLR cluster across eight different *A. thaliana* accessions. Tall rectangles represent NLR genes, and short rectangles non-NLR genes. Colors indicate close sequence relatedness. There is a highly variable number of blue NLR genes in the cluster (including the functionally defined members *RPP4*, *RPP5* and *SNC1*), while the adjacent singleton *RLM3* gene (magenta) shows P/A polymorphism. The *RPP4*/*RPP5* cluster has also been invaded by an unrelated NLR gene (yellow). Finally, some of the non-NLR genes (pink, green) flanking the cluster are duplicated as well. **C**. Average number of genes per NLR gene cluster across eight *A. thaliana* accessions. Most have on average fewer than three genes, while a few are both highly expanded and highly variable in numbers. Data for B and C from (Jiao and Schneeberger 2020).

Plant Mating System and Inbreeding Depression

Apart from the functional and structural properties of NLR and non-NLR *R* genes, a third factor which could influence the occurrence of hybrid incompatibilities following the Bateson-Dobzhansky-Muller Model (BDM) is the plant's mating system. The hybrid incompatibility cases I describe in the first two projects are based on deleterious epistatic interactions between two diverging loci in *A. thaliana*, a predominantly selfing species. We studied whether these kinds of BDM incompatibilities occur in natural outcrossing *A. arenosa*

populations. We found that although abnormal and likely deleterious phenotypes are common in natural *A. arenosa* plants these are not necessarily the product of hybrid incompatibilities but, at least in some cases, the result of inbreeding depression. This may in part be expected, since inbreeding depression has been observed and is thought to be quite common in wild populations of both plant and animal species (Keller and Waller 2002). The negative fitness effect inbreeding depression has on an individual, is believed to be the major force counteracting the transition from an outcrossing to a selfing mating system (D. Charlesworth and Charlesworth 1987). We cannot exclude that BDM hybrid incompatibilities exist in *A. arenosa*, but unlike in *A. thaliana,* inbreeding depression also plays a larger role in giving rise to deleterious phenotypes.

Genetic Diversity and Disease at Immune Loci in Animals

In plants, postzygotic genetic incompatibilities found in hybrids are often due to divergent elements of the plant immune system that, upon meeting in a hybrid, trigger autoimmunity. To date, animal NLR loci have not been identified to cause hybrid incompatibility. This may be somewhat expected, since animal NLRs are generally less expanded than in plants, humans for example only have 22 NLRs (Yifei Zhong, Kinio, and Saleh 2013). This is likely a result of animal NLRs functioning as cytosolic PRRs that recognize conserved PAMP molecules rather than rapidly evolving pathogen effectors (A. Bentham et al. 2017). Only in the absence of an adaptive immune system have animal NLR expansions been observed, like in corals and sea urchins (Hibino et al. 2006; Hamada et al. 2013). Although genetic BDM-like incompatibilities have not been identified, mutations and misregulations in animal NLRs have been associated with inflammatory and autoimmune diseases (Yifei Zhong, Kinio, and Saleh 2013).

Loci that may be under similar selective pressures as plant NLR genes are vertebrate MHC complexes, where pathogen-driven selection maintains high levels of genetic variation (Radwan et al. 2020). MHC loci enable “self” and “non-self” recognition in T-cells, which is critical in initiating an immune response against pathogens, and where miseregulation has also been associated with autoimmunity (Trowsdale and Knight 2013). In primates, it has been shown that while some level of heterozygosity in MHC alleles is beneficial (Penn, Damjanovich, and Potts 2002), too much divergence can also be detrimental by increasing the risk of developing multiple autoimmune diseases (Matzaraki et al. 2017). In both plant NLR and vertebrate MHC loci it seems that there is a diversity optimum, where too much variation increases the risk of autoimmunity, and too little results in compromised defenses against pathogen attacks.

Something to keep in mind when discussing about hybrid incompatibilities in outcrossing species, which includes almost all vertebrates, is that all individuals are by default hybrids, since they are progeny of a cross between two different individuals. So, just like with divergence at plant immune loci, deleterious effects resulting from high heterozygosity at MHC loci would, in principle, represent cases of hybrid incompatibility in animals.

Capturing Plant NLR Diversity Through Pan-NLRomes

From what we have discussed so far, the fact that there is tremendous variation in NLR genes, especially in those that are found in clusters, has been emphasized. In recent years, it has become abundantly clear that one cannot not speak of a single linear sequence as “the genome” of a species. In plants, one of the earliest indications that this concept was misleading came from comparing haplotypes of individual NLR clusters (Botella et al. 1998; McDowell et al. 1998; Noël et al. 1999; H. Kuang et al. 2004; Srichumpa et al. 2005). Even when comparing relatively few haplotypes, extensive variation at all levels was observed: in the coding sequence of a single gene, in copy number and in genomic location. However, while we in principle know the molecular processes that can create the observed differences (point mutations, transposon insertions, deletions, duplications, other types of chromosomal rearrangements, gene conversion, illegitimate recombination events etc.) (**Fig 5**), we still largely do not know the true extent of NLR diversity within a species, both in terms of presence and absence of individual genes nor in terms of haplotype diversity, nor of the underlying evolutionary forces generating and differentially maintaining this diversity. Two decisive factors in answering these questions will be reconstructing the evolutionary history of NLR loci, and comparing haplotypes that have survived in a population. The only way to do this, is by assembling more and more genomes from the same species and comparing their NLR gene content, resulting in the complete set of NLRs of a species, its pan-NLRome (**Fig 11**).

From initial pan-genomic studies, which look at the entire gene and allele repertoire of a species (Tettelin et al. 2005), it was observed that NLR genes are overrepresented in the variable fraction of gene content across accessions, called shell and cloud genes (Golicz et al. 2016; Montenegro et al. 2017; Hurgobin et al. 2018) (**Fig 11**). For example, while only 19% of the *Brassica oleracea* pan-genome was composed of genes missing from the reference, the number of NLR genes missing from the reference was almost 60% (Golicz et al. 2016). Similarly, 50% of the 307 NLR genes found across 53 *B. napus* accessions were absent in the Darmor-bzh reference (Hurgobin et al. 2018). These findings have been confirmed with more

recent and more complete long-read assemblies of multiple accessions, for example in soybean (Y. Liu et al. 2020).

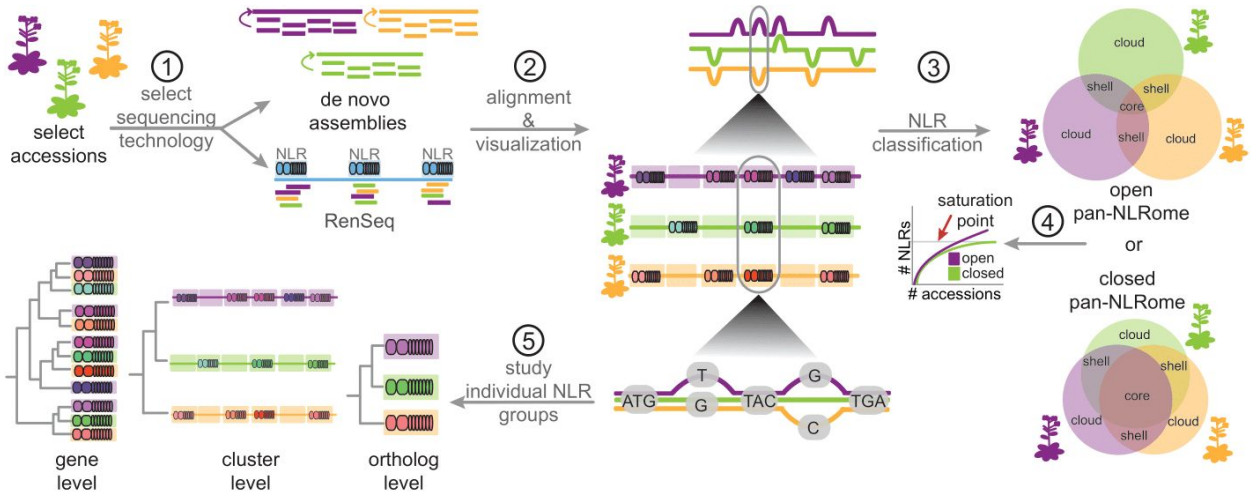


Fig 11. Potential workflow of the construction, visualization and analysis of a pan-NLRome. First, accessions to be sequenced are selected in a manner that maximizes diversity in order to approach a saturated pan-NLRome as quickly as possible (1). Then either the whole genome or the NLR complement is sequenced and assembled (2). The assemblies from each genome are then aligned to each other and potentially combined into a genome graph (3). Core and accessory genes are identified and the saturation point of the NLRome is assessed, to determine whether the NLRome is open or closed (4). Finally, individual NLR groups can be studied, e.g., one can analyze the phylogeny of a certain cluster across accessions, by comparing individual alleles in those accessions that share the same genes, the structure of whole clusters, or the entire set of genes present in all clusters (5).

Two recent studies focused specifically on the pan-NLRome with the goal of capturing a substantial fraction of inter- and intraspecific NLR diversity. The first study compared 64 *A. thaliana* accessions (Van de Weyer et al. 2019), while the second study spanned 16 accessions from five different *Solanum* species, plus single accessions of *Nicotiana benthamiana* and *Capsicum annuum*, also from the Solanaceae family (Seong et al. 2020). Both the *A. thaliana* and the Solanaceae pan-NLRome analyses confirmed that although there are apparently so-called core NLR genes that are present in all or almost all accessions, these account only for a minority of all NLR genes (Van de Weyer et al. 2019; Seong et al. 2020) (**Fig 11**). In addition, it was shown that the known NLR diversity in the *A. thaliana* study with 64 accessions quadrupled the number of NLR domain architectures known from the single reference genome. This may appear daunting, but it was also observed that a set of about 40 maximally diverse accessions would have been sufficient to discover the vast majority of NLR genes, reaching its so-called saturation point (Van de Weyer et al. 2019) (**Fig 11**). At a first glance, reaching saturation levels with relatively few individuals might perhaps appear surprising, but it likely

reflects strong selection for NLR genes, which are therefore more likely to spread faster throughout the entire population, and are therefore more evenly distributed.

Taken together, pan-genomic and pan-NLRome studies will speed up the identification of intraspecific NLR diversity, which although vast, seems to be finite. In the context of hybrid necrosis, by assessing the evolutionary history of incompatible NLR loci it will become clearer whether both incompatible alleles were present in a single population at some point, and if so for how long and at which frequencies. In the case of the risk *DM10* and *DM11* alleles, for example, it could be studied whether selective forces really do prevent their co-occurrence in local populations and if so since when.

Turning NLR Sequence Knowledge into Functional Knowledge

The ultimate goal of capturing NLR diversity through pan-NLRome studies, is to accelerate the discovery of NLR specificity in disease resistance. This has value in an evolutionary and ecological context, but also a practical value for plant breeding. The first step in capturing NLR diversity is the generation of species-wide NLR gene inventory, which is, as previously discussed, clearly within reach, and is reflected by the rapidly increasing number of high-quality plant pan-genomes released in this present year alone (Alonge et al. 2020; Y. Liu et al. 2020; Song et al. 2020; Jiao and Schneeberger 2020; Zhou et al. 2020; K. Cao et al. 2020; Haberer et al. 2020). The second step, using the pan-NLRomes to better understand how NLR diversity is distributed and to discern the evolutionary history of NLRs in a species, is more challenging, but we are optimistic that this can be solved as well. A highly innovative approach to parsing NLR diversity has recently been introduced, with a focus on revealing potential functional sites in NLRs that will aid in the rational design of novel or broad-spectrum disease resistance (Prigozhin and Krasileva, 2020). What remains is to comprehensively assign function to all existing NLRs. The most common known molecular activities of NLRs are (i) detecting a specific pathogen, or a specific pathogen effector, and (ii) enabling the functioning of other, pathogen-detecting NLRs (helper function). An additional activity for which there is less direct evidence so far, is NLR expression being part of broad-spectrum disease resistance after a pathogen attack.

For comprehensive assignment of NLR gene function, more direct functional tests need to be pursued. The simplest cases for direct functional tests are those where a function is already known for a particular NLR, and based on this knowledge a system can be devised for rapid investigation of sequence-related NLRs. An example is the *RPP13* allelic series in *A. thaliana*,

where transient expression was used to determine which *RPP13* alleles recognized which allele of the matching effector *ATR13* (R. L. Allen et al. 2004; Rebecca L. Allen et al. 2008; Krasileva et al. 2011). A caveat of this system is that sequence-related NLRs, even those that are found in the same genomic location, often do not always recognize the same effectors, or even the same pathogens. For example in the CNLs *RPP8*/*HRT*/*RCY1* in *A. thaliana*, where *RPP8* in the Ler accession endows plants with resistance to the Emco5 isolate of *Hpa*, *HRT* in the Dijon-17 accession to turnip crinkle virus, and *RCY1* in the C24 accession to the yellow strain of cucumber mosaic virus (McDowell et al. 1998; Cooley et al. 2000; Takahashi et al. 2002). By knowing the biochemical functions of these three encoded proteins (e.g., if they act as guards, do they all have the same guardee?), and by understanding the evolutionary history of this subfamily of NLR genes (MacQueen et al. 2019), the functions of these and of other alleles would be easier to predict.

In the context of hybrid incompatibility, many of the incompatible NLR and non-NLR alleles have either no known function in disease resistance, or at least do not have the same function as other alleles for which a function is known. For example, the *RPW8.1* allele in the Ms-0 accession that is known to confer broad-spectrum disease resistance to filamentous pathogens (S. Xiao et al. 2001; Ma et al. 2014), is not the same as the *RPW8.1* allele that I found causes incompatibility with certain *RPP7* alleles. Similarly, the *RPP7* allele that causes incompatibility, is not the same as the *RPP7* allele that confers resistance to downy mildew isolate Hiks1 (Slusarenko and Schlaich 2003; Li et al. 2020). By identifying the molecular activities of these genes and alleles, it could be studied whether alleles that cause incompatibility tend to have a particular function, like for example in basal vs. strain-specific disease resistance or if they tend to be non-functional and simply act as sequence reservoirs instead.

Apart from the fact that NLR sequence similarity does not necessarily equal functional similarity, a second challenge in identifying the function of NLRs lies in finding those which are truncated. Truncated genes are often more difficult to identify due to their lower overall percentage of sequence similarity when compared to a full-length allele. Even if truncated alleles are identified, assigning them a function is challenging since, unlike in many other proteins, truncations do not only not necessarily render NLRs non-functional, as we have seen for the truncated DM10 variants, which retain the ability to trigger cell death, but can also functionally differ from full-length proteins. An example of this can be seen in wheat, where YrSP, a truncated CNL that is missing most of its LRR domain, not only retained the ability to confer resistance to yellow

rust, but also developed recognition specificities differing from its highly similar (99.8% identity) and apparent full-length protein Yr5 (Marchal et al. 2018).

A more general approach to test for the function of an NLR, would be introducing these newly identified NLR genes and alleles into susceptible plant backgrounds, followed by testing with different races of the pathogen of interest (Yang et al. 2013; Zhang et al. 2015; L. Wang et al. 2019). Direct functional testing of NLRs by expressing them in a foreign background may, however, not always be so straightforward since NLRs often require other genetic components for functioning. Many NLRs act as sensor NLRs and rely on helper NLRs for conferring disease resistance (Jubic et al. 2019; Feehan et al. 2020; Wu et al. 2017). In the case of paired NLRs, where both are required for functioning, this limitation could be circumvented by introducing always both members of a pair, which are easily found due to their characteristic head-to-head genomic arrangement. More challenging are NLR networks that include unlinked NLRs, such as those found in the complex NRC immune network in the Solanaceae (Wu et al. 2017).

The hope is that, after collecting a vast collection of NLR genes and alleles, we will arrive at a point where NLR function can be predicted from sequence alone by combining knowledge of protein-protein interactions with other sources of information such as structures of NLRs and their interactors, as well as knowledge of host protein modifications by pathogen effectors. This way, predictive models of NLR function which may come into reach in the next decade. Finally, with global knowledge of NLRs and matching effectors in hand, mapping the co-occurrence such pairs in hosts and pathogens in time and space should reveal the true extent of past and ongoing arms races between hosts and their pathogens, as well as of the extent of a by-product of this arms race: hybrid incompatibility.

Hybrid Incompatibility and Current Breeding Practices

Hybridization through intercrossing inbred lines is widespread in current breeding practices, making the study of epistatic interactions among parental lines relevant for future crop crossing schemes. In addition, introgression of beneficial traits such as disease resistance from a different cultivar or from a wild relative into elite lines is also common, and hybrid necrosis is thought to occasionally occur as a by-product of these attempts of enhancing disease resistance. Avoiding potential incompatibilities such as those occurring between diverging members of the plant immune system should therefore be a goal, so that yield penalties are minimized and disease resistance is maximized.

Since a plant often has to cope with limited resources, allocation of these fluctuates depending on the plant's environmental pressures. Usually, there is a trade-off between immunity and growth, and maintaining resistance genes comes at a cost. This has been experimentally shown, for example, in the NLR genes *RPM1* and *RPS5*, which confer resistance to *Pseudomonas* bacteria, and where plants carrying a resistant allele at this locus show a decreased seed-set when compared to plants carrying a susceptible allele or no allele (Tian et al. 2003; Karasov et al. 2014). In incompatible hybrids, the balance between growth and immunity is disturbed, and while necrotic hybrids often show reduced growth and fertility, at least in some incompatibility cases, the hybrid is also more resistant to pathogens (Todesco et al. 2014). This same yield-resistance trade-off has been observed in crop breeding panels (Calvo-Baltanás, Wang, and Chae 2020).

Disease resistance is often rapidly overcome when single resistance genes are introgressed into elite lines, in rice for example, resistance conferred by a single resistance gene is expected to last only from three to ten years (Devi et al. 2015). In order to circumvent this issue, resistance gene pyramiding is often done, where multiple resistance genes are stacked to promote more durable and more broad-spectrum disease resistance. The rice cultivar Tetep presents a naturally occuring example of broad-spectrum disease resistance to the blast fungus thanks to the presence of multiple NLRs (L. Wang et al. 2019). The recapitulation of this naturally-occurring phenomenon of multiple resistance genes leading to broad-spectrum resistance is currently been used in a wide variety of commercial crop lines in species like rice, potato, wheat, soybean and tomato (Yamanaka and Hossain 2019; Hanson et al. 2016; N. Xiao et al. 2016; Brunner et al. 2010; Ghislain et al. 2019). However, negative epistasis among pyramided resistance genes can sometimes occur, leading to crop yield reductions. For example, an extensive study of different inbred rice lines found that a reason for the high yields in these plants, is the accumulation of susceptible alleles which have been actively selected for over many generations (Huang et al. 2015). This is likely the result of avoiding fitness penalties caused by the presence of resistance genes, and also of avoiding potential hybrid incompatibilities these genes might cause (Calvo-Baltanás, Wang, and Chae 2020). Negative epistasis among *R* genes may not only lead to reduced yield, but also to reduced disease resistance. For example, mismatches among different pyramided *Pm3* alleles in wheat can suppress each other's powdery mildew resistance (Stirnweis et al. 2014), and *Pm3* can also repress resistance of its rye-ortholog *Pm8* when introduced into the same background (Hurni et al. 2014). As datasets on genetic incompatibilities grow, at some point it will likely be possible

for incompatibilities between two loci to be predicted beforehand when designing crossing schemes in breeding. If it is known that two loci are prone to be incompatible with each other, and the allelic variants at these loci which cause incompatibility are identified, a link can be made between parental genotypes and the hybrid progeny's phenotype, and then the hybrid phenotype could be predicted from parental genotype combinations without actually having to cross these. These predictions would work similarly to the hybrid incompatibility cases presented here, where by knowing the *RPP7* and the *HR4* or *RPW8.1* alleles, or the *DM10* and *DM11* alleles present in each parent, one could already say beforehand how the hybrid phenotype will look like. Such predictive power over hybrid phenotypes would be highly valuable in plant breeding by saving time and money. However, while on the right track, we are currently still a long way from reaching widespread phenotypic predictions in crop systems. Recent advances in sequencing technologies and the creation of the aforementioned pan-genomes and pan-NLRomes will help in this ambitious endeavor though.

Conclusion

Hybrid incompatibility cases due to epistatic interactions between diverging elements of the plant immune system present a unique opportunity to study the mechanistic functioning and evolutionary dynamics between incompatible loci. The aims of the three projects I presented in this thesis were to: (i) Identify further hybrid incompatibility cases in *A. thaliana*; (ii) Characterize the underlying genetic interactions between the incompatible loci; (iii) Study the evolutionary dynamics that ultimately resulted in these incompatibilities and (iv) Compare the incidence of hybrid incompatibilities between a selfing and a outcrossing plant species. These aims were achieved. I identified a third incompatibility case involving *RPP7* and *RPW8* and also identified incompatibility between *DM10* and *DM11* as the underlying cause of a severe case of hybrid necrosis. I also identified and characterized the underlying genes for these incompatibility cases and established that length variations present in *RPW8* and *DM10*, whether it be repeat length variations or truncations, are enough to cause incompatibilities, and in the case of *RPW8*, to even quantitatively modulate the severity of the hybrid phenotype. In addition, I showed that some incompatible alleles are common and geographically widespread, indicating that some benefit is likely associated with them. Lastly, I showed that abnormal phenotypes are common in natural *A. arenosa* populations, and at least some, are likely the result of inbreeding depression rather than hybrid incompatibilities following the BDM model. Taken together, insights into the causes and consequences of hybrid incompatibilities between divergent

elements of the plant immune system and between two different *Arabidopsis* species were gained.

References

Adachi, Hiroaki, Mauricio P. Contreras, Adeline Harant, Chih-Hang Wu, Lida Derevnina, Toshiyuki Sakai, Cian Duggan, et al. 2019. “An N-Terminal Motif in NLR Immune Receptors Is Functionally Conserved across Distantly Related Plant Species.” *eLife* 8:e49956.

Alcázar, Rubén, Ana V. García, Jane E. Parker, and Matthieu Reymond. 2009. “Incremental Steps toward Incompatibility Revealed by Arabidopsis Epistatic Interactions Modulating Salicylic Acid Pathway Activation.” *Proceedings of the National Academy of Sciences* 106 (1): 334–39.

Allen, Rebecca L., Julia C. Meitz, Rachel E. Baumber, Sharon A. Hall, Sarah C. Lee, Laura E. Rose, and Jim L. Beynon. 2008. “Natural Variation Reveals Key Amino Acids in a Downy Mildew Effector That Alters Recognition Specificity by an Arabidopsis Resistance Gene.” *Molecular Plant Pathology* 9 (4): 511–23.

Allen, R. L., P. D. Bittner-Eddy, L. J. Grenville-Briggs, J. C. Meitz, A. P. Rehmany, L. E. Rose, and J. L. Beynon. 2004. “Host-Parasite Coevolutionary Conflict between Arabidopsis and Downy Mildew.” *Science* 306 (5703): 1957–60.

Alonge, Michael, Xingang Wang, Matthias Benoit, Sebastian Soyk, Lara Pereira, Lei Zhang, Hamsini Suresh, et al. 2020. “Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato.” *Cell* 182 (1): 145-161 .

An, Dong, Yong Zhou, Changsheng Li, Qiao Xiao, Tao Wang, Yating Zhang, Yongrui Wu, et al. 2019. “Plant Evolution and Environmental Adaptation Unveiled by Long-Read Whole-Genome Sequencing of Spirodela.” *Proceedings of the National Academy of Sciences of the United States of America* 116 (38): 18893–99.

Arunkumar, Ramesh, Rob W. Ness, Stephen I. Wright, and Spencer C. H. Barrett. 2015. “The Evolution of Selfing Is Accompanied by Reduced Efficacy of Selection and Purging of Deleterious Mutations.” *Genetics* 199 (3): 817–29.

Asai, Shuta, Oliver J. Furzer, Volkan Cevik, Dae Sung Kim, Naveed Ishaque, Sandra Goritschnig, Brian J. Staskawicz, Ken Shirasu, and Jonathan D. G. Jones. 2018. “A Downy Mildew Effector Evades Recognition by Polymorphism of Expression and Subcellular Localization.” *Nature Communications* 9 (1): 5192.

Baggs, Erin, J. Grey Monroe, Anil S. Thanki, Ruby O’Grady, Christian Schudoma, Wilfried Haerty, and Ksenia V. Krasileva. 2020. “Convergent Loss of an EDS1/PAD4 Signaling Pathway in Several Plant Lineages Reveals Co-Evolved Components of Plant Immunity and Drought Response.” *The Plant Cell*, 32(7): 2158-2177.

Bakker, Erica G., Christopher Toomajian, Martin Kreitman, and Joy Bergelson. 2006. “A Genome-Wide Survey of R Gene Polymorphisms in Arabidopsis.” *The Plant Cell* 18 (8): 1803–18.

Barragan, Cristina A., Rui Wu, Sang-Tae Kim, Wanyan Xi, Anette Habring, Jörg Hagmann, Anna-Lena Van de Weyer, et al. 2019. “RPW8/HR Repeats Control NLR Activation in Arabidopsis Thaliana.” *PLoS Genetics* 15 (7): e1008313.

Bent, A. F., B. N. Kunkel, D. Dahlbeck, K. L. Brown, R. Schmidt, J. Giraudat, J. Leung, and B. J. Staskawicz. 1994. “RPS2 of Arabidopsis Thaliana: A Leucine-Rich Repeat Class of Plant Disease Resistance Genes.” *Science* 265 (5180): 1856–60.

Bentham, Adam, Hayden Burdett, Peter A. Anderson, Simon J. Williams, and Bostjan Kobe. 2017. “Animal NLRs Provide Structural Insights into Plant NLR Function.” *Annals of Botany* 119 (5): 827–702.

Bentham, Adam R., Rafal Zdrzalek, Juan Carlos De la Concepcion, and Mark J. Banfield. 2018. “Uncoiling CNLs: Structure/Function Approaches to Understanding CC Domain Function in Plant NLRs.” *Plant & Cell Physiology* 59 (12): 2398–2408.

Berkey, Robert, Yi Zhang, Xianfeng Ma, Harlan King, Qiong Zhang, Wenming Wang, and Shunyuan Xiao. 2017. “Homologues of the RPW8 Resistance Protein Are Localized to the Extrahaustorial Membrane That Is Likely Synthesized De Novo.” *Plant Physiology* 173 (1): 600–613.

Biezen, Erik A. van der, Cecilie T. Freddie, Katherine Kahn, Jane E. Parker, and Jonathan D. G. Jones. 2002. “Arabidopsis RPP4 Is a Member of the RPP5 Multigene Family of TIR-NB-LRR Genes and Confers Downy Mildew Resistance through Multiple Signalling Components.” *The Plant Journal: For Cell and Molecular Biology* 29 (4): 439–51.

Boller, Thomas, and Georg Felix. 2009. “A Renaissance of Elicitors: Perception of Microbe-Associated Molecular Patterns and Danger Signals by Pattern-Recognition Receptors.” *Annual Review of Plant Biology* 60: 379–406.

Bomblies, Kirsten, Janne Lempe, Petra Epple, Norman Warthmann, Christa Lanz, Jeffery L. Dangl, and Detlef Weigel. 2007. “Autoimmune Response as a Mechanism for a Dobzhansky-Muller-Type Incompatibility Syndrome in Plants.” *PLoS Biology* 5 (9): e236.

Bonardi, Vera, Saijun Tang, Anna Stallmann, Melinda Roberts, Karen Cherkis, and Jeffery L. Dangl. 2011. “Expanded Functions for a Family of Plant Intracellular Immune Receptors beyond Specific Recognition of Pathogen Effectors.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (39): 16463–68.

Botella, M. A., J. E. Parker, L. N. Frost, P. D. Bittner-Eddy, J. L. Beynon, M. J. Daniels, E. B. Holub, and J. D. Jones. 1998. “Three Genes of the Arabidopsis RPP1 Complex Resistance Locus Recognize Distinct Peronospora Parasitica Avirulence Determinants.” *The Plant Cell* 10 (11): 1847–60.

Broz, Petr, and Vishva M. Dixit. 2016. “Inflammasomes: Mechanism of Assembly, Regulation and Signalling.” *Nature Reviews. Immunology* 16 (7): 407–20.

Brunner, Susanne, Severine Hurni, Philipp Streckeisen, Gabriele Mayr, Mario Albrecht, Nabila Yahiaoui, and Beat Keller. 2010. “Intragenic Allele Pyramiding Combines Different Specificities of Wheat Pm3 Resistance Alleles.” *The Plant Journal: For Cell and Molecular Biology* 64 (3): 433–45.

Caicedo, A. L., B. A. Schaal, and B. N. Kunkel. 1999. “Diversity and Molecular Evolution of the RPS2 Resistance Gene in Arabidopsis Thaliana.” *Proceedings of the National Academy of Sciences of the United States of America* 96 (1): 302–6.

Caldwell, Ralph M., and L. E. Compton. 1943. “Complementary Lethal Genes in Wheat Causing a Progressive Lethal Necrosis of Seedlings.” *The Journal of Heredity* 34 (3): 67–70.

Calvo-Baltanás, Vanesa, Jinge Wang, and Eunyoung Chae. 2020. “Hybrid Incompatibility of the Plant Immune System: An Opposite Force to Heterosis Equilibrating Hybrid Performances.”

Cao, Jun, Korbinian Schneeberger, Stephan Ossowski, Torsten Günther, Sebastian Bender, Joffrey Fitz, Daniel Koenig, et al. 2011. “Whole-Genome Sequencing of Multiple Arabidopsis Thaliana Populations.” *Nature Genetics* 43 (10): 956–63.

Cao, Ke, Zhen Peng, Xing Zhen, Yong Li, Kuozhan Liu, Pere Arus, Gengrui Zhu, et al. 2020. “Pan-Genome Analyses of Peach and Its Wild Relatives Provide Insights into the Genetics of Disease Resistance and Species Adaptation.” BioRxiv https://doi.org/10.1101/2020.07.13.200204.

Cesari, Stella. 2018. “Multiple Strategies for Pathogen Perception by Plant Immune Receptors.” *The New Phytologist* 219 (1): 17–24.
Chae, Eunyoung, Kirsten Bomblies, Sang-Tae Kim, Darya Karelina, Maricris Zaidem, Stephan Ossowski, Carmen Martín-Pizarro, et al. 2014. “Species-Wide Genetic Incompatibility Analysis Identifies Immune Genes as Hot Spots of Deleterious Epistasis.” *Cell* 159 (6): 1341–51.
Charlesworth, D., and B. Charlesworth. 1987. “Inbreeding Depression and Its Evolutionary Consequences.” *Annual Review of Ecology and Systematics* 18: 237–68.
Charlesworth, Deborah, and John H. Willis. 2009. “The Genetics of Inbreeding Depression.” *Nature Reviews. Genetics* 10 (11): 783–96.
Chen, Chen, Hao Chen, You-Shun Lin, Jin-Bo Shen, Jun-Xiang Shan, Peng Qi, Min Shi, et al. 2014. “A Two-Locus Interaction Causes Interspecific Hybrid Weakness in Rice.” *Nature Communications* 5: 3357.
Chisholm, Stephen T., Gitta Coaker, Brad Day, and Brian J. Staskawicz. 2006. “Host-Microbe Interactions: Shaping the Evolution of the Plant Immune Response.” *Cell* 124 (4): 803–14.
Choi, Kyuha, Carsten Reinhard, Heïdi Serra, Piotr A. Ziolkowski, Charles J. Underwood, Xiaohui Zhao, Thomas J. Hardcastle, et al. 2016. “Recombination Rate Heterogeneity within Arabidopsis Disease Resistance Genes.” *PLoS Genetics* 12 (7): e1006179.
Choi, K., X. Zhao, K. A. Kelly, O. Venn, J. D. Higgins, N. E. Yelina, T. J. Hardcastle, et al. 2013. “Arabidopsis Meiotic Crossover Hot Spots Overlap with H2A.Z Nucleosomes at Gene Promoters.” *Nature Genetics* 45 (11): 1327–36.
Collier, Sarah M., Louis-Philippe Hamel, and Peter Moffett. 2011. “Cell Death Mediated by the N-Terminal Domains of a Unique and Highly Conserved Class of NB-LRR Protein.” *Molecular Plant-Microbe Interactions: MPMI* 24 (8): 918–31.
Cooley, M. B., S. Pathirana, H. J. Wu, P. Kachroo, and D. F. Klessig. 2000. “Members of the Arabidopsis HRT/RPP8 Family of Resistance Genes Confer Resistance to Both Viral and Oomycete Pathogens.” *The Plant Cell* 12 (5): 663–76.
Dangl, Jeffery L., Diana M. Horvath, and Brian J. Staskawicz. 2013. “Pivoting the Plant Immune System from Dissection to Deployment.” *Science* 341 (6147): 746–51.
Dangl, J. L., and J. D. Jones. 2001. “Plant Pathogens and Integrated Defence Responses to Infection.” *Nature* 411 (6839): 826–33.
Daskalov, Asen, Birgit Habenstein, Raimon Sabaté, Mélanie Berbon, Denis Martinez, Stéphane Chaignepain, Bénédicte Coulary-Salin, Kay Hofmann, Antoine Loquet, and Sven J. Saupe. 2016. “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 of the United States of America* 113 (10): 2720–25.
Deng, Jieqiong, Lei Fang, Xiefei Zhu, Baoliang Zhou, and Tianzhen Zhang. 2019. “A CC-NBS-LRR Gene Induces Hybrid Lethality in Cotton.” *Journal of Experimental Botany* 70 (19): 5145–56.
Devi, S. J. S. Rama, Kuldeep Singh, B. Umakanth, B. Vishalakshi, P. Renuka, K. Vijay Sudhakar, M. S. Prasad, B. C. Viraktamath, V. Ravindra Babu, and M. S. Madhav. 2015. “Development and Identification of Novel Rice Blast Resistant Sources and Their Characterization Using Molecular Markers.” *Rice Science* 22 (6): 300–308.
Duxbury, Zane, Shanshan Wang, Craig I. MacKenzie, Jeannette L. Tenthorey, Xiaoxiao Zhang, Sung Un Huh, Lanxi Hu, et al. 2020. “Induced Proximity of a TIR Signaling Domain on a Plant-Mammalian NLR Chimera Activates Defense in Plants.” *Proceedings of the National Academy of Sciences of the United States of America* 117 (31): 18832–39.
Dyrka, Witold, Marina Lamacchia, Pascal Durrens, Bostjan Kobe, Asen Daskalov, Matthieu

Paoletti, David J. Sherman, and Sven J. Saupe. 2014. “Diversity and Variability of NOD-like Receptors in Fungi.” *Genome Biology and Evolution* 6 (12): 3137–58.
Essuman, Kow, Daniel W. Summers, Yo Sasaki, Xianrong Mao, Aaron DiAntonio, and Jeffrey Milbrandt. 2017. “The SARM1 Toll/Interleukin-1 Receptor Domain Possesses Intrinsic NAD+ Cleavage Activity That Promotes Pathological Axonal Degeneration.” *Neuron* 93 (6): 1334–43.e5.
Feehan, Joanna M., Baptiste Castel, Adam R. Bentham, and Jonathan Dg Jones. 2020. “Plant NLRs Get by with a Little Help from Their Friends.” *Current Opinion in Plant Biology* 56: 99–108.
Flor, H. H. 1971. “Current Status of the Gene for Gene Hypothesis.” *Annual Review of Phytopathology* 9: 275–95.
Gao, Yuxia, Wenqiang Wang, Tian Zhang, Zhen Gong, Huayao Zhao, and Guan-Zhu Han. 2018. “Out of Water: The Origin and Early Diversification of Plant R-Genes.” *Plant Physiology* 177 (1): 82–89.
Ghislain, Marc, Arinaitwe Abel Byarugaba, Eric Magembe, Anne Njoroge, Cristina Rivera, María Lupe Román, José Carlos Tovar, et al. 2019. “Stacking Three Late Blight Resistance Genes from Wild Species Directly into African Highland Potato Varieties Confers Complete Field Resistance to Local Blight Races.” *Plant Biotechnology Journal* 17 (6): 1119–29.
Golicz, Agnieszka A., Philipp E. Bayer, Guy C. Barker, Patrick P. Edger, Hyeran Kim, Paula A. Martinez, Chon Kit Kenneth Chan, et al. 2016. “The Pangenome of an Agronomically Important Crop Plant Brassica Oleracea.” *Nature Communications* 6:13390.
Gos, Gesseca, Tanja Slotte, and Stephen I. Wright. 2012. “Signatures of Balancing Selection Are Maintained at Disease Resistance Loci Following Mating System Evolution and a Population Bottleneck in the Genus Capsella.” *BMC Evolutionary Biology* 12:152.
Haberer, Georg, Nadia Kamal, Eva Bauer, Heidrun Gundlach, Iris Fischer, Michael A. Seidel, Manuel Spannagl, et al. 2020. “European Maize Genomes Highlight Intraspecies Variation in Repeat and Gene Content.” *Nature Genetics* 52 (9) 950-957.
Hamada, Mayuko, Eiichi Shoguchi, Chuya Shinzato, Takeshi Kawashima, David J. Miller, and Nori Satoh. 2013. “The Complex NOD-like Receptor Repertoire of the Coral Acropora Digitifera Includes Novel Domain Combinations.” *Molecular Biology and Evolution* 30 (1): 167–76.
Hanson, Peter, Shu-Fen Lu, Jaw-Fen Wang, Wallace Chen, Lawrence Kenyon, Chee-Wee Tan, Kwee Lian Tee, et al. 2016. “Conventional and Molecular Marker-Assisted Selection and Pyramiding of Genes for Multiple Disease Resistance in Tomato.” *Scientia Horticulturae* 201: 346–54.
Hatsugai, Noriyuki, Daisuke Igarashi, Keisuke Mase, You Lu, Yayoi Tsuda, Suma Chakravarthy, Hai-Lei Wei, et al. 2017. “A Plant Effector-Triggered Immunity Signaling Sector Is Inhibited by Pattern-Triggered Immunity.” *The EMBO Journal* 36 (18): 2758–69.
Henk, A. D., R. F. Warren, and R. W. Innes. 1999. “A New Ac-like Transposon of Arabidopsis Is Associated with a Deletion of the RPS5 Disease Resistance Gene.” *Genetics* 151 (4): 1581–89.
Hibino, Taku, Mariano Loza-Coll, Cynthia Messier, Audrey J. Majeske, Avis H. Cohen, David P. Terwilliger, Katherine M. Buckley, et al. 2006. “The Immune Gene Repertoire Encoded in the Purple Sea Urchin Genome.” *Developmental Biology* 300 (1): 349–65.
Hill, W. G., and A. Robertson. 1966. “The Effect of Linkage on Limits to Artificial Selection.” *Genetical Research* 8 (3): 269–94.
Hoorn, Renier A. L. van der, and Sophien Kamoun. 2008. “From Guard to Decoy: A New Model for Perception of Plant Pathogen Effectors.” *The Plant Cell.* 20 (8): 2009–17.

Horsefield, Shane, Hayden Burdett, Xiaoxiao Zhang, Mohammad K. Manik, Yun Shi, Jian Chen, Tiancong Qi, et al. 2019. “NAD+ Cleavage Activity by Animal and Plant TIR Domains in Cell Death Pathways.” *Science* 365 (6455): 793–99.
Horton, M., A. M. Hancock, Y. S. Huang, C. Toomajian, S. Atwell, A. Auton, W. Muliyati, et al. 2012. “Genome-Wide Pattern of Genetic Variation in Worldwide Arabidopsis Thaliana Accessions from the RegMap Panel.” *Nature Genetics* 44 (2): 212–16.
Huang, Xuehui, Shihua Yang, Junyi Gong, Yan Zhao, Qi Feng, Hao Gong, Wenjun Li, et al. 2015. “Genomic Analysis of Hybrid Rice Varieties Reveals Numerous Superior Alleles That Contribute to Heterosis.” *Nature Communications* 6 (2): 6258.
Hurgobin, Bhavna, Agnieszka A. Golicz, Philipp E. Bayer, Chon-Kit Kenneth Chan, Soodeh Tirnaz, Aria Dolatabadian, Sarah V. Schiessl, et al. 2018. “Homoeologous Exchange Is a Major Cause of Gene Presence/absence Variation in the Amphidiploid Brassica Napus.” *Plant Biotechnology Journal* 16 (7): 1265–74.
Hurni, Severine, Susanne Brunner, Daniel Stirnweis, Gerhard Herren, David Peditto, Robert A. McIntosh, and Beat Keller. 2014. “The Powdery Mildew Resistance Gene Pm8 Derived from Rye Is Suppressed by Its Wheat Ortholog Pm3.” *The Plant Journal: For Cell and Molecular Biology* 79 (6): 904–13.
Jeuken, Marieke J. W., Ningwen W. Zhang, Leah K. McHale, Koen Pelgrom, Erik den Boer, Pim Lindhout, Richard W. Michelmore, Richard G. F. Visser, and Rients E. Niks. 2009. “Rin4 Causes Hybrid Necrosis and Race-Specific Resistance in an Interspecific Lettuce Hybrid.” *The Plant Cell* 21 (10): 3368–78.
Jiang, Xueyuan, and Raquel Assis. 2017. “Natural Selection Drives Rapid Functional Evolution of Young Drosophila Duplicate Genes.” *Molecular Biology and Evolution* 34 (12): 3089–98.
Jiao, Wen-Biao, and Korbinian Schneeberger. 2020. “Chromosome-Level Assemblies of Multiple Arabidopsis Genomes Reveal Hotspots of Rearrangements with Altered Evolutionary Dynamics.” *Nature Communications* 11 (1): 989.
Jia, Yanxiao, Yang Yuan, Yanchun Zhang, Sihai Yang, and Xiaohui Zhang. 2015. “Extreme Expansion of NBS-Encoding Genes in Rosaceae.” *BMC Genetics* 16 (May): 48.
Johal, G. S., and S. P. Briggs. 1992. “Reductase Activity Encoded by the HM1 Disease Resistance Gene in Maize.” *Science* 258 (5084): 985–87.
Jones, Jonathan D. G., and Jeffery L. Dangl. 2006. “The Plant Immune System.” *Nature* 444 (7117): 323–29.
Jubic, Lance M., Svenja Saile, Oliver J. Furzer, Farid El Kasmi, and Jeffery L. Dangl. 2019. “Help Wanted: Helper NLRs and Plant Immune Responses.” *Current Opinion in Plant Biology* 50: 82–94.
Karasov, Talia L., Joel M. Kniskern, Liping Gao, Brody J. DeYoung, Jing Ding, Ullrich Dubiella, Ruben O. Lastra, et al. 2014. “The Long-Term Maintenance of a Resistance Polymorphism through Diffuse Interactions.” *Nature* 512 (7515): 436–40.
Keller, Lukas F., and Donald M. Waller. 2002. “Inbreeding Effects in Wild Populations.” *Trends in Ecology & Evolution* 17 (5): 230–41.
Koenig, Daniel, Jörg Hagmann, Rachel Li, Felix Bemm, Tanja Slotte, Barbara Neuffer, Stephen I. Wright, and Detlef Weigel. 2019. “Long-Term Balancing Selection Drives Evolution of Immunity Genes in Capsella.” *eLife* 8.43606.
Kourelis, Jiorgos, and Renier A. L. van der Hoorn. 2018. “Defended to the Nines: 25 Years of Resistance Gene Cloning Identifies Nine Mechanisms for R Protein Function.” *The Plant Cell* 30 (2): 285–99.
Krasileva, K. V., C. Zheng, L. Leonelli, S. Goritschnig, D. Dahlbeck, and B. J. Staskawicz. 2011. “Global Analysis of Arabidopsis/downy Mildew Interactions Reveals Prevalence of

Incomplete Resistance and Rapid Evolution of Pathogen Recognition." *PloS One* 6 (12): e28765.
Krüger, Julia, Colwyn M. Thomas, Catherine Golstein, Mark S. Dixon, Matthew Smoker, Saijun Tang, Lonneke Mulder, and Jonathan D. G. Jones. 2002. "A Tomato Cysteine Protease Required for Cf-2-Dependent Disease Resistance and Suppression of Autonecrosis." *Science* 296 (5568): 744–47.
Kuang, Hanhui, Herman J. van Eck, Delphine Sicard, Richard Michelmore, and Eviatar Nevo. 2008. "Evolution and Genetic Population Structure of Prickly Lettuce (Lactuca Serriola) and Its RGC2 Resistance Gene Cluster." *Genetics* 178 (3): 1547–58.
Kuang, H., S. S. Woo, B. C. Meyers, E. Nevo, and R. W. Michelmore. 2004. "Multiple Genetic Processes Result in Heterogeneous Rates of Evolution within the Major Cluster Disease Resistance Genes in Lettuce." *The Plant Cell* 16 (11): 2870–94.
Lee, Rachelle R. Q., and Eunyoung Chae. 2020. "Variation Patterns of NLR Clusters in Arabidopsis Thaliana Genomes." *Plant Communications*, June, 100089.
Li, Lei, Anette Habring, Kai Wang, and Detlef Weigel. 2020. "Atypical Resistance Protein RPW8/HR Triggers Oligomerization of the NLR Immune Receptor RPP7 and Autoimmunity." *Cell Host & Microbe*, 27(3) 405-417.
Lin, Xiao, Yu Zhang, Hanhui Kuang, and Jiongjiong Chen. 2013. "Frequent Loss of Lineages and Deficient Duplications Accounted for Low Copy Number of Disease Resistance Genes in Cucurbitaceae." *BMC Genomics* 14:335.
Liu, F., M. King, C. Toth, E. Nogales, and B. J. Staskawicz. 2020. "Structure of the Activated Roq1 Resistosome Directly Recognizing the Pathogen Effector XopQ." *bioRxiv*. https://www.biorxiv.org/content/10.1101/2020.08.13.246413v1.
Liu, Jinyan, Bo Min Kim, Yo-Hei Kaneko, Tsuyoshi Inukai, and Chikara Masuta. 2015. "Identification of the TuNI Gene Causing Systemic Necrosis in Arabidopsis Ecotype Ler Infected with Turnip Mosaic Virus and Characterization of Its Expression." *Journal of General Plant Pathology: JGPP* 81 (3): 180–91.
Liu, Yucheng, Huilong Du, Pengcheng Li, Yanting Shen, Hua Peng, Shulin Liu, Guo-An Zhou, et al. 2020. "Pan-Genome of Wild and Cultivated Soybeans." *Cell* 182 (1) 162-176.
MacQueen, Alice, Dacheng Tian, Wenhan Chang, Eric Holub, Martin Kreitman, and Joy Bergelson. 2019. "Population Genetics of the Highly Polymorphic RPP8 Gene Family." *Genes* 10 (9) 691.
Marchal, Clemence, Jianping Zhang, Peng Zhang, Paul Fenwick, Burkhard Steuernagel, Nikolai M. Adamski, Lesley Boyd, et al. 2018. "BED-Domain-Containing Immune Receptors Confer Diverse Resistance Spectra to Yellow Rust." *Nature Plants* 4 (9): 662–68.
Matzaraki, Vasiliki, Vinod Kumar, Cisca Wijmenga, and Alexandra Zhernakova. 2017. "The MHC Locus and Genetic Susceptibility to Autoimmune and Infectious Diseases." *Genome Biology* 18 (1): 76.
Mauricio, R., E. A. Stahl, T. Korves, D. Tian, M. Kreitman, and J. Bergelson. 2003. "Natural Selection for Polymorphism in the Disease Resistance Gene RPS2 of Arabidopsis Thaliana." *Genetics* 163 (2): 735–46.
Ma, Xian-Feng, Yan Li, Jin-Long Sun, Ting-Ting Wang, Jing Fan, Yang Lei, Yan-Yan Huang, et al. 2014. "Ectopic Expression of RESISTANCE TO POWDERY MILDEW8.1 Confers Resistance to Fungal and Oomycete Pathogens in Arabidopsis." *Plant & Cell Physiology* 55 (8): 1484–96.
McDowell, J. M., M. Dhandaydham, T. A. Long, M. G. Aarts, S. Goff, E. B. Holub, and J. L. Dangl. 1998. "Intragenic Recombination and Diversifying Selection Contribute to the Evolution of Downy Mildew Resistance at the RPP8 Locus of Arabidopsis." *The Plant Cell*

10 (11): 1861–74.
Meyers, B. C., D. B. Chin, K. A. Shen, S. Sivaramakrishnan, D. O. Lavelle, Z. Zhang, and R. W. Michelmore. 1998. “The Major Resistance Gene Cluster in Lettuce Is Highly Duplicated and Spans Several Megabases.” *The Plant Cell* 10 (11): 1817–32.
Meyers, Blake C., Alexander Kozik, Alyssa Griego, Hanhui Kuang, and Richard W. Michelmore. 2003. “Genome-Wide Analysis of NBS-LRR-Encoding Genes in Arabidopsis.” *The Plant Cell* 15 (4): 809–34.
Michael, T. P., E. Ernst, N. Hartwick, P. Chu, and D. Bryant. 2020. “Genome and Time-of-Day Transcriptome of Wolffia Australiana Link Morphological Extreme Minimization with Un-Gated Plant Growth.” *bioRxiv*. https://www.biorxiv.org/content/10.1101/2020.03.31.018291
Michelmore, R. W., and B. C. Meyers. 1998. “Clusters of Resistance Genes in Plants Evolve by Divergent Selection and a Birth-and-Death Process.” *Genome Research* 8 (11): 1113–30.
Mindrinos, M., F. Katagiri, G. L. Yu, and F. M. Ausubel. 1994. “The A. Thaliana Disease Resistance Gene RPS2 Encodes a Protein Containing a Nucleotide-Binding Site and Leucine-Rich Repeats.” *Cell* 78 (6): 1089–99.
Montenegro, Juan D., Agnieszka A. Golicz, Philipp E. Bayer, Bhavna Hurgobin, Hueytyng Lee, Chon-Kit Kenneth Chan, Paul Visendi, et al. 2017. “The Pangenome of Hexaploid Bread Wheat.” *The Plant Journal: For Cell and Molecular Biology* 90 (5): 1007–13.
Munch, David, Vikas Gupta, Asger Bachmann, Wolfgang Busch, Simon Kelly, Terry Mun, and Stig Uggerhøj Andersen. 2018. “The Brassicaceae Family Displays Divergent, Shoot-Skewed NLR Resistance Gene Expression.” *Plant Physiology* 176 (2): 1598–1609.
Ngou, B. P. M., H. K. Ahn, P. Ding, and J. D. G. Jones. 2020. “Mutual Potentiation of Plant Immunity by Cell-Surface and Intracellular Receptors.” *bioRxiv*. https://www.biorxiv.org/content/10.1101/2020.04.10.034173
Noël, L., T. L. Moores, E. A. van Der Biezen, M. Parniske, M. J. Daniels, J. E. Parker, and J. D. Jones. 1999. “Pronounced Intraspecific Haplotype Divergence at the RPP5 Complex Disease Resistance Locus of Arabidopsis.” *The Plant Cell* 11 (11): 2099–2112.
Ohno, Susumu. 1970. *Evolution by Gene Duplication*. Springer, Berlin, Heidelberg.
Orr, H. A. 1996. “Dobzhansky, Bateson, and the Genetics of Speciation.” *Genetics* 144 (4): 1331–35.
Penn, Dustin J., Kristy Damjanovich, and Wayne K. Potts. 2002. “MHC Heterozygosity Confers a Selective Advantage against Multiple-Strain Infections.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (17): 11260–64.
Prigozhin, Daniil M., and Ksenia V. Krasileva. “Intraspecies Diversity Reveals a Subset of Highly Variable Plant Immune Receptors and Predicts Their Binding Sites.” bioRxiv https://www.biorxiv.org/content/10.1101/2020.07.10.190785v.
Radwan, Jacek, Wiesław Babik, Jim Kaufman, Tobias L. Lenz, and Jamie Winternitz. 2020. “Advances in the Evolutionary Understanding of MHC Polymorphism.” *Trends in Genetics: TIG* 36 (4): 298–311.
Rowan, Beth A., Darren Heavens, Tatiana R. Feuerborn, Andrew J. Tock, Ian R. Henderson, and Detlef Weigel. 2019. “An Ultra High-Density Arabidopsis Thaliana Crossover Map That Refines the Influences of Structural Variation and Epigenetic Features.” *Genetics* 213 (3) 771-787.
Sáenz-Mata, Jorge, and Juan Francisco Jiménez-Bremont. 2012. “HR4 Gene Is Induced in the Arabidopsis-Trichoderma Atroviride Beneficial Interaction.” *International Journal of Molecular Sciences* 13 (7): 9110–28.
Sandstedt, Gabrielle D., Carrie A. Wu, and Andrea L. Sweigart. 2020. “Evolution of Multiple

Postzygotic Barriers between Species in the Mimulus Tilingii Species Complex." bioRvix. https://doi.org/10.1101/2020.08.07.241489.
Seo, Eunyoung, Seungill Kim, Seon-In Yeom, and Doil Choi. 2016. "Genome-Wide Comparative Analyses Reveal the Dynamic Evolution of Nucleotide-Binding Leucine-Rich Repeat Gene Family among Solanaceae Plants." *Frontiers in Plant Science* 7: 1205.
Seong, Kyungyong, Eunyoung Seo, Kamil Witek, Meng Li, and Brian Staskawicz. 2020. "Evolution of NLR Resistance Genes with Noncanonical N-Terminal Domains in Wild Tomato Species." *The New Phytologist*, 227(5) 1530-43.
Seuring, Carolin, Jason Greenwald, Christian Wasmer, Roger Wepf, Sven J. Saupe, Beat H. Meier, and Roland Riek. 2012. "The Mechanism of Toxicity in HET-S/HET-S Prion Incompatibility." *PLoS Biology* 10 (12): e1001451.
Shao, Zhu-Qing, Jia-Yu Xue, Qiang Wang, Bin Wang, and Jian-Qun Chen. 2019. "Revisiting the Origin of Plant NBS-LRR Genes." *Trends in Plant Science* 24 (1): 9–12.
Sicard, Adrien, Christian Kappel, Emily B. Josephs, Young Wha Lee, Cindy Marona, John R. Stinchcombe, Stephen I. Wright, and Michael Lenhard. 2015. "Divergent Sorting of a Balanced Ancestral Polymorphism Underlies the Establishment of Gene-Flow Barriers in Capsella." *Nature Communications* 6: 7960.
Slusarenko, Alan J., and Nikolaus L. Schlaich. 2003. "Downy Mildew of Arabidopsis Thaliana Caused by Hyaloperonospora Parasitica (formerly Peronospora Parasitica)." *Molecular Plant Pathology* 4 (3): 159–70.
Soanes, Darren M., and Nicholas J. Talbot. 2010. "Comparative Genome Analysis Reveals an Absence of Leucine-Rich Repeat Pattern-Recognition Receptor Proteins in the Kingdom Fungi." *PloS One* 5 (9): e12725.
Song, Jia-Ming, Zhilin Guan, Jianlin Hu, Chaocheng Guo, Zhiquan Yang, Shuo Wang, Dongxu Liu, et al. 2020. "Eight High-Quality Genomes Reveal Pan-Genome Architecture and Ecotype Differentiation of Brassica Napus." *Nature Plants* 6 (1): 34–45.
Srichumpa, Payorm, Susanne Brunner, Beat Keller, and Nabila Yahiaoui. 2005. "Allelic Series of Four Powdery Mildew Resistance Genes at the Pm3 Locus in Hexaploid Bread Wheat." *Plant Physiology* 139 (2): 885–95.
Staal, Jens, Maria Kaliff, Svante Bohman, and Christina Dixelius. 2006. "Transgressive Segregation Reveals Two Arabidopsis TIR-NB-LRR Resistance Genes Effective against Leptosphaeria Maculans, Causal Agent of Blackleg Disease." *The Plant Journal: For Cell and Molecular Biology* 46 (2): 218–30.
Stahl, E. A., G. Dwyer, R. Mauricio, M. Kreitman, and J. Bergelson. 1999. "Dynamics of Disease Resistance Polymorphism at the Rpm1 Locus of Arabidopsis." *Nature* 400 (6745): 667–71.
Stirnweis, Daniel, Samira D. Milani, Susanne Brunner, Gerhard Herren, Gabriele Buchmann, David Peditto, Tina Jordan, and Beat Keller. 2014. "Suppression among Alleles Encoding Nucleotide-Binding-Leucine-Rich Repeat Resistance Proteins Interferes with Resistance in F1 Hybrid and Allele-Pyramided Wheat Plants." *The Plant Journal: For Cell and Molecular Biology* 79 (6): 893–903.
Takahashi, Hideki, Jennifer Miller, Yukine Nozaki, Megumi Takeda, Jyoti Shah, Shu Hase, Masato Ikegami, Yoshio Ehara, S. P. Dinesh-Kumar, and Sukamto. 2002. "RCY1, an Arabidopsis Thaliana RPP8/HRT Family Resistance Gene, Conferring Resistance to Cucumber Mosaic Virus Requires Salicylic Acid, Ethylene and a Novel Signal Transduction Mechanism." *The Plant Journal: For Cell and Molecular Biology* 32 (5): 655–67.
Tarr, D. Ellen K., and Helen M. Alexander. 2009. "TIR-NBS-LRR Genes Are Rare in Monocots: Evidence from Diverse Monocot Orders." *BMC Research Notes* 2:197.
Tettelin, Hervé, Vega Masignani, Michael J. Cieslewicz, Claudio Donati, Duccio Medini, Naomi

L. Ward, Samuel V. Angiuoli, et al. 2005. “Genome Analysis of Multiple Pathogenic Isolates of Streptococcus Agalactiae: Implications for the Microbial ‘Pan-Genome.’” *Proceedings of the National Academy of Sciences of the United States of America* 102 (39): 13950–55.
Tian, D., H. Araki, E. Stahl, J. Bergelson, and M. Kreitman. 2002. “Signature of Balancing Selection in Arabidopsis.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (17): 11525–30.
Tian, D., M. B. Traw, J. Q. Chen, M. Kreitman, and J. Bergelson. 2003. “Fitness Costs of R-Gene-Mediated Resistance in Arabidopsis Thaliana.” *Nature* 423 (6935): 74–77.
Todesco, Marco, Sang-Tae Kim, Eunyoung Chae, Kirsten Bomblies, Maricris Zaidem, Lisa M. Smith, Detlef Weigel, and Roosa A. E. Laitinen. 2014. “Activation of the Arabidopsis Thaliana Immune System by Combinations of Common ACD6 Alleles.” *PLoS Genetics* 10 (7): e1004459.
Toruño, Tania Y., Mingzhe Shen, Gitta Coaker, and David Mackey. 2019. “Regulated Disorder: Posttranslational Modifications Control the RIN4 Plant Immune Signaling Hub.” *Molecular Plant-Microbe Interactions: MPMI* 32 (1): 56–64.
Trowsdale, John, and Julian C. Knight. 2013. “Major Histocompatibility Complex Genomics and Human Disease.” *Annual Review of Genomics and Human Genetics* 14: 301–23.
Uehling, Jessie, Aurélie Deveau, and Mathieu Paoletti. 2017. “Do Fungi Have an Innate Immune Response? An NLR-Based Comparison to Plant and Animal Immune Systems.” *PLoS Pathogens* 13 (10): e1006578.
Urbach, Jonathan M., and Frederick M. Ausubel. 2017. “The NBS-LRR Architectures of Plant R-Proteins and Metazoan NLRs Evolved in Independent Events.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (5): 1063–68.
Van de Weyer, Anna-Lena, Freddy Monteiro, Oliver J. Furzer, Marc T. Nishimura, Volkan Cevik, Kamil Witek, Jonathan D. G. Jones, Jeffery L. Dangl, Detlef Weigel, and Felix Bemm. 2019. “A Species-Wide Inventory of NLR Genes and Alleles in Arabidopsis Thaliana.” *Cell* 178 (5): 1260–72.e14.
Van Ghelder, Cyril, Geneviève J. Parent, Philippe Rigault, Julien Prunier, Isabelle Giguère, Sébastien Caron, Juliana Stival Sena, et al. 2019. “The Large Repertoire of Conifer NLR Resistance Genes Includes Drought Responsive and Highly Diversified RNLs.” *Scientific Reports* 9 (1): 11614.
Wang, Guoxun, Brice Roux, Feng Feng, Endrick Guy, Lin Li, Nannan Li, Xiaojuan Zhang, 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 (3): 285–95.
Wang, Huayi, Liming Sun, Lijing Su, Josep Rizo, Lei Liu, Li-Feng Wang, Fu-Sheng Wang, and Xiaodong Wang. 2014. “Mixed Lineage Kinase Domain-like Protein MLKL Causes Necrotic Membrane Disruption upon Phosphorylation by RIP3.” *Molecular Cell* 54 (1): 133–46.
Wang, Jizong, Meijuan Hu, Jia Wang, Jinfeng Qi, Zhifu Han, Guoxun Wang, Yijun Qi, Hong-Wei Wang, Jian-Min Zhou, and Jijie Chai. 2019. “Reconstitution and Structure of a Plant NLR Resistosome Conferring Immunity.” *Science* 364 (6435).
Wang, J. Z., J. Wang, M. J. Hu, H. W. Wang, J. M. Zhou, and J. J. Chai. 2019. “Ligand-Triggered Allosteric ADP Release Primes a Plant NLR Complex.” *Science* 364 (6435).
Wang, Long, Lina Zhao, Xiaohui Zhang, Qijun Zhang, Yanxiao Jia, Guan Wang, Simin Li, Dacheng Tian, Wen-Hsiung Li, and Sihai Yang. 2019. “Large-Scale Identification and Functional Analysis of NLR Genes in Blast Resistance in the Tetep Rice Genome Sequence.” *Proceedings of the National Academy of Sciences of the United States of*

America 116 (37): 18479–87.
Wang, Wenming, Alessandra Devoto, John G. Turner, and Shunyuan Xiao. 2007. “Expression of the Membrane-Associated Resistance Protein RPW8 Enhances Basal Defense against Biotrophic Pathogens.” *Molecular Plant-Microbe Interactions: MPMI* 20 (8): 966–76.
Wan, Li, Kow Essuman, Ryan G. Anderson, Yo Sasaki, Freddy Monteiro, Eui-Hwan Chung, Erin Osborne Nishimura, et al. 2019. “TIR Domains of Plant Immune Receptors Are NAD+-Cleaving Enzymes That Promote Cell Death.” *Science* 365 (6455): 799–803.
Whitham, S., S. P. Dinesh-Kumar, D. Choi, R. Hehl, C. Corr, and B. Baker. 1994. “The Product of the Tobacco Mosaic Virus Resistance Gene N: Similarity to Toll and the Interleukin-1 Receptor.” *Cell* 78 (6): 1101–15.
Wicker, Thomas, Nabila Yahiaoui, and Beat Keller. 2007. “Illegitimate Recombination Is a Major Evolutionary Mechanism for Initiating Size Variation in Plant Resistance Genes.” *The Plant Journal: For Cell and Molecular Biology* 51 (4): 631–41.
Williams, Simon J., Kee Hoon Sohn, Li Wan, Maud Bernoux, Panagiotis F. Sarris, Cecile Segonzac, Thomas Ve, et al. 2014. “Structural Basis for Assembly and Function of a Heterodimeric Plant Immune Receptor.” *Science* 344 (6181): 299–303.
Wu, Chih-Hang, Ahmed Abd-El-Haliem, Tolga O. Bozkurt, Khaoula Belhaj, Ryohei Terauchi, Jack H. Vossen, and Sophien Kamoun. 2017. “NLR Network Mediates Immunity to Diverse Plant Pathogens.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (30): 8113–18.
Xiao, Ning, Yunyu Wu, Cunhong Pan, Ling Yu, Yu Chen, Guangqing Liu, Yuhong Li, et al. 2016. “Improving of Rice Blast Resistances in Japonica by Pyramiding Major R Genes.” *Frontiers in Plant Science* 7: 1918.
Xiao, S., S. Ellwood, O. Calis, E. Patrick, T. Li, M. Coleman, and J. G. Turner. 2001. “Broad-Spectrum Mildew Resistance in Arabidopsis Thaliana Mediated by RPW8.” *Science* 291 (5501): 118–20.
Xiao, S., B. C. Emerson, K. Ratanasut, E. Patrick, C. O’Neil, I. Bancroft, and J. G. Turner. 2004. “Origin and Maintenance of a Broad-Spectrum Disease Resistance Locus in Arabidopsis.” *Molecular Biology and Evolution* 21: 12.
Yamamoto, Eiji, Tomonori Takashi, Yoichi Morinaka, Shaoyang Lin, Jianzhong Wu, Takashi Matsumoto, Hidemi Kitano, Makoto Matsuoka, and Motoyuki Ashikari. 2010. “Gain of Deleterious Function Causes an Autoimmune Response and Bateson-Dobzhansky-Muller Incompatibility in Rice.” *Molecular Genetics and Genomics: MGG* 283 (4): 305–15.
Yamanaka, Naoki, and Md M. Hossain. 2019. “Pyramiding Three Rust-resistance Genes Confers a High Level of Resistance in Soybean (Glycine Max).” Edited by Ram Singh. *Plant Breeding = Zeitschrift Fur Pflanzenzuchtung* 138 (6): 686–95.
Yang, S., J. Li, X. Zhang, Q. Zhang, J. Huang, J. Q. Chen, D. L. Hartl, and D. Tian. 2013. “Rapidly Evolving R Genes in Diverse Grass Species Confer Resistance to Rice Blast Disease.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (46): 18572–77.
Yuan, M., Z. Jiang, G. Bi, K. Nomura, M. Liu, and S. Y. He. 2020. “Pattern-Recognition Receptors Are Required for NLR-Mediated Plant Immunity.” *bioRxiv*. https://www.biorxiv.org/content/10.1101/2020.04.10.031294v1.abstract.
Yue, Jia-Xing, Blake C. Meyers, Jian-Qun Chen, Dacheng Tian, and Sihai Yang. 2012. “Tracing the Origin and Evolutionary History of Plant Nucleotide-Binding Site-Leucine-Rich Repeat (NBS-LRR) Genes.” *The New Phytologist* 193 (4): 1049–63.
Zhang, Xiaohui, Sihai Yang, Jiao Wang, Yanxiao Jia, Ju Huang, Shengjun Tan, Yan Zhong, et al. 2015. “A Genome-Wide Survey Reveals Abundant Rice Blast R Genes in Resistant

Cultivars." *The Plant Journal: For Cell and Molecular Biology* 84 (1): 20–28.
Zhong, Yan, and Zong-Ming Max Cheng. 2016. "A Unique RPW8-Encoding Class of Genes That Originated in Early Land Plants and Evolved through Domain Fission, Fusion, and Duplication." *Scientific Reports* 6: 32923.
Zhong, Yifei, Anna Kinio, and Maya Saleh. 2013. "Functions of NOD-Like Receptors in Human Diseases." *Frontiers in Immunology* 4: 333.
Zhou, Yong, Dmytro Chebotarov, Dave Kudrna, Victor Llaca, Seunghee Lee, Shanmugam Rajasekar, Nahed Mohammed, et al. 2020. "A Platinum Standard Pan-Genome Resource That Represents the Population Structure of Asian Rice." *Scientific Data* 7 (1): 113.

PLOS GENETICS

RESEARCH ARTICLE

RPW8/HR repeats control NLR activation in *Arabidopsis thaliana*

Cristina A. Barragan[1], Rui Wu[1], Sang-Tae Kim[1¤a], Wanyan Xi[1], Anette Habring[1], Jörg Hagmann[1¤b], Anna-Lena Van de Weyer[1], Maricris Zaidem[1¤c], William Wing Ho Ho[1,2], George Wang[1], Ilja Bezrukov[1], Detlef Weigel[1]*, Eunyoung Chae[1,3]*

1 Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany, **2** Melbourne Integrative Genomics, The University of Melbourne, Parkville, Victoria, Australia, **3** Department of Biological Sciences, National University of Singapore, Singapore

¤a Current address: Center for Genome Engineering, Institute for Basic Science, Daejeon, South Korea
¤b Current address: Computomics GmbH, Tübingen, Germany
¤c Current address: Center for Genomics and Systems Biology, New York University, New York, New York, United States of America
* weigel@tue.mpg.de (DW); dbsce@nus.edu.sg (EC)

OPEN ACCESS

Citation: Barragan CA, Wu R, Kim S-T, Xi W, Habring A, Hagmann J, et al. (2019) RPW8/HR repeats control NLR activation in *Arabidopsis thaliana*. PLoS Genet 15(7): e1008313. https://doi.org/10.1371/journal.pgen.1008313

Editor: Gitta Coaker, University of California Davis, UNITED STATES

Received: April 24, 2019

Accepted: July 17, 2019

Published: July 25, 2019

Data Availability Statement: DNA sequences have been deposited with GenBank under accession numbers MK598747 and MK604929-MK604934. All the other relevant data are within the manuscript and its Supporting Information files.

Funding: This study was supported by a Marie Curie postdoctoral fellowship (2014-655295, https://ec.europa.eu/research/mariecurieactions/actions/individual-fellowships_en) to RW, the Academic Research Fund from the National University of Singapore (R-154-000-B33-114, https://www.nrf.gov.sg/) to EC, ERC Advanced

Abstract

In many plant species, conflicts between divergent elements of the immune system, especially nucleotide-binding oligomerization domain-like receptors (NLR), can lead to hybrid necrosis. Here, we report deleterious allele-specific interactions between an NLR and a non-NLR gene cluster, resulting in not one, but multiple hybrid necrosis cases in *Arabidopsis thaliana*. The NLR cluster is *RESISTANCE TO PERONOSPORA PARASITICA 7* (*RPP7*), which can confer strain-specific resistance to oomycetes. The non-NLR cluster is *RESISTANCE TO POWDERY MILDEW 8* (*RPW8*) / *HOMOLOG OF RPW8* (*HR*), which can confer broad-spectrum resistance to both fungi and oomycetes. RPW8/HR proteins contain at the N-terminus a potential transmembrane domain, followed by a specific coiled-coil (CC) domain that is similar to a domain found in pore-forming toxins MLKL and HET-S from mammals and fungi. C-terminal to the CC domain is a variable number of 21- or 14-amino acid repeats, reminiscent of regulatory 21-amino acid repeats in fungal HET-S. The number of repeats in different RPW8/HR proteins along with the sequence of a short C-terminal tail predicts their ability to activate immunity in combination with specific RPP7 partners. Whether a larger or smaller number of repeats is more dangerous depends on the specific RPW8/HR autoimmune risk variant.

Author summary

In many plant species, conflicts between divergent elements of the immune system can cause hybrids to express autoimmunity, a generally deleterious syndrome known as hybrid necrosis. We are investigating multiple hybrid necrosis cases in *Arabidopsis thaliana* that are caused by allele-specific interactions between different variants at two unlinked resistance (R) gene clusters, *RESISTANCE TO PERONOSPORA PARASITICA 7* (*RPP7*) and *RESISTANCE TO POWDERY MILDEW 8* (*RPW8*)/*HOMOLOG OF RPW8*

Grant IMMUNEMESIS (340602, https://erc.europa.eu/funding/advanced-grants), the Deutsche Forschungsgemeinschaft through the Collaborative Research Center (CRC1101, https://www.dfg.de/en/research_funding/programmes/coordinated_programmes/collaborative_research_centres/), and The Max Planck Society (https://www.mpg.de/en) to DW. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

(*HR*). The *RPP7* locus encodes intracellular nucleotide binding site-leucine rich repeat (NLR) immune receptors that can confer strain-specific resistance to oomycetes, while the *RPW8*/*HR* locus encodes atypical resistance proteins, of which some can confer broad-spectrum resistance to filamentous pathogens. There is extensive structural variation in the *RPW8/HR* cluster, both at the level of gene copy number and at the level of C-terminal, 21- or 14-amino acid long RPW8/HR repeats. We demonstrate that the number of RPW8/HR repeats and the short C-terminal tail correlate, in an allele-specific manner, with the severity of hybrid necrosis when these alleles are combined with *RPP7* variants. We discuss these findings in light of sequence similarity between RPW8/HR and pore-forming toxins MLKL and HET-S from mammals and fungi.

Introduction

The combination of divergent parental genomes in hybrids can produce new phenotypes not seen in either parent. At one end of the spectrum is hybrid vigor, with progeny being superior to the parents, while at the other end there is hybrid weakness, with progeny being inferior to the parents, and in the most extreme cases being sterile or unable to survive.

In plants, a particularly conspicuous set of hybrid incompatibilities is associated with autoimmunity, often with substantial negative effects on hybrid fitness [1–3]. Studies of hybrid autoimmunity in several species, often expressed as hybrid necrosis, have revealed that the underlying genetics tends to be simple, with often only one or two major-effect loci. Where known, at least one of the causal loci encodes an immune protein, often an intracellular nucleotide binding site-leucine-rich repeat (NLR) protein [4–13]. The gene family encoding NLR immune receptors is the most variable gene family in plants, both in terms of inter- and intraspecific variation [14–17]. Many NLR proteins function as major disease resistance (R) proteins, with the extravagant variation at these loci being due to a combination of maintenance of very old alleles by long-term balancing selection and rapid evolution driven by strong diversifying selection [18–20]. The emergence of new variants is favored by many NLR genes being organized in tandem clusters, which can spawn new alleles as well as copy number variation by illegitimate recombination, and by the presence of leucine-rich repeats in NLR genes, which can lead to expansion and contraction of coding sequences [21–23]. Cluster expansion has been linked to diversification and adaptation in a range of systems [24–26]. Several complex plant NLR loci provide excellent examples of cluster rearrangement increasing pathogen recognition specificities [19]. Substantial efforts have been devoted to decomposing the complexity of the plant immune system and interactions between its components.

While many plant disease R genes are members of the NLR family, some feature different molecular architectures. One of these is *RESISTANCE TO POWDERY MILDEW 8 (RPW8)* in *Arabidopsis thaliana*, which was initially identified based on an allele that confers resistance to multiple powdery mildew isolates [27] and later shown also to provide resistance to oomycetes [28,29]. The namesake *RPW8* gene is located in a gene cluster of variable size and composition that includes multiple *RPW8*-like genes as well as *HOMOLOG OF RPW8 (HR)* genes [27,30,31]. The reference accession Col-0, which is susceptible to powdery mildew, has four *HR* genes, but no *RPW8* gene, whereas the resistant accession Ms-0 carries *RPW8.1* and *RPW8.2* along with three *HR* genes [27]. Several RPW8 proteins from *A. thaliana* and *Brassica spp*. become localized to the extra-haustorial membrane upon powdery mildew infection, highlighting their potential function at the host-microbe interface [29,32,33]. NLRs are distinguished by N-terminal Toll/interleukin-1 receptor (TIR) or coiled-coil (CC) domains, which,

when overexpressed alone, can often activate immune signaling [34,35]. A subset of CC-NLRs (CNLs) has a diagnostic type of coiled-coil domain, termed CC_R to indicate that this domain is being shared with RPW8/HR proteins. The latter have an N-terminal extension that might be a transmembrane domain as well as C-terminal repeats of unknown activity [36,37]. It has been noted that the CC_R domain is similar to a portion of the animal mixed-lineage kinase domain-like (MLKL) protein that forms a multi-helix bundle [38] as well as the HeLo and HELL domains of fungi, which also form multi-helix bundles [39–41]. Many fungal HeLo domain proteins have a prion-forming domain that consists of C-terminal 21-amino acid repeats. This domain can form amyloids and thereby affect oligomerization and activity of these proteins [39–43].

We have previously reported hybrid necrosis due to incompatible alleles at the *RPW8/HR* locus and at the complex *RECOGNITION OF PERONOSPORA PARASITICA 7* (*RPP7*) locus, which encodes a canonical CNL and which has alleles that provide race-specific resistance to the oomycete *Hyaloperonospora arabidopsidis* [44,45]. Here, we investigate in detail three independent cases of incompatible *RPW8/HR* and *RPP7*-like alleles, and show that two are caused by members of the fast-evolving RPW8.1/HR4 clade. We describe how variation in the number of C-terminal repeats and the short C-terminal tail predict the degree of incompatibility between two common *RPW8.1/HR4* alleles and corresponding *RPP7*-like alleles.

Results

Distinct pairs of *RPP7* and *RPW8/HR* alleles cause hybrid necrosis

In a systematic intercrossing and genetic mapping program among 80 *A. thaliana* accessions, a series of genomic regions involved in hybrid incompatibility were identified [10]. The underlying genes were termed *DANGEROUS MIX* (*DM*) loci. One instance, between the *DM6* and *DM7* regions, stood out because it is responsible for two phenotypically distinct hybrid necrosis cases (Fig 1A) [10]. Strong candidates, as previously inferred from a combination of mapping, gene knockdown and transformation with genomic constructs, suggested that *DM6* corresponds to the *RPP7* cluster, and *DM7* to the *RPW8/HR* cluster. We recently found an additional case of incompatibility between the *DM6* and *DM7* regions, with a third distinctive phenotype (Figs 1A and 2A). In addition to phenotypic differences between the three *DM6–DM7* F_1 hybrids, test crosses confirmed that each case was caused by different combinations of *DM6* and *DM7* alleles, as only certain combinations resulted in hybrid necrosis (Fig 1B).

To corroborate the evidence from mapping experiments that *DM6* alleles of Mrk-0 and ICE79 were *RPP7* homologs, we designed ten artificial microRNAs (amiRNAs) based on sequences from the Col-0 reference accession. AmiRNAs targeting a subclade of five *RPP7* homologs that make up the second half of the *RPP7* cluster in Col-0, suppressed hybrid necrosis in all three crosses, Mrk-0 x KZ10, Lerik1-3 x Fei-0 and ICE79 x Don-0 (S1 Fig and S1 Table). These rescue experiments, together with the above-mentioned test crosses, indicate that specific *RPP7* homologs in Mrk-0, Lerik1-3 and ICE79 correspond to different *DM6* alleles that cause hybrid necrosis in combination with specific *DM7* alleles from other accessions.

A common set of *RPW8/HR* haplotypes affecting hybrid performances in F_1 and F_2 progeny

In the mentioned set of diallelic F_1 crosses among 80 accessions [10], we noted that the *DM6* carrier Lerik1-3 was incompatible with several other accessions, suggesting that these have *DM7* (*RPW8/HR*) hybrid necrosis risk alleles that are similar to the one in Fei-0. Crosses with

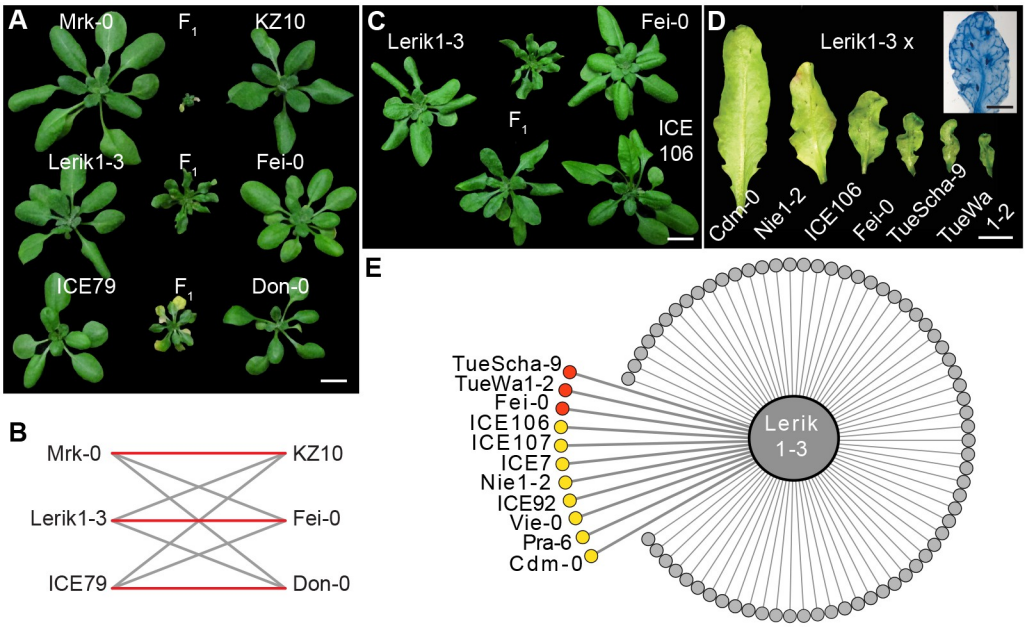


Fig 1. *DM6–DM7* hybrid necrosis cases. (A) Morphological variation in three independent *DM6–DM7* hybrid necrosis cases. **(B)** Red lines indicate necrosis in F_1 hybrids, grey indicates normal progeny. **(C, D)** Variation in morphology in two *DM6–DM7* cases sharing the same *DM6* allele in Lerik1-3. **(C)** Entire rosettes of four-week-old plants. **(D)** Abaxial sides of eighth leaves of six-week-old plants. Inset shows Trypan Blue stained leaf of Lerik1-3 x Fei-0 F_1. **(E)** Summary phenotypes in crosses of Lerik1-3 to 80 other accessions. Red is strong necrosis in F_1, and yellow is mild necrosis in F_1 or necrosis only observable in F_2. Scale bars indicate 1 cm.

https://doi.org/10.1371/journal.pgen.1008313.g001

TueScha-9 and TueWa1-2 produced hybrids that looked very similar to Lerik1-3 x Fei-0 progeny, with localized spots of cell death spreading across the leaf lamina along with leaf crinkling and dwarfism (Fig 1D and S2 Fig). Similar spots of cell death and leaf crinkling were observed in crosses of Lerik1-3 to ICE106 and ICE107, although these were not as dwarfed (Fig 1C and 1D and S2 Fig).

Hybrid necrosis often becomes more severe when the causal loci are homozygous [5,7,10,12]. To explore whether Lerik1-3 might cause milder forms of hybrid necrosis that are missed in the F_1 generation, we surveyed several F_2 populations involving Lerik1-3. Six segregated necrotic plants with very similar phenotypes (Fig 1D and 1E and S2 Fig). This makes all together for 11 incompatible accessions, which are spread over much of Eurasia (Fig 1E).

The F_2 segregation ratios suggested that the effects of the *DM7* allele from ICE106/ICE107 are intermediate between those of the Fei-0/TueWa1-2/TueScha-9 alleles and the Cdm-0/Nie-0 alleles (Table 1). Alternatively, the hybrid phenotypes might be affected by background modifiers, such that identical *DM7* alleles produce a different range of phenotypes in combination with $DM6^{Lerik1\text{-}3}$.

Because the phenotypic variation among hybrid necrosis cases involving Lerik1-3 could involve loci other than *DM6* and *DM7*, we carried out linkage mapping with Lerik1-3 x ICE106 and Lerik1-3 x ICE107 crosses. We combined genotyping information from Lerik1-3 x ICE106 and Lerik1-3 x ICE107 F_2 and F_3 individuals for mapping, because the genomes of ICE106 and ICE107, which come from closeby collection sites, are very similar and because the two crosses produce very similar F_1 hybrid phenotypes, suggesting that the responsible alleles are likely to be identical. We used F_3 populations to better distinguish different phenotypic classes, since we did not know the number of causal genes nor their genetic behavior.

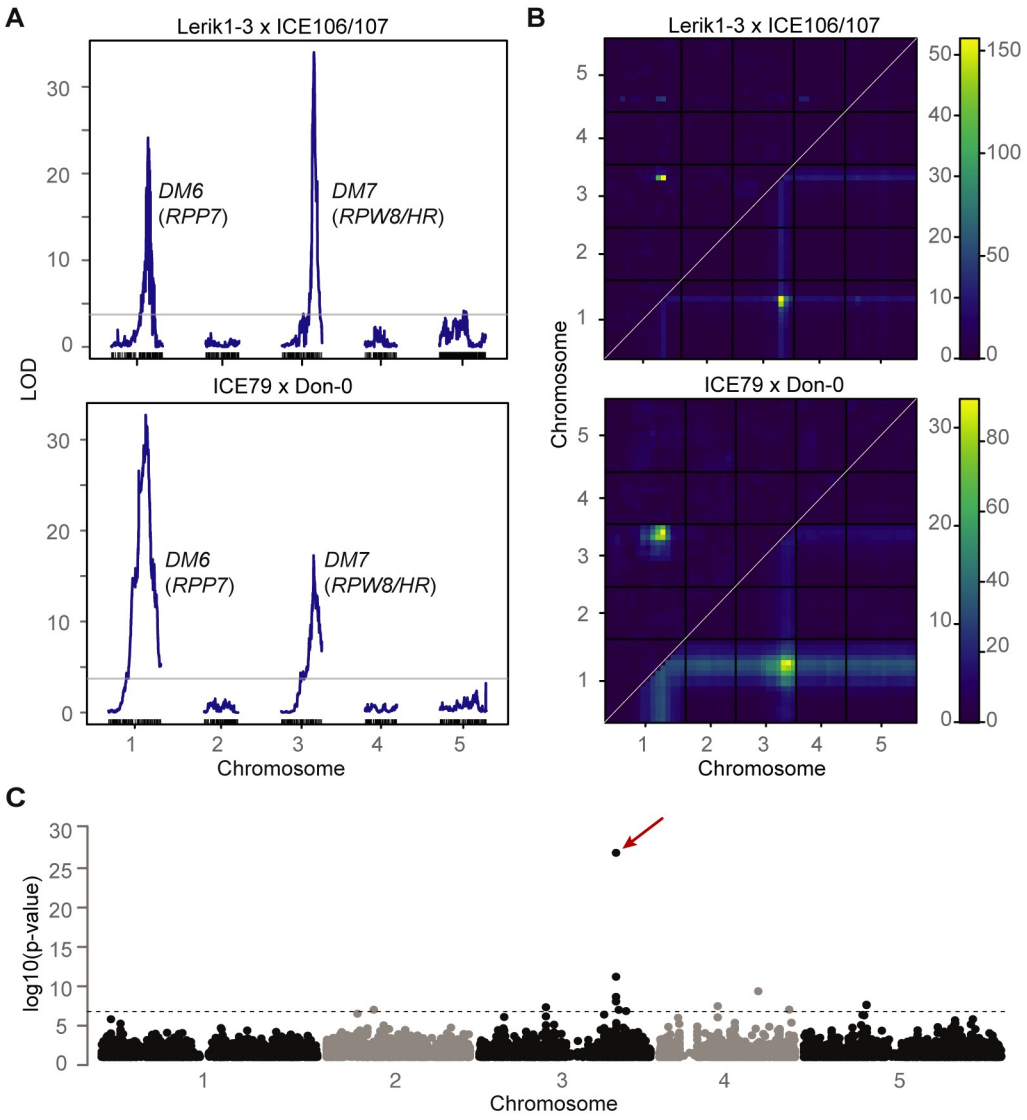


Fig 2. Mapping of two *DM6* (*RPP7* cluster)–*DM7* (*RPW8*/*HR* cluster) hybrid necrosis cases. (A) QTL analyses. The QTL on chromosome 1 includes *RPP7* from Lerik1-3 and ICE79 (21.37–22.07 and 21.50–21.98 Mb), and the QTL on chromosome 3 *RPW8*/*HR* from ICE106/ICE107 and Don-0 (18.59–19.09 Mb, 18.61–19.06 Mb). The horizontal lines indicate 0.05 significance thresholds established after 1,000 permutations. **(B)** Heat map for two-dimensional, two-QTL model genome scans. Upper left triangles indicate epistasis scores (LOD*i*) and lower right triangles joint two-locus scores (LOD*f*). Scales for LOD*i* on left and for LOD*f* on right. **(C)** Manhattan plot for a GWAS of necrosis in hybrid progeny of Lerik1-3 crossed to 80 other accessions (see S2 Table). The hit in the *RPW8*/*HR* region (red arrow) stands out, but it is possible that some of the other hits that pass the significance threshold (Bonferroni correction, 5% familywise error) identify modifiers of the *DM6*–*DM7* interaction.

https://doi.org/10.1371/journal.pgen.1008313.g002

QTL analysis confirmed that the *DM6* and *DM7* genomic regions are linked to hybrid necrosis in these crosses (Fig 2A and 2B).

To narrow down the *DM7* mapping interval, we took advantage of having 11 accessions that produced hybrid necrosis in combination with Lerik1-3, and 69 accessions (including Lerik1-3 itself) that did not. We performed GWAS with Lerik1-3-dependent hybrid necrosis as a binary trait [46]. The by far most strongly associated marker was immediately downstream of *HR4*, the last member of the *RPW8*/*HR* cluster in Col-0 (Fig 2C and S2 Table). An amiRNA

Table 1. F_2 segregation ratios at 16˚C.

Cross	n[a]	Phenotype				Model[d]	χ^2
		Normal[b]	F_1-like[b]	Enhanced[c]			
Fei-0/Lerik1-3	384	178	107	99		I	0.85
TueWa1-2/Lerik1-3	138	66	42	30		I	0.36
TueScha-9/Lerik1-3	193	92	44	57		I	0.42
Lerik1-3/ICE106	265	121	67	62	15	II	0.89
Lerik1-3/ICE107	291	204		70	17	III	0.88
Cdm-0/Lerik1-3	260	173		71	16	III	0.68
Nie-0/Lerik1-3	227	170		57		IV	0.59

[a]. If the model had a class of dead segregants that could not be counted, *n* was estimated to include the dead individuals for χ^2 calculation.

[b]. In the bottom three populations, F_1 phenotypes were nearly indistinguishable from normal ones and therefore both classes were combined.

[c]. More severe than F_1 hybrids with distinct *DM6–DM7* phenotypes. For milder cases, the enhanced phenotypic classes were separated into two groups, with a rosette diameter of 1 cm as threshold. The rightmost numbers indicate the most severe class.

[d]. Best-fit models using F_2 segregation analyses with incompatibility alleles indicated as "A" and "B".

I: two-loci-semi-dominant; AaBb F_1-like; AABb and AaBB stronger than F_1; AABB dead and not countable.

II: two-loci-semi-dominant; AaBb F_1-like; AABb stronger than F_1; AaBB and AABB almost dead, but countable.

III: two-loci-semi-dominant; Aabb and aaBb (normal) and AaBb (F_1-like) not easily distinguished; AABb and AaBB stronger than F_1; AABB almost dead, but countable.

IV: two-loci-semi-dominant; Aabb and aaBb (normal) and AaBb (F_1-like) not easily distinguished; AABb and AaBB stronger than F_1; AABB dead and not countable.

https://doi.org/10.1371/journal.pgen.1008313.t001

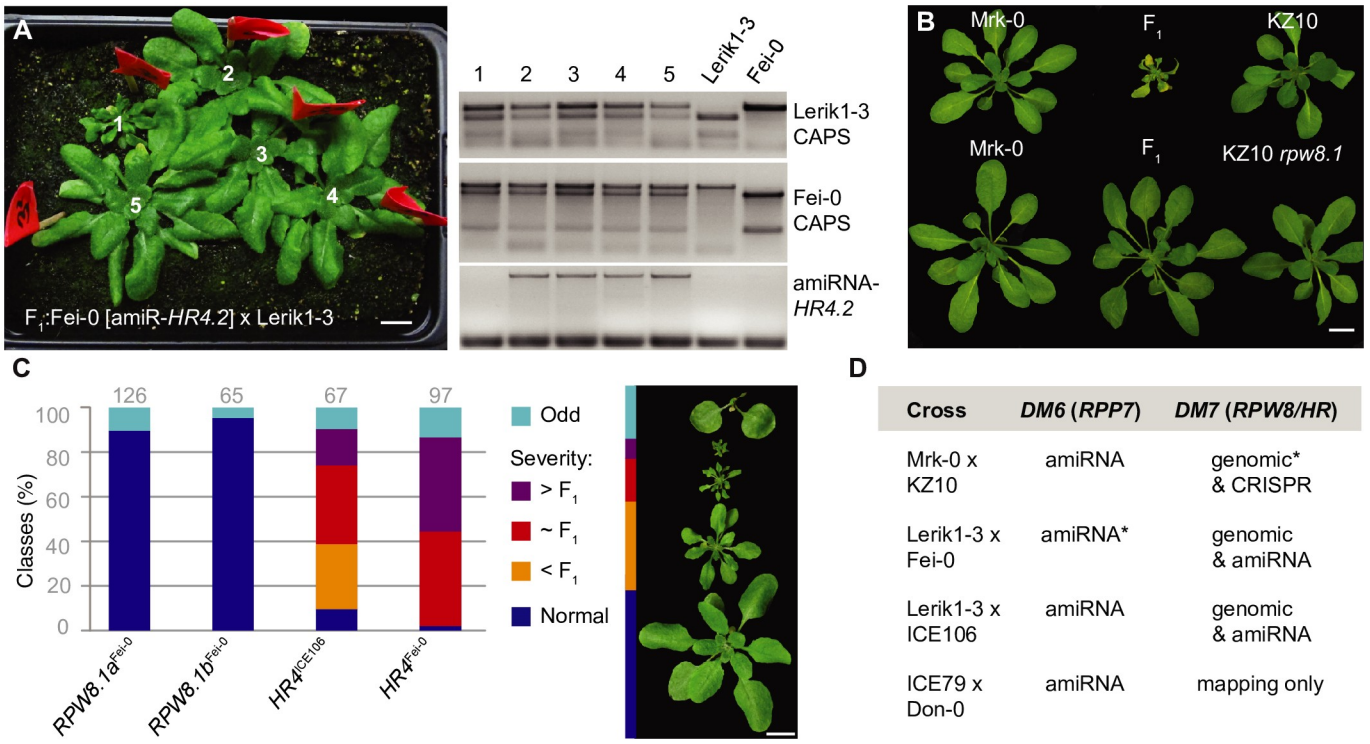
matching *HR4* sequences from Col-0 fully rescued both the strong necrosis in Lerik1-3 x Fei-0 and the weaker necrosis in Lerik1-3 x ICE106 (Fig 3A and S3 Table). We confirmed the causality of another member of the *RPW8*/*HR* cluster in the KZ10 x Mrk-0 case with a CRISPR/Cas9-induced mutation of $RPW8.1^{KZ10}$ (Fig 3B and S3 Fig).

In Col-0, but not in all *A. thaliana* accessions, resistance to *H. arabidopsidis* Hiks1 maps to the *RPP7* cluster [47,48]. The *RPP7*-like hybrid necrosis risk allele carrier Lerik1-3 was resistant to Hiks1 as well, but Fei-0 and ICE106 were not. Resistance was inherited in a dominant manner (S4 Fig and S4 Table). We further used seven different amiRNAs against *RPP7* homologs, three of which had suppressed hybrid necrosis in combination with $HR4^{Fei-0}$ (S1 Table), to test whether *RPP7* homologs underlie Hiks1 resistance in Lerik1-3. That none of the amiRNAs reduced Hiks1 resistance indicates minimally that there is no simple correspondence between the *RPP7*-like hybrid necrosis risk allele and the Hiks1 resistance gene. We also asked whether *HR4* is required for *RPP7*-mediated Hiks1 resistance in Col-0. Two independent *hr4* CRISPR/Cas9 knockout lines in Col-0 (S3 Fig) remained completely resistant to Hiks1 (S4 Fig and S4 Table), indicating that *HR4* in Col-0 is dispensable for *RPP7*-mediated resistance to Hiks1.

Structural variation of the *RPW8*/*HR* cluster

For reasons of convenience, we assembled the *RPW8*/*HR* cluster from TueWa1-2 instead of Fei-0; accession TueWa1-2 interacted with *RPP7*-like gene from Lerik1-3 in the same manner as Fei-0; the strong necrosis in Lerik1-3 x TueWa1-2 was rescued with the same amiRNA as in Lerik1-3 x Fei-0 (S3 Table), and TueWa1-2 had an *HR4* allele that was identical in sequence to $HR4^{Fei-0}$. We found that the *RPW8*/*HR* cluster from TueWa1-2 had at least 13 *RPW8*/*HR*-like genes, several of which were very similar to each other (Fig 4A). For example, there were at least four copies of *RPW8.3*-like genes with 93 to 99.8% sequence similarity, and two identical *RPW8.1* genes, named *RPW8.1a*, followed by distinct *RPW8*/*HR* copies.

Recapitulation experiments had identified $HR4^{Fei-0}$ (identical to $HR4^{TueWa1-2}$ and $HR4^{TueScha-9}$) and $HR4^{ICE106}$ as causal for hybrid necrosis (Fig 3C and 3D). We analyzed the



Cross	*DM6* (*RPP7*)	*DM7* (*RPW8/HR*)
Mrk-0 x KZ10	amiRNA	genomic* & CRISPR
Lerik1-3 x Fei-0	amiRNA*	genomic & amiRNA
Lerik1-3 x ICE106	amiRNA	genomic & amiRNA
ICE79 x Don-0	amiRNA	mapping only

Fig 3. Confirmation of causal genes in *RPW8/HR* cluster. (A) Rescue of hybrid necrosis in Lerik1-3 x Fei-0 F_1 plants with an amiRNA against *HR4*. Fei-0 parents were T_1 transformants. PCR genotyping of numbered plants from left shown on the right. Only plant 1, which does not carry the amiRNA, is necrotic and dwarfed. **(B)** Rescue of hybrid necrosis in Mrk-0 x KZ10 F_1 plants by CRISPR/Cas9-targeted mutagenesis on $RPW8.1^{KZ10}$. **(C)** Recapitulation of hybrid necrosis in Lerik1-3 T_1 plants transformed with indicated genomic fragments from Fei-0 and ICE106. Representative phenotypes on right. Numbers of T_1 plants examined given on top. **(D)** Summary of rescue and recapitulation experiments. Asterisks refer to published experiments [10]. Scale bars indicate 1 cm.

https://doi.org/10.1371/journal.pgen.1008313.g003

phylogenetic relationship of the *RPW8/HR* genes in TueWa1-2 with the ones from published *RPW8/HR* clusters in *A. thaliana*, in *A. lyrata* and in *Brassica* spp. [10,30,31,49]. In *A. thaliana*, *RPW8/HR* genes seem to have undergone at least three duplication events, with the first one generating a new *A. thaliana* specific clade, which gave rise to independent *RPW8.1/HR4* and *RPW8.2/RPW8.3* duplications.

The *RPW8/HR* cluster of TueWa1-2 consists of *RPW8/HR* members from both the ancestral and the two *A. thaliana* specific clades, an arrangement that has not been observed before. Using species-wide data [50], we found that accessions carrying Col-0-like *HR4* alleles have simple cluster configurations, while accessions with *HR4* genes resembling hybrid necrosis alleles have more complex configurations (Fig 4A). The tagging SNPs found in GWAS (Fig 4A and S2 Table) were mostly found to be associated with the complex clusters, suggesting that the tagging SNPs are linked to structural variation in the distal region of the *RPW8/HR* cluster (Fig 4B).

Causality of RPW8/HR C-terminal repeats

To further narrow down the mutations that cause autoimmunity, we compared $RPW8.1^{KZ10}$ and $HR4^{Fei-0}$ with other *RPW8/HR* alleles from the global *A. thaliana* collection [50]. Some *RPW8.1* alleles have intragenic duplications of a sequence encoding a 21-amino acid repeat (QWDDIKEIKAKISEMDTKLA[D/E]) at the C-terminal end of the protein [31]. In HR4, there is a related 14-amino acid repeat (IQV[H/D]QW[T/I]DIKEMKA). Both RPW8.1 and

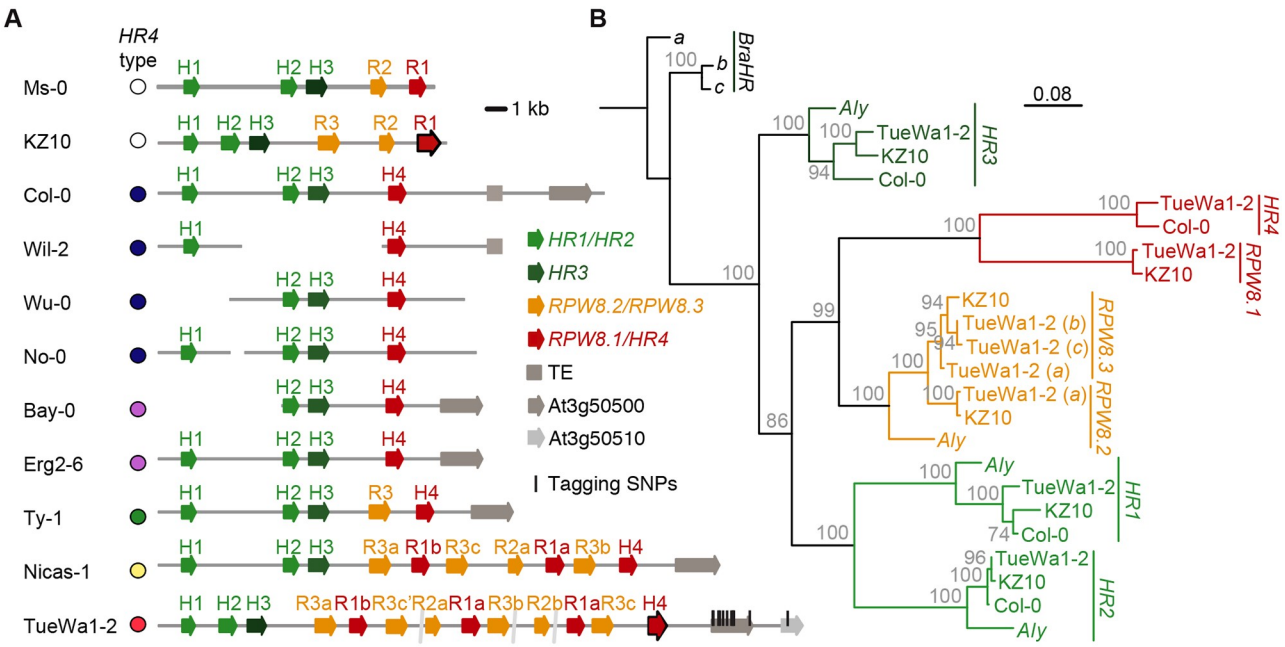


Fig 4. Structural variation of the *RPW8/HR* cluster. (A) The *RPW8/HR* cluster in different accessions. The extreme degree of recent duplications in TueWa1-2, with the same *HR4* hybrid necrosis risk allele as Fei-0, did not allow for closure of the assembly from PCR products; assembly gaps are indicated. Color coding of *HR4* alleles according to Fig 6. *HR4* and *RPW8.1* form a distinct clade from other *RPW8*s. Tagging SNPs found in GWAS marked in TueWa1-2 *RPW8/HR* cluster as black vertical lines. **(B)** Maximum likelihood tree of *RPW8/HR* genes from three *A. thaliana* accessions and the *A. lyrata* and *B. rapa* reference genomes. Branch lengths in nucleotide substitutions are indicated. Bootstrap values (out of 100) are indicated on each branch.

https://doi.org/10.1371/journal.pgen.1008313.g004

HR4 repeats are predicted to fold into extended alpha-helices, but only RPW8.1 repeats appear to have the potential to form coiled coils [51].

The number of repeats varies in both *RPW8.1* and *HR4* between hybrid necrosis risk and non-risk alleles. To experimentally test the effect of repeat number variation and other polymorphisms, we generated a series of derivatives in which we altered the number of repeats and swapped different portions of the coding sequences between the $RPW8.1^{KZ10}$ risk and $RPW8.1^{Ms\text{-}0}$ non-risk alleles, and between the $HR4^{Fei\text{-}0}$ and $HR4^{ICE106}$ risk and the $HR4^{Col\text{-}0}$ non-risk alleles (Fig 5A).

A 1.4 kb promoter fragment of $RPW8.1^{KZ10}$ and a 1.2 kb promoter fragment of $HR4^{Fei\text{-}0}$ in combination with coding sequences of risk alleles were sufficient to induce hybrid necrosis (Figs 3C, 5A and 5B). To simplify discussion of the chimeras, the N-terminal portion was labeled with the initial of the accession in italics ("*M*", "*K*", etc.), complete repeats were labeled with different capital letters to distinguish sequence variants ("A", "B", etc.), the partial repeat in KZ10 with a lowercase letter ("c"), and the C-terminal tails with Greek letters ("α", "β", etc.).

In $RPW8.1^{KZ10}$, there are two complete repeats and one partial repeat, while $RPW8.1^{Ms\text{-}0}$ has only one repeat (Fig 5A). Modifying the number of repeats in *RPW8.1* affected the frequency and severity of necrosis in T_1 plants in a Mrk-0 background, which carries the interacting *RPP7*-like allele, dramatically. Deletion of the first full repeat in $RPW8.1^{KZ10}$ ("*K*-Bcβ", with the KZ10 configuration being "*K*-BBcβ") substantially reduced the number of plants that died in the first three weeks of growth. The additional deletion of the partial repeat ("*K*-Bβ") reduced death and necrosis even further (Fig 5A). That *K*-Bβ still produces some necrosis, even though its repeat structure is the same as in the inactive *K*-Aα suggests that the

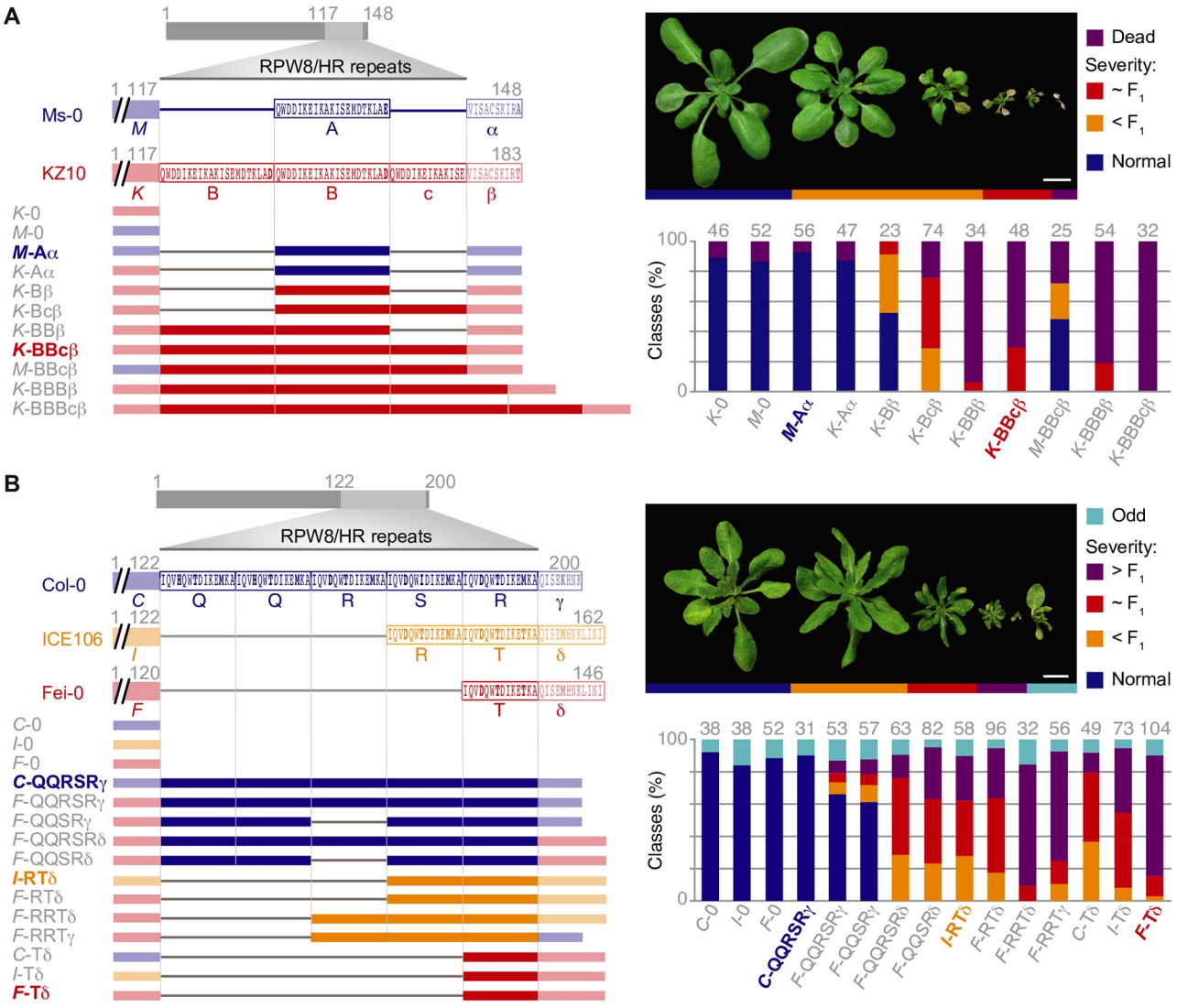


Fig 5. Necrosis-inducing activity of *RPW8.1* and *HR4* chimeras. N-terminal portions indicated with the initial of the accession in italics ("*K*", "*M*", etc.), complete repeats indicated with regular capital letters ("A", "B", etc.), the partial repeat in KZ10 with a lowercase letter ("c"), and the C-terminal tails with Greek letters ("α", "β", etc.). Non-repeat portions are semi-transparent. Repeats with identical amino acid sequences have the same letter designation. Numbers indicate amino acid positions. Constructs on the left, and distribution across phenotypic classes in T_1 transformants on the right, with *n* given on top of each column. Natural alleles labeled in color and bold. RPW8/HR repeats indicated as light grey boxes. **(A)** *RPW8.1* chimeras, driven by the $RPW8.1^{KZ10}$ promoter, were introduced into Mrk-0, which carries the corresponding incompatible *RPP7*-like allele. **(B)** *HR4* chimeras, driven by the $HR4^{Fei-0}$ promoter, were introduced into Lerik1-3, which carries the corresponding incompatible *RPP7*-like allele. Scale bars indicate 1 cm.

https://doi.org/10.1371/journal.pgen.1008313.g005

polymorphism in the C-terminal tail makes some contribution to necrosis activity. It is less likely that the polymorphism in the repeats play a role, as there is only a very conservative aspartate-glutamate difference between A and B repeats.

In contrast to repeat shortening, the extension of the partial repeat ("*K*-BBBβ") or addition of a full repeat ("*K*-BBBcβ") increased the necrosis-inducing activity of $RPW8.1^{KZ10}$, such that almost all T_1 plants died without making any true leaves. However, it appears that not all repeats function equally, as removal of the partial repeat slightly increased necrosis-inducing activity ("*K*-BBβ"). Polymorphisms in the N-terminal non-repeat region seemed to contribute to necrosis, as swaps of the N-terminal Ms-0 fragment ("*M*-BBcβ" or "*M*-BBBβ") induced

weaker phenotypes than the corresponding variants with the N-terminal fragment from KZ10. Nevertheless, we note that the normal KZ10 repeat configuration was sufficient to impart substantial necrosis-inducing activity on a chimera in which the N-terminal half was from Ms-0, which is distinguished from KZ10 by nine nonsynonymous substitutions outside the repeats.

Compared to the RPW8.1 situation, the relationship between HR4 repeat length and necrosis-inducing activity is more complex. The natural alleles suggested a negative correlation of repeat number with necrosis-inducing activity when crossed to Lerik1-3, since the non-risk HR4 allele from Col-0 has five full repeats, while weaker risk alleles such as the one from ICE106 have two, and the strong risk allele from Fei-0 has only one (Fig 5B). Addition of a full repeat to HR4^{Fei-0} ("*F*-RTδ", with the original Fei-0 configuration being "*F*-Tδ") reduced its activity to a level similar to that of HR4^{ICE106} ("*I*-RTδ"). Deletion of a full repeat from HR4^{ICE106} ("*I*-Tδ") modestly increased HR4 activity (Fig 5B). Together, the chimera analyses indicated that the quantitative differences between crosses of Fei-0 and ICE106 to Lerik1-3 (Fig 1 and S2 Fig) are predominantly due to variation in HR4 repeat number. This is further supported by the necrosis-inducing activity of a chimera in which the repeats in the Col-0 non-risk allele were replaced with those from *HR4*$^{Fei-0}$ ("*C*-Tδ", with the original Col-0 configuration being "*C*-QQRSRγ") (Fig 5B and S5 Fig). However, repeat number alone is not the only determinant of necrosis-inducing activity of *HR4* in combination with *RPP7-like*$^{Lerik1-3}$. Adding another repeat to the "*F*-RTδ" chimera, resulting in "*F*-RRTδ", increased the activity of *HR4*$^{Fei-0}$ again, perhaps suggesting that there is an optimal length for *HR4* to interact with the cognate *RPP7*.

Unlike RPW8.1, the C-terminal tails of HR4 proteins beyond the RPW8/HR repeats (fragments "γ" and "δ") differ in length between hybrid necrosis-risk and non-risk variants (Fig 5B). Swapping only these two fragments affected HR4 activity substantially, and converted two chimeras with weak necrosis-inducing activity ("*F*-QQRSRγ" to "*F*-QQRSRδ" and "*F*-QQSRγ" to "*F*-QQSRδ") into chimeras with activity resembling that of HR4^{ICE106} (which is "*I*-RTδ").

Taken together, the swap experiments led us to conclude that naturally occurring variation in the configuration of RPW8/HR repeats play a major role in quantitatively modulating the severity of autoimmune phenotypes when these *RPW8*/*HR* variants are combined with *RPP7* alleles from Mrk-0 and Lerik1-3. At least in the case of HR4, we could show directly that the short C-terminal tail also affects the hybrid phenotype, while for RPW8.1 this seems likely as well, given that the repeats between different alleles differ less from each other than the tails.

Prediction of *RPP7*-dependent hybrid performance using *RPW8.1*/*HR4* haplotypes

To obtain a better picture of *RPW8.1*/*HR4* variation, we remapped the raw reads from the 1001 Genomes project to the longest *RPW8.1* and *HR4* alleles, *RPW8.1*^{KZ10} and *HR4*$^{Col-0}$, as references (S5 and S6 Tables). The results suggested that *HR4*-carrying accessions are more rare than those carrying *RPW8.1* alleles (285 vs. 903 out of 1,221 accessions). The short, necrosis-linked, *HR4* risk alleles (Fig 6A) were predicted to be as frequent as the long non-risk variants (Fig 6A and 6B and S5 Table), whereas for *RPW8.1*, only seven accessions were predicted to have the long *RPW8.1*^{KZ10}-type risk variant (Fig 6A and S6 Table).

To confirm the short read-based length predictions, *RPW8.1* was PCR amplified from 28 accessions and *HR4* from 113 accessions (Fig 6A–6D and S5 and S6 Tables). This not only confirmed that the Illumina predictions were accurate, but also revealed new variants with different arrangements of *HR4* repeats, although none were as short as *HR4*$^{Fei-0}$ or *HR4*^{ICE106} (Fig 6A and 6B). The short necrosis-risk *HR4* variants are found across much of the global range of *A. thaliana* (Fig 6C), whereas the much rarer necrosis-risk *RPW8.1*^{KZ10}-like variant was

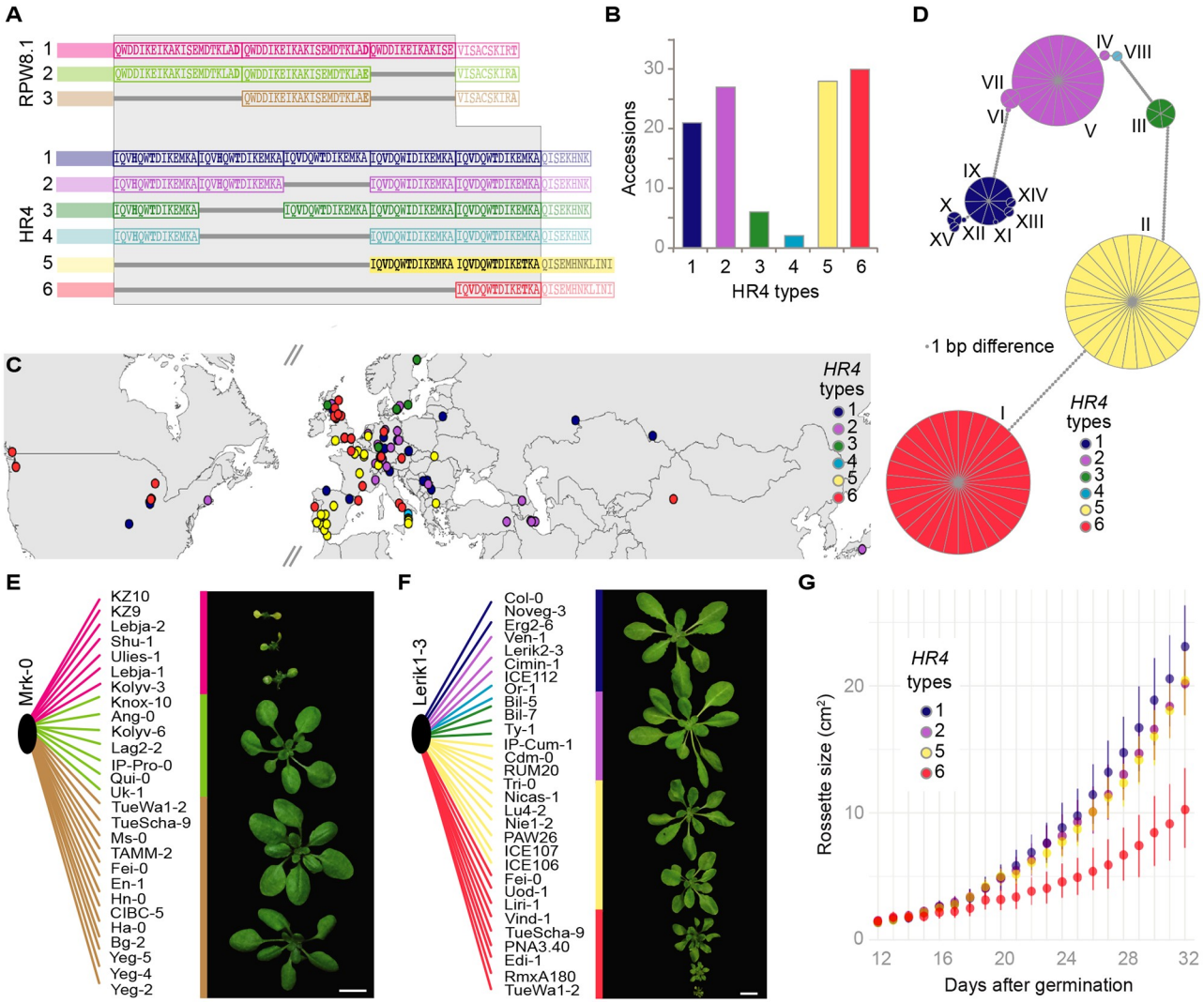


Fig 6. Sequence variation of a large collection of *RPW8.1* and *HR4* alleles. **(A)** Repeat polymorphisms in RPW8.1 and HR4 proteins (grey background). N-terminal regions and tails are semi-transparent. **(B)** Distribution of *HR4* types across 113 Sanger sequenced alleles (see S5 Table). **(C)** Distribution of *HR4* allele types in Eurasia and North America. **(D)** Haplotype network of *HR4* alleles, with a 1-bp minimum difference. **(E)** F_1 progeny of Mrk-0 crossed to accessions with different *RPW8.1* alleles. Short *RPW8.1* variants do not induce hybrid necrosis. **(F)** F_1 progeny of Lerik1-3 crossed to accessions with different *HR4* alleles. The shortest *HR4* alleles (red) cause strong hybrid necrosis, the second shortest *HR4* alleles (yellow) cause mild hybrid necrosis. **(G)** Rosette growth of F_1 progeny from Lerik1-3 and accessions carrying different *HR4* alleles. The shortest *HR4* allele causes a strong growth reduction, while the second-shortest *HR4* allele has a milder effect. Scale bars indicate 1 cm.

https://doi.org/10.1371/journal.pgen.1008313.g006

exclusive to Central Asia. We also observed that sequences of the two short *HR4* types were more conserved than the longer ones, with each short type belonging to a single haplotype, while the long necrosis-risk *HR4* alleles belonged to multiple haplotypes (Fig 6D).

The extensive information on *RPW8.1/HR4* haplotypes allowed us to use test crosses to determine whether interaction with either causal *RPP7*-like genes from Mrk-0 or Lerik1-3 is predictable from sequence, specifically from repeat number (Fig 6E and 6F). As expected, accessions with the longest, Type 1, $RPW8.1^{KZ10}$-like alleles (Fig 6E, **pink**) produced necrotic hybrid progeny when crossed to Mrk-0, whereas accessions carrying the two shorter Type 2 and 3 alleles did not (Fig 6E and S7 Table). The situation was similar for *HR4*; all but two of the tested accessions with the shortest $HR4^{Fei\text{-}0}$-like alleles (Fig 6F, **red**) produced strongly

necrotic progeny when crossed to Lerik1-3, while accessions carrying the second shortest $HR4^{ICE106}$-like alleles (Fig 6F and S8 Table) produced more mildly affected progeny. Hybrid progeny of Lerik1-3 and accessions carrying other *HR4* alleles did not show any signs of necrosis (Fig 6F). Necrosis was correlated with reduction in overall size of plants, which in turn correlated with RPW8.1/HR4 repeat length (Fig 6F and S9 Table). Finally, $HR4^{Fei-0}$-like alleles in two accessions caused a mild phenotype similar to $HR4^{ICE106}$, suggesting the presence of genetic modifiers that partially suppress autoimmune symptoms.

Discussion

The *RPW8*/*HR* cluster is remarkably variable in terms of copy number, reminiscent of many multi-gene clusters carrying NLR-type *R* genes [16]. While the first three genes in the cluster, *HR1*, *HR2* and *HR3*, are generally well conserved, there is tremendous variation in the number of the other genes in the cluster, including *RPW8.1*/*HR4*. Nevertheless, that the *HR4* hybrid necrosis-risk allele is not rare and widely distributed, accounting for half of all *HR4* carriers (Fig 6B and 6C), suggests that it might provide adaptive benefits, as postulated before for *ACD6* hybrid necrosis-risk alleles [12].

The N-terminal portion of RPW8 and HR proteins can be homology modeled on a multi-helix bundle in the animal MLKL protein [38], which in turn shares structural similarity with fungal HeLo and HELL domain proteins [41]. In both cases, the N-terminal portions can insert into membranes (with somewhat different mechanisms proposed for the two proteins), thereby disrupting membrane integrity and triggering cell death [40,52–54]. For both proteins, insertion is regulated by sequences immediately C-terminal to the multi-helix bundle [40,52–56]. It is tempting to speculate that the RPW8/HR repeats and the C-terminal tail, which together make up the C-terminal portions of the proteins, similarly regulate activity of RPW8.1 and HR4. In agreement, our chimera studies, where we exchanged and varied the number of RPW8/HR repeats and swapped the C-terminal tail, indeed point to the C-terminal portion of RPW8/HR proteins having a regulatory role. A positive regulator of RPW8-mediated disease resistance, a 14-3-3 protein, interacts specifically with the C-terminal portion of RPW8.2, consistent with this part of the protein controlling RPW8/HR activity [57]. Perhaps even more intriguing is the fact that in many fungal HeLo domains this C-terminal region is a prion-forming domain composed of 21-amino acid repeats. RPW8.1 also has 21-amino acid repeats, while HR4 has 14-amino acid repeats, but in both cases these were not interrupted by a spacer, as in the fungal proteins. In fungal HET-S and related proteins, the repeats exert regulatory function by forming amyloids and thereby causing the proteins to oligomerize [39–43]. While it remains to be investigated whether the RPW8/HR repeats and the C-terminal tail function in a similar manner, their potential regulatory function makes them a possible target for pathogen effectors. In such a scenario, at least some RPP7 proteins might act as guards for RPW8/HR proteins and sense their modification by pathogen effectors [16,58].

Can we conclude from the MLKL homology that RPW8 and HR proteins form similar pores as MLKL? Unfortunately, this is not immediately obvious, as a different mechanism has been suggested for fungal proteins with HeLo and HELL domains [39–41]. For MLKL, it has been suggested that the multi-helix bundle directly inserts into the membrane, whereas for the fungal protein, it has been proposed that the multi-helix bundle regulates the ability of an N-terminal transmembrane domain to insert into the membrane. An N-terminal transmembrane domain has been predicted for RPW8 [27], but although RPW8 proteins can be membrane associated [33,59], the insertion of this domain into the membrane has not been directly demonstrated.

We have shown that differences in protein structure, rather than expression patterns or levels, are key to the genetic interaction between RPW8/HR and RPP7. While we do not know whether the proteins interact directly, allele-specific genetic interactions are often an indicator of direct interaction between the gene products [60]. Moreover, reminiscent of RPW8/HR and RPP7 interaction, the activity of the fungal HeLo domain protein HET-S is regulated by an NLR protein [42].

Finally, we would like to emphasize that our observations do not necessarily imply that *RPP7* and *RPW8*/*HR* genes are obligatory partners. First, we found that *HR4* is not required for *RPP7*-dependent *Hpa* Hiks1 resistance in Col-0. Second, previous genetic studies have revealed both overlap and differences in the downstream signaling requirements of *RPP7* and *RPW8*/*HR* genes [44,61].

In conclusion, we have described in detail an intriguing case of hybrid necrosis in *A. thaliana*, where three different pairs of alleles at a conventional complex NLR resistance gene cluster, *RPP7*, and alleles at another complex, but non-NLR resistance gene cluster, *RPW8*/*HR*, interact to trigger autoimmunity in the absence of pathogens. Our findings suggest that within the immune system, conflict does not occur randomly, but that certain pairs of loci are more likely to misbehave than others. Finally, that genes of the *RPW8*/*HR* cluster can confer broad-spectrum disease resistance, while at least one *RPP7* member can confer race-specific resistance, provides yet another link between different arms of the plant immune system [62].

Materials and methods

Plant material

Stock numbers of accessions used are listed in Supplementary Material. All plants were stratified in the dark at 4˚C for 4–6 days prior to planting on soil. Late flowering accessions were vernalized for six weeks under short day conditions (8 h light) at 4˚C as seedlings. All plants were grown in long days (16 h light) at 16˚C or 23˚C at 65% relative humidity under Cool White fluorescent light of 125 to 175 μmol m^{-2} s^{-1}. Transgenic seeds were selected either with 1% BASTA (Sigma-Aldrich), or by mCherry fluorescence. Constructs are listed in S10 Table.

RAPA phenotyping

Images were acquired daily in top view using two cameras per tray. Cameras were equipped with OmniVision OV5647 sensors with a resolution of 5 megapixels. Each camera was attached to a Raspberry Pi computer (Revision 1.2, Raspberry Pi Foundation, UK) [63]. Images of individual plants were extracted using a predefined mask for each plant. Segmentation of plant leaves and background was then performed by removing the background voxels then a GrabCut-based automatic postprocessing was applied [64]. Lastly, unsatisfactory segmentations were manually corrected. The leaf area of each plant was then calculated based on the segmented plant images.

Histology

Cotyledons from 18 day-old seedlings were collected and 1 ml of lactophenol Trypan Blue solution (20 mg Trypan Blue, 10 g phenol, 10 ml lactic acid, 10 ml glycerol and 10 ml water) diluted 1: 2 in 96% ethanol was added for 1 hour at 70˚C. Trypan Blue was removed, followed by the addition of 1 ml 2.5g/ml chloral hydrate and an overnight incubation. The following day, the de-stained cotyledons were transferred to 50% glycerol and mounted on slides.

Pathology

The *Hyaloperonospora arabidopsidis* isolate Hiks1 was maintained by weekly subculturing on susceptible Ws-0 *eds1-1* plants [47]. To assay resistance of susceptibility, 12- to 13-day old seedlings were inoculated with 5 x 10^4 spores/ml. Sporangiophores were counted 5 days after infection.

Constructs and transgenic lines

Genomic fragments were PCR amplified, cloned into pGEM®-T Easy (Promega, Madison, WI, USA), and either directly transferred to binary vector pMLBart or Gateway vectors pJLBlue and pFK210. amiRNAs [65] against members of the *RPP7* and *RPW8*/*HR* clusters were designed using the WMD3 online tool (http://wmd3.weigelworld.org/), and placed under the CaMV 35S promoter in the binary vector pFK210 derived from pGreen [66]. amiRNA constructs were introduced into plants using *Agrobacterium*-mediated transformation [67]. T_1 transformants were selected on BASTA, and crossed to incompatible accessions. For the chimeras, promoters and 5' coding sequences were PCR amplified from genomic DNA, repeat and tail sequences were synthesized using Invitrogen's GeneArt gene synthesis service, all were cloned into pBlueScript. The three parts, promoter, 5' and 3' coding sequences, were assembled using Greengate cloning [68] in the backbone vector pMCY2 [69]. Quality control was done by Sanger sequencing. Transgenic T_1 plants were selected based on mCherry seed fluorescence. For CRISPR/Cas9 constructs, sgRNAs targeting *HR4* or *RPW8.1* were designed on the Chopchop website (http://chopchop.cbu.uib.no/), and assembled using a Greengate reaction into supervector pRW006 (pEF005-sgRNA-shuffle-in [70] Addgene plasmid #104441). mCherry positive T_2 transformants were screened for CRISPR/Cas9-induced mutations by Illumina MiSeq based sequencing of barcoded 250-bp amplicons. Non-transgenic homozygous T_3 lines were selected based on absence of fluorescence in seed coats.

Genotyping-by-sequencing and QTL mapping

Genomic DNA was isolated from Lerik1-3 x ICE106/ICE107 F_2 and F_3 individuals and from ICE79 x Don-0 F_2 individuals using a Biosprint 96 instrument and the BioSprint 96 DNA Plant Kit (Qiagen, Hilden, Germany). The individuals represented all classes of segregating phenotypes. Genotyping-by-Sequencing (GBS) using RAD-seq was used to genotype individuals in the mapping populations with *Kpn*I tags [71]. Briefly, libraries were single-end sequenced on a HiSeq 3000 instrument (Illumina, San Diego, USA) with 150 bp reads. Reads were processed with SHORE [72] and mapped to the *A. thaliana* Col-0 reference genome. QTL was performed using R/qtl with the information from 330 individuals and 2,989 markers for the Lerik1-3 x ICE106/107 populations, and 304 individuals and 2,207 markers for the ICE79 x Don-0 population. The severity of the hybrid phenotype was scored as a quantitative trait.

GWAS

Lerik1-3-dependent hybrid necrosis in F_1 progeny from crosses with 80 accessions [10] was scored as 1 or 0. The binary trait with accession information was submitted to the easyGWAS platform [46], using the FaSTLMM algorithm. A -$\log_{10}$(p-value) was calculated for every SNP along the five *A. thaliana* chromosomes.

RPP7 phylogeny

The NB domain was predicted using SMART (http://smart.embl-heidelberg.de/). NB amino acid sequences were aligned using MUSCLE (70). A maximum-likelihood tree was generated

using the BLOSUM62 model in RaxML (71). Topological robustness was assessed by bootstrapping 1,000 replicates.

RPW8.1/*HR4* length prediction

Short reads from the 1001 Genomes project (http://1001genomes.org) were mapped using SHORE [72] with 5 mismatches allowed per read. Sequences of the *RPW8*/*HR* clusters from Col-0 and KZ10 were provided as references and the covered region for $RPW8.1^{KZ10}$ and $HR4^{Col\text{-}0}$ was retrieved.

RPW8.1/*HR4* sequence analysis

Overlapping fragments covering the *HR4*/*RPW8.1* genomic region were PCR amplified from different *A. thaliana* accessions (oligonucleotides in S11 Table). Fragments were cloned and Sanger sequenced. A maximum-likelihood tree of coding portions of exons and introns was computed using RaxML [73] and visualized with Figtree.

Population genetic analysis

The geographical distribution of the 113 accessions carrying different *HR4* alleles was plotted using R (version 0.99.903). Packages maps, mapdata, mapplots and scales were used. A haplotype network was built using a cDNA alignment of 113 *HR4* alleles from different accessions. The R packages used were ape (dist.dna function) and pegas (haploNet function).

Oligonucleotides

See S11 Table.

Supporting information

S1 Fig. Role of the *RPP7* cluster in *DM6*–*DM7* dependent hybrid necrosis. Related to Fig 1. (A) *RPP7* cluster in the Col-0 reference genome. The left portion of the cluster consists of three NLR genes, *At1g58390*, *At1g58400* and *At1g58410* (green arrows). The right portion includes five NLR genes, *At1g58602*, *At1g58807*, *At1g58848*, *At1g59124* and *At1g59218* (brown arrows). Twenty-two non-NLR genes in this region are not shown. (B) Maximum-likelihood tree of NLR genes in the *RPP7* cluster based on the NB domain. *At1g59124* and *At1g58807* sequences are identical, as are *At1g59218* and *At1g58848*. Same colors as in (A). Bootstrap values (out of 100) are indicated on each branch. (C) Representative rescue experiment using an amiRNA construct targeting *RPP7* homologs (see S1 Table). ICE79 was transformed with the amiRNA construct EK21 and T1 plants were crossed to Don-0, resulting in rescued and non-rescued plants segregating in the F1 progeny. Parental genotypes were confirmed with CAPS markers, shown below. Five-week old plants grown in 16˚C are shown.
(TIF)

S2 Fig. Phenotypic variation in Lerik1-3 F_1 hybrids. Related to Fig 1. Major differences were observed in rosette size of F_1 hybrids (A) and spotted cell death on the abaxial side of leaves (B). Scale bar represents 1cm (A) and 1mm (B). Plants were five weeks old.
(TIF)

S3 Fig. *HR4* and *RPW8.1* CRISPR/Cas9 knockout lines. Related to Fig 3 and S4 Fig. (A) Two alleles of *HR4* in Col-0 with a 1-bp insertion (#8/18) or a 19-bp deletion (#8/6) were identified by amplicon sequencing. (B) An allele of *RPW8.1* in KZ10 with a 1-bp insertion was recovered. The stop codons are marked with an asterisk and the first amino acid after a frameshifting

event is in bold.
(TIF)

S4 Fig. Resistance and susceptibility to *H. arabidopsidis* isolate Hiks1. (A) Trypan Blue stained cotyledons 5 days after infection. Lerik1-3 is resistant, while Fei-0 and ICE106 are fully susceptible. The F_1 hybrids Lerik1-3 x Fei-0 and Lerik1-3 x ICE106 appear to be less resistant than Lerik1-3. Ws-0 *eds1-1* is a positive infection control. (B) Two different *hr4* loss-of-function alleles (see S3 Fig) are as resistant as Col-0 wild-type plants. *eds1-1* and *rpp7-15* are positive infection controls.
(TIF)

S5 Fig. Hybrid necrosis by introduction of chimeras. Related to Fig 5. Effects of chimeric *HR4* transgenes introduced into Lerik1-3, with negative and positive controls shown to the left and right. Scale bar represents 1cm. Five week-old plants are shown.
(TIF)

S6 Fig. Predicted lengths of *HR4* and *RPW8.1* coding sequences from remapping of short reads from the 1001 Genomes project. Related to Fig 6. (A) *HR4* type assignments based on information from Sanger sequencing. (B) *RPW8.1* type based on information from Sanger sequencing.
(TIF)

S1 Table. Rescue of hybrid necrosis by amiRNAs against *RPP7* homologs. Related to Fig 1. AmiRNAs were designed based on NLR sequences of the *RPP7* cluster in Col-0 (Table S1) using WMD3 (http://wmd3.weigelworld.org/). Constructs were introduced into Mrk-0, Lerik1-3 or ICE79, and T_1 lines were crossed to incompatible parents. Hybrid necrosis was scored at 16˚C. Examples of F_1 plants are shown in S1 Fig.
(TIF)

S2 Table. GWAS hits on chromosome 3 from Lerik1-3 x 80 accessions panel and tagging SNPs present in accessions carrying different HR4 types. Related to Fig 2. Location of *HR4* (*At3g50480*) is 18,733,287 to 18,734,180 bp on chromosome 3 of the reference Col-0 genome. The next protein-coding gene is *At3g50500* (18,741,805 to 18,743,904 bp), with *At3g50490* (18,738,630 to 18,739,261 bp) encoding a transposable element (see Fig 4A). SNPs in bold italics differ from the Col-0 reference.
(TIF)

S3 Table. Rescue effects of amiRNAs targeting *RPW8* homologs. Related to Figs 1 and 3. AmiRNAs were designed based on sequence information of *RPW8*/*HR* clusters from Col-0, Ms-0 and KZ10. Constructs were introduced into Fei-0 or ICE106, and T_1 lines were crossed to the incompatible accession Lerik1-3. Hybrid necrosis was scored at 16˚C. Parental genotypes and the presence of amiRNA constructs were confirmed by PCR genotyping (see Fig 3A).
(TIF)

S4 Table. Resistance to the *H. arabidopsidis* isolate Hiks1. Related to S4 Fig. *strong resistance: no conidiophores; weak resistance: 1–5 conidiophores/cotyledon, with some sporulation; very weak resistance: 6–19 conidiophores/cotyledon, with low to medium sporulation; no resistance: >20 conidiophores/cotyledon, heavy sporulation. †See S1 Table for amiRNA key.
(TIF)

S5 Table. Accessions for *HR4* survey. Related to Fig 6. Covered region indicates the length of $HR4^{Col\text{-}0}$ (894 bp) covered by reads from the 1001 Genomes Project (http://1001genomes.org),

allowing for five mismatches. *HR4* types are categorized according to the number of *RPW8/HR* repeats, and the haplotype is based on the entire *HR4* coding sequence.
(TIF)

S6 Table. Accessions for *RPW8.1* survey. Related to Fig 6.
(TIF)

S7 Table. Hybrid necrosis in F_1 plants of Mrk-0 F_1 crossed to other accessions. Related to Fig 6. Strong hybrid necrosis equals what is observed in KZ10 x Mrk-0 hybrids.
(TIF)

S8 Table. Hybrid necrosis in F_1 plants of Lerik1-3 crossed to other accessions. Related to Fig 6. Strong hybrid necrosis equals what is observed in Lerik1-3 x Fei-0 F_1 hybrids.
(TIF)

S9 Table. Accessions and hybrids in which growth was analyzed with the automated phenotyping platform RAPA. Related to Fig 6.
(TIF)

S10 Table. Constructs.
(TIF)

S11 Table. Oligonucleotides used for amplifying *RPW8.1/HR4* genomic fragments and swap constructs. Related to Figs 3 and 5.
(TIF)

Acknowledgments

We thank Jane Parker for the *H. arabidopsidis* Hiks1 isolate, Katrin Fritschi and Camilla Kleinhempel for technical support, and Gautam Shirsekar for help with the pathology assays and discussions.

Author Contributions

Conceptualization: Cristina A. Barragan, Detlef Weigel, Eunyoung Chae.

Data curation: Eunyoung Chae.

Formal analysis: Cristina A. Barragan, Sang-Tae Kim, Jörg Hagmann, Anna-Lena Van de Weyer, George Wang, Ilja Bezrukov, Eunyoung Chae.

Funding acquisition: Rui Wu, Detlef Weigel, Eunyoung Chae.

Investigation: Cristina A. Barragan, Rui Wu, Wanyan Xi, Anette Habring, Maricris Zaidem, William Wing Ho Ho, Eunyoung Chae.

Methodology: Cristina A. Barragan, Eunyoung Chae.

Project administration: Detlef Weigel.

Supervision: Detlef Weigel.

Validation: Cristina A. Barragan, Eunyoung Chae.

Writing – original draft: Cristina A. Barragan, Eunyoung Chae.

Writing – review & editing: Detlef Weigel.

References

1. Bomblies K, Weigel D. Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. Nat Rev Genet. 2007; 8: 382–393. https://doi.org/10.1038/nrg2082 PMID: 17404584
2. Chen C, Zhiguo E, Lin H-X. Evolution and Molecular Control of Hybrid Incompatibility in Plants. Front Plant Sci. 2016; 7: 1135.
3. Vaid N, Laitinen RAE. Diverse paths to hybrid incompatibility in Arabidopsis. Plant J. 2019; 97: 199–213. https://doi.org/10.1111/tpj.14061 PMID: 30098060
4. Krüger J, Thomas CM, Golstein C, Dixon MS, Smoker M, Tang S, et al. A tomato cysteine protease required for Cf-2-dependent disease resistance and suppression of autonecrosis. Science. 2002; 296: 744–747. https://doi.org/10.1126/science.1069288 PMID: 11976458
5. Bomblies K, Lempe J, Epple P, Warthmann N, Lanz C, Dangl JL, et al. Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. PLoS Biol. 2007; 5: e236. https://doi.org/10.1371/journal.pbio.0050236 PMID: 17803357
6. Jeuken MJW, Zhang NW, McHale LK, Pelgrom K, den Boer E, Lindhout P, et al. Rin4 causes hybrid necrosis and race-specific resistance in an interspecific lettuce hybrid. Plant Cell. 2009; 21: 3368–3378. https://doi.org/10.1105/tpc.109.070334 PMID: 19855048
7. Alcázar R, García AV, Parker JE, Reymond M. Incremental steps toward incompatibility revealed by Arabidopsis epistatic interactions modulating salicylic acid pathway activation. Proc Natl Acad Sci U S A. 2009; 106: 334–339. https://doi.org/10.1073/pnas.0811734106 PMID: 19106299
8. Alcázar R, García AV, Kronholm I, de Meaux J, Koornneef M, Parker JE, et al. Natural variation at Strubbelig Receptor Kinase 3 drives immune-triggered incompatibilities between Arabidopsis thaliana accessions. Nat Genet. 2010; 42: 1135–1139. https://doi.org/10.1038/ng.704 PMID: 21037570
9. Yamamoto E, Takashi T, Morinaka Y, Lin S, Wu J, Matsumoto T, et al. Gain of deleterious function causes an autoimmune response and Bateson-Dobzhansky-Muller incompatibility in rice. Mol Genet Genomics. 2010; 283: 305–315. https://doi.org/10.1007/s00438-010-0514-y PMID: 20140455
10. Chae E, Bomblies K, Kim S-T, Karelina D, Zaidem M, Ossowski S, et al. Species-wide Genetic Incompatibility Analysis Identifies Immune Genes as Hot Spots of Deleterious Epistasis. Cell. Elsevier; 2014; 159: 1341–1351.
11. Chen C, Chen H, Lin Y-S, Shen J-B, Shan J-X, Qi P, et al. A two-locus interaction causes interspecific hybrid weakness in rice. Nat Commun. 2014; 5: 3357. https://doi.org/10.1038/ncomms4357 PMID: 24556665
12. Todesco M, Kim ST, Chae E, Bomblies K, Zaidem M, Smith LM, et al. Activation of the Arabidopsis thaliana immune system by combinations of common ACD6 alleles. PLoS Genet. 2014; 10: e1004459. https://doi.org/10.1371/journal.pgen.1004459 PMID: 25010663
13. Sicard A, Kappel C, Josephs EB, Lee YW, Marona C, Stinchcombe JR, et al. Divergent sorting of a balanced ancestral polymorphism underlies the establishment of gene-flow barriers in Capsella. Nat Commun. 2015; 6: 7960. https://doi.org/10.1038/ncomms8960 PMID: 26268845
14. Cesari S. Multiple strategies for pathogen perception by plant immune receptors. New Phytol. 2017; https://doi.org/10.1111/nph.14877 PMID: 29131341
15. Zhang X, Dodds PN, Bernoux M. What Do We Know About NOD-Like Receptors in Plant Immunity? Annu Rev Phytopathol. 2017; 55: 205–229. https://doi.org/10.1146/annurev-phyto-080516-035250 PMID: 28637398
16. Monteiro F, Nishimura MT. Structural, Functional, and Genomic Diversity of Plant NLR Proteins: An Evolved Resource for Rational Engineering of Plant Immunity. Annu Rev Phytopathol. 2018; 56: 243–267. https://doi.org/10.1146/annurev-phyto-080417-045817 PMID: 29949721
17. Kourelis J, van der Hoorn RAL. Defended to the Nines: 25 Years of Resistance Gene Cloning Identifies Nine Mechanisms for R Protein Function. Plant Cell. American Society of Plant Biologists; 2018; 30: 285–299.
18. Gos G, Slotte T, Wright SI. Signatures of balancing selection are maintained at disease resistance loci following mating system evolution and a population bottleneck in the genus Capsella. BMC Evol Biol. 2012; 12: 152. https://doi.org/10.1186/1471-2148-12-152 PMID: 22909344
19. Jacob F, Vernaldi S, Maekawa T. Evolution and Conservation of Plant NLR Functions. Front Immunol. 2013; 4: 297. https://doi.org/10.3389/fimmu.2013.00297 PMID: 24093022
20. Karasov TL, Horton MW, Bergelson J. Genomic variability as a driver of plant-pathogen coevolution? Curr Opin Plant Biol. 2014; 18: 24–30. https://doi.org/10.1016/j.pbi.2013.12.003 PMID: 24491596
21. Wicker T, Yahiaoui N, Keller B. Illegitimate recombination is a major evolutionary mechanism for initiating size variation in plant resistance genes. Plant J. 2007; 51: 631–641. https://doi.org/10.1111/j.1365-313X.2007.03164.x PMID: 17573804

22. Nagy ED, Bennetzen JL. Pathogen corruption and site-directed recombination at a plant disease resistance gene cluster. Genome Res. 2008; 18: 1918–1923. https://doi.org/10.1101/gr.078766.108 PMID: 18719093

23. Baggs E, Dagdas G, Krasileva KV. NLR diversity, helpers and integrated domains: making sense of the NLR IDentity. Curr Opin Plant Biol. 2017; 38: 59–67. https://doi.org/10.1016/j.pbi.2017.04.012 PMID: 28494248

24. Trowsdale J. The gentle art of gene arrangement: the meaning of gene clusters. Genome Biol. 2002; 3: COMMENT2002.

25. Lemons D, McGinnis W. Genomic evolution of Hox gene clusters. Science. 2006; 313: 1918–1922. https://doi.org/10.1126/science.1132040 PMID: 17008523

26. Zeng G, Zhang P, Zhang Q, Zhao H, Li Z, Zhang X, et al. Duplication of a Pks gene cluster and subsequent functional diversification facilitate environmental adaptation in Metarhizium species. PLoS Genet. 2018; 14: e1007472. https://doi.org/10.1371/journal.pgen.1007472 PMID: 29958281

27. Xiao S, Ellwood S, Calis O, Patrick E, Li T, Coleman M, et al. Broad-spectrum mildew resistance in Arabidopsis thaliana mediated by RPW8. Science. 2001; 291: 118–120. https://doi.org/10.1126/science.291.5501.118 PMID: 11141561

28. Wang W, Devoto A, Turner JG, Xiao S. Expression of the membrane-associated resistance protein RPW8 enhances basal defense against biotrophic pathogens. Mol Plant Microbe Interact. 2007; 20: 966–976. https://doi.org/10.1094/MPMI-20-8-0966 PMID: 17722700

29. Ma X-F, Li Y, Sun J-L, Wang T-T, Fan J, Lei Y, et al. Ectopic expression of RESISTANCE TO POWDERY MILDEW8.1 confers resistance to fungal and oomycete pathogens in Arabidopsis. Plant Cell Physiol. 2014; 55: 1484–1496. https://doi.org/10.1093/pcp/pcu080 PMID: 24899552

30. Xiao S, Emerson B, Ratanasut K, Patrick E, O'Neill C, Bancroft I, et al. Origin and maintenance of a broad-spectrum disease resistance locus in Arabidopsis. Mol Biol Evol. 2004; 21: 1661–1672. https://doi.org/10.1093/molbev/msh165 PMID: 15155802

31. Orgil U, Araki H, Tangchaiburana S, Berkey R, Xiao S. Intraspecific genetic variations, fitness cost and benefit of RPW8, a disease resistance locus in Arabidopsis thaliana. Genetics. 2007; 176: 2317–2333. https://doi.org/10.1534/genetics.107.070565 PMID: 17565954

32. Wang W, Wen Y, Berkey R, Xiao S. Specific targeting of the Arabidopsis resistance protein RPW8.2 to the interfacial membrane encasing the fungal Haustorium renders broad-spectrum resistance to powdery mildew. Plant Cell. 2009; 21: 2898–2913. https://doi.org/10.1105/tpc.109.067587 PMID: 19749153

33. Berkey R, Zhang Y, Ma X, King H, Zhang Q, Wang W, et al. Homologues of the RPW8 Resistance Protein Are Localized to the Extrahaustorial Membrane that Is Likely Synthesized De Novo. Plant Physiol. 2017; 173: 600–613. https://doi.org/10.1104/pp.16.01539 PMID: 27856916

34. Wróblewski T, Spiridon L, Martin EC, Petrescu A-J, Cavanaugh K, Truco MJ, et al. Genome-wide functional analyses of plant coiled-coil NLR-type pathogen receptors reveal essential roles of their N-terminal domain in oligomerization, networking, and immunity. PLoS Biol. 2018; 16: e2005821. https://doi.org/10.1371/journal.pbio.2005821 PMID: 30540748

35. El Kasmi F, Nishimura MT. Structural insights into plant NLR immune receptor function. Proc Natl Acad Sci U S A. 2016; 113: 12619–12621. https://doi.org/10.1073/pnas.1615933113 PMID: 27803318

36. Collier SM, Hamel L-P, Moffett P. Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. Mol Plant Microbe Interact. 2011; 24: 918–931. https://doi.org/10.1094/MPMI-03-11-0050 PMID: 21501087

37. Zhong Y, Cheng Z-MM. A unique RPW8-encoding class of genes that originated in early land plants and evolved through domain fission, fusion, and duplication. Sci Rep. 2016; 6: 32923. https://doi.org/10.1038/srep32923 PMID: 27678195

38. Bentham AR, Zdrzalek R, De la Concepcion JC, Banfield MJ. Uncoiling CNLs: Structure/Function Approaches to Understanding CC Domain Function in Plant NLRs. Plant Cell Physiol. 2018; 59: 2398–2408. https://doi.org/10.1093/pcp/pcy185 PMID: 30192967

39. Greenwald J, Buhtz C, Ritter C, Kwiatkowski W, Choe S, Maddelein M-L, et al. The mechanism of prion inhibition by HET-S. Mol Cell. 2010; 38: 889–899. https://doi.org/10.1016/j.molcel.2010.05.019 PMID: 20620958

40. Seuring C, Greenwald J, Wasmer C, Wepf R, Saupe SJ, Meier BH, et al. The mechanism of toxicity in HET-S/HET-s prion incompatibility. PLoS Biol. 2012; 10: e1001451. https://doi.org/10.1371/journal.pbio.1001451 PMID: 23300377

41. Daskalov A, Habenstein B, Sabaté R, Berbon M, Martinez D, Chaignepain S, et al. Identification of a novel cell death-inducing domain reveals that fungal amyloid-controlled programmed cell death is related to necroptosis. Proc Natl Acad Sci U S A. 2016; 113: 2720–2725. https://doi.org/10.1073/pnas.1522361113 PMID: 26903619

42. Daskalov A, Habenstein B, Martinez D, Debets AJM, Sabaté R, Loquet A, et al. Signal transduction by a fungal NOD-like receptor based on propagation of a prion amyloid fold. PLoS Biol. 2015; 13: e1002059. https://doi.org/10.1371/journal.pbio.1002059 PMID: 25671553

43. Daskalov A, Dyrka W, Saupe SJ. Theme and variations: evolutionary diversification of the HET-s functional amyloid motif. Sci Rep. 2015; 5: 12494. https://doi.org/10.1038/srep12494 PMID: 26219477

44. McDowell JM, Cuzick A, Can C, Beynon J, Dangl JL, Holub EB. Downy mildew (Peronospora parasitica) resistance genes in Arabidopsis vary in functional requirements for NDR1, EDS1, NPR1 and salicylic acid accumulation. Plant J. 2000; 22: 523–529. PMID: 10886772

45. Tsuchiya T, Eulgem T. An alternative polyadenylation mechanism coopted to the Arabidopsis RPP7 gene through intronic retrotransposon domestication. Proc Natl Acad Sci U S A. 2013; 110: E3535–43. https://doi.org/10.1073/pnas.1312545110 PMID: 23940361

46. Grimm DG, Roqueiro D, Salomé PA, Kleeberger S, Greshake B, Zhu W, et al. easyGWAS: A Cloud-Based Platform for Comparing the Results of Genome-Wide Association Studies. Plant Cell. 2017; 29: 5–19. https://doi.org/10.1105/tpc.16.00551 PMID: 27986896

47. Holub EB, Beynon JL. Symbiology of Mouse-Ear Cress (Arabidopsis Thaliana) and Oomycetes. In: Andrews JH, Tommerup IC, Callow JA, editors. Advances in Botanical Research. Academic Press; 1997. pp. 227–273.

48. Nemri A, Atwell S, Tarone AM, Huang YS, Zhao K, Studholme DJ, et al. Genome-wide survey of Arabidopsis natural variation in downy mildew resistance using combined association and linkage mapping. Proc Natl Acad Sci U S A. 2010; 107: 10302–10307. https://doi.org/10.1073/pnas.0913160107 PMID: 20479233

49. Jorgensen TH, Emerson BC. Functional variation in a disease resistance gene in populations of Arabidopsis thaliana. Mol Ecol. 2008; 17: 4912–4923. https://doi.org/10.1111/j.1365-294X.2008.03960.x PMID: 19140981

50. 1001 Genomes Consortium. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell. 2016; 166: 481–491. https://doi.org/10.1016/j.cell.2016.05.063 PMID: 27293186

51. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol. 2018; 430: 2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007 PMID: 29258817

52. Hildebrand JM, Tanzer MC, Lucet IS, Young SN, Spall SK, Sharma P, et al. Activation of the pseudokinase MLKL unleashes the four-helix bundle domain to induce membrane localization and necroptotic cell death. Proc Natl Acad Sci U S A. 2014; 111: 15072–15077. https://doi.org/10.1073/pnas.1408987111 PMID: 25288762

53. Su L, Quade B, Wang H, Sun L, Wang X, Rizo J. A plug release mechanism for membrane permeation by MLKL. Structure. 2014; 22: 1489–1500. https://doi.org/10.1016/j.str.2014.07.014 PMID: 25220470

54. Chen X, Li W, Ren J, Huang D, He W-T, Song Y, et al. Translocation of mixed lineage kinase domain-like protein to plasma membrane leads to necrotic cell death. Cell Res. 2014; 24: 105–121. https://doi.org/10.1038/cr.2013.171 PMID: 24366341

55. Murphy JM, Czabotar PE, Hildebrand JM, Lucet IS, Zhang J-G, Alvarez-Diaz S, et al. The pseudokinase MLKL mediates necroptosis via a molecular switch mechanism. Immunity. 2013; 39: 443–453. https://doi.org/10.1016/j.immuni.2013.06.018 PMID: 24012422

56. Wang H, Sun L, Su L, Rizo J, Liu L, Wang L-F, et al. Mixed lineage kinase domain-like protein MLKL causes necrotic membrane disruption upon phosphorylation by RIP3. Mol Cell. 2014; 54: 133–146. https://doi.org/10.1016/j.molcel.2014.03.003 PMID: 24703947

57. Yang X, Wang W, Coleman M, Orgil U, Feng J, Ma X, et al. Arabidopsis 14-3-3 lambda is a positive regulator of RPW8-mediated disease resistance. Plant J. 2009; 60: 539–550. https://doi.org/10.1111/j.1365-313X.2009.03978.x PMID: 19624472

58. Jones JDG, Vance RE, Dangl JL. Intracellular innate immune surveillance devices in plants and animals. Science. 2016; 354. https://doi.org/10.1126/science.aaf6395 PMID: 27934708

59. Wang W, Zhang Y, Wen Y, Berkey R, Ma X, Pan Z, et al. A comprehensive mutational analysis of the Arabidopsis resistance protein RPW8.2 reveals key amino acids for defense activation and protein targeting. Plant Cell. 2013; 25: 4242–4261. https://doi.org/10.1105/tpc.113.117226 PMID: 24151293

60. Phizicky EM, Fields S. Protein-protein interactions: methods for detection and analysis. Microbiol Mol Biol Rev. American Society for Microbiology; 1995; 59: 94–123.

61. Xiao S, Calis O, Patrick E, Zhang G, Charoenwattana P, Muskett P, et al. The atypical resistance gene, RPW8, recruits components of basal defence for powdery mildew resistance in Arabidopsis. Plant J. 2005; 42: 95–110. https://doi.org/10.1111/j.1365-313X.2005.02356.x PMID: 15773856

62. Thomma BP, Nürnberger T, Joosten MH. Of PAMPs and effectors: the blurred PTI-ETI dichotomy. Plant Cell. 2011; 23: 4–15. https://doi.org/10.1105/tpc.110.082602 PMID: 21278123

63. Vasseur F, Bresson J, Wang G, Schwab R, Weigel D. Image-based methods for phenotyping growth dynamics and fitness components in Arabidopsis thaliana. Plant Methods. 2018; 14: 63. https://doi.org/10.1186/s13007-018-0331-6 PMID: 30065776

64. Cheng MM, Prisacariu VA, Zheng S, Torr PHS, Rother C. DenseCut: Densely Connected CRFs for Realtime GrabCut. Comput Graph Forum. 2015; 34: 193–201.

65. Schwab R, Ossowski S, Riester M, Warthmann N, Weigel D. Highly specific gene silencing by artificial microRNAs in Arabidopsis. Plant Cell. 2006; 18: 1121–1133. https://doi.org/10.1105/tpc.105.039834 PMID: 16531494

66. Hellens RP, Edwards EA, Leyland NR, Bean S, Mullineaux PM. pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. Plant Mol Biol. 2000; 42: 819–832. PMID: 10890530

67. Weigel D, Glazebrook J. Arabidopsis: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2002.

68. Lampropoulos A, Sutikovic Z, Wenzl C, Maegele I, Lohmann JU, Forner J. GreenGate—a novel, versatile, and efficient cloning system for plant transgenesis. PLoS One. 2013; 8: e83043. https://doi.org/10.1371/journal.pone.0083043 PMID: 24376629

69. Emami S, Yee M-C, Dinneny JR. A robust family of Golden Gate Agrobacterium vectors for plant synthetic biology. Front Plant Sci. 2013; 4: 339. https://doi.org/10.3389/fpls.2013.00339 PMID: 24032037

70. Wu R, Lucke M, Jang Y-T, Zhu W, Symeonidi E, Wang C, et al. An efficient CRISPR vector toolbox for engineering large deletions in Arabidopsis thaliana. Plant Methods. 2018; 14: 65. https://doi.org/10.1186/s13007-018-0330-7 PMID: 30083222

71. Rowan BA, Seymour DK, Chae E, Lundberg DS, Weigel D. Methods for Genotyping-by-Sequencing. In: White SJ, Cantsilieris S, editors. Genotyping: Methods and Protocols. New York, NY: Springer New York; 2017. pp. 221–242.

72. Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D. Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Res. 2008; 18: 2024–2033. https://doi.org/10.1101/gr.080200.108 PMID: 18818371

73. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30: 1312–1313. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623

Supplemental Information

RPW8/HR Repeats Predict NLR-dependent Hybrid Performance

Cristina A. Barragan, Rui Wu, Sang-Tae Kim, Wanyan Xi, Anette Habring, Jörg Hagmann, Anna-Lena Van de Weyer, Maricris Zaidem, William Wing Ho Ho, George Wang, Ilja Bezrukov, Detlef Weigel, Eunyoung Chae

Supplemental Figures

Supplemental Tables

Supplemental Experimental Procedures

Supplemental References

Supplemental Figures

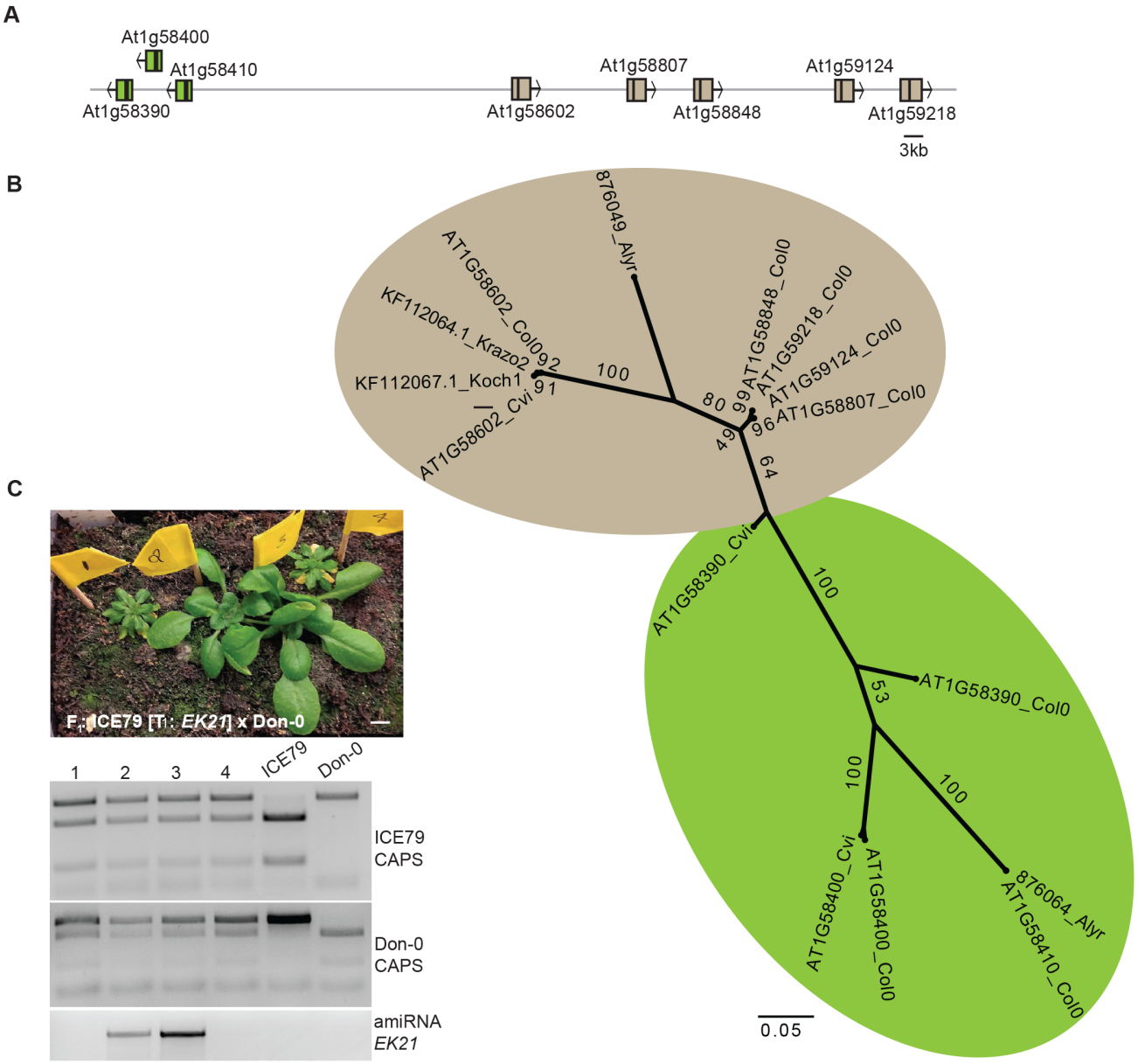


Fig S1. Role of the *RPP7* cluster in *DM6-DM7* dependent hybrid necrosis. Related to Fig 1.
(A) *RPP7* cluster in the Col-0 reference genome. The left portion of the cluster consists of three *NLR* genes, *At1g58390*, *At1g58400* and *At1g58410* (green arrows). The right portion includes five *NLR* genes, *At1g58602*, *At1g58807*, *At1g58848*, *At1g59124* and *At1g59218* (brown arrows). Twenty-two non-NLR genes in this region are not shown. **(B)** Maximum-likelihood tree of *NLR* genes in the *RPP7* cluster based on the NB domain. *At1g59124* and *At1g58807* sequences are identical, as are *At1g59218* and *At1g58848*. Same colors as in (A). Bootstrap values (out of 100) are indicated on each branch. **(C)** Representative rescue experiment using an amiRNA construct targeting *RPP7* homologs (see Table S1). ICE79 was transformed with the amiRNA construct *EK21* and T_1 plants were crossed to Don-0, resulting in rescued and non-rescued plants segregating in the F_1 progeny. Parental genotypes were confirmed with CAPS markers, shown below. Five-week old plants grown in 16°C are shown.

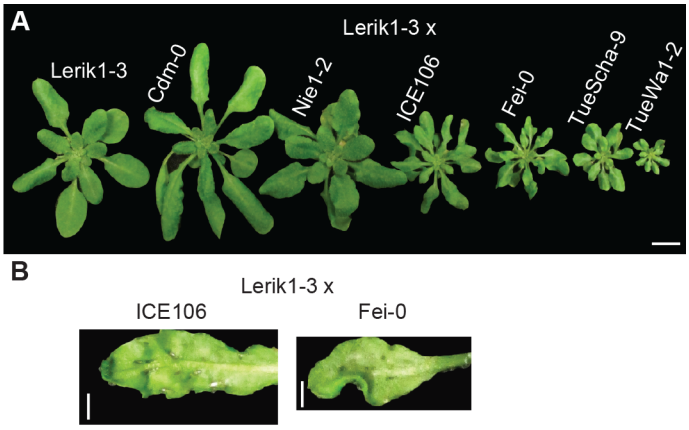


Fig S2. Phenotypic variation in Lerik1-3 F_1 hybrids. Related to Fig 1.
Major differences were observed in rosette size of F_1 hybrids **(A)** and spotted cell death on the abaxial side of leaves **(B)**. Scale bar represents 1cm **(A)** and 1mm **(B)**. Plants were five weeks old.

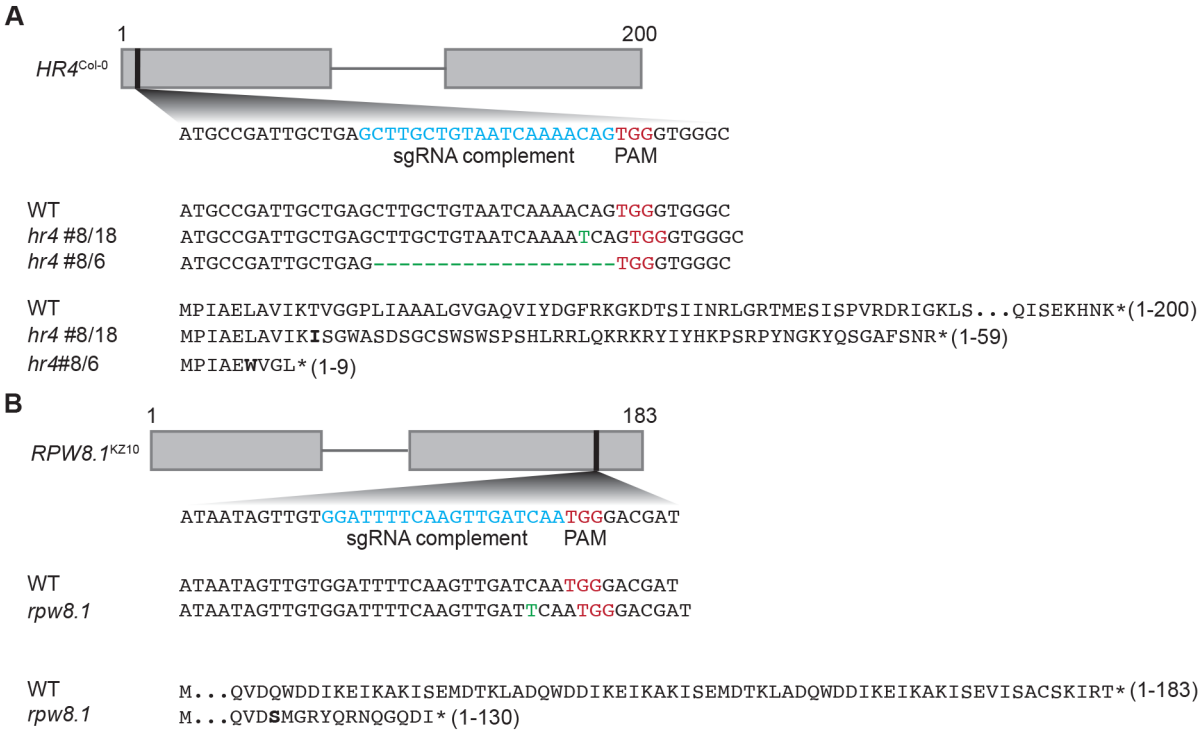


Fig S3. *HR4* and *RPW8.1* CRISPR/Cas9 knockout lines. Related to Fig 3 and Fig S4.
(A) Two alleles of *HR4* in Col-0 with a 1-bp insertion (#8/18) or a 19-bp deletion (#8/6) were identified by amplicon sequencing. **(B)** An allele of *RPW8.1* in KZ10 with a 1-bp insertion was recovered. The stop codons are marked with an asterisk and the first amino acid after a frameshifting event is in bold.

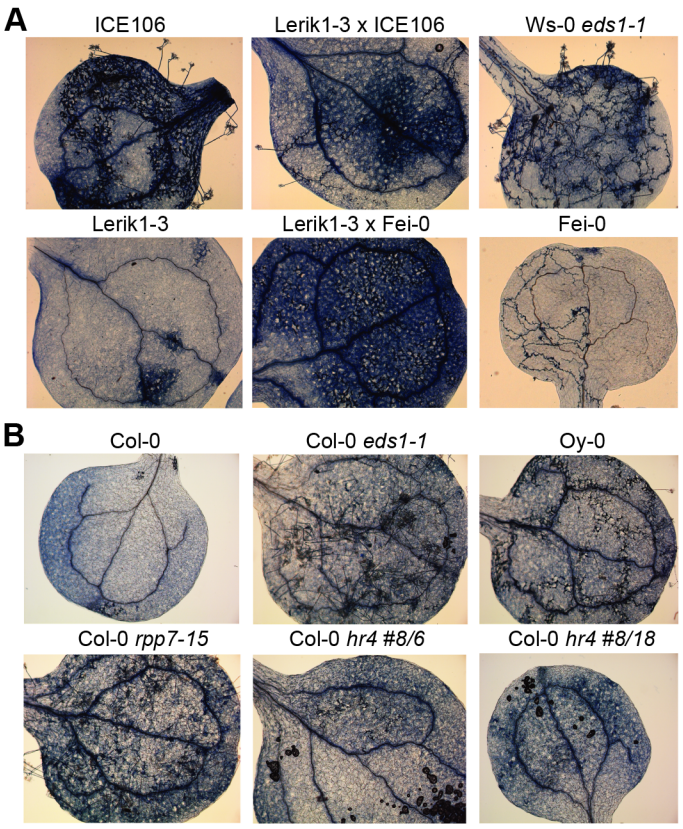


Fig S4. Resistance and susceptibility to *H. arabidopsidis* isolate Hiks1.
(A) Trypan Blue stained cotyledons 5 days after infection. Lerik1-3 is resistant, while Fei-0 and ICE106 are fully susceptible. The F_1 hybrids Lerik1-3 x Fei-0 and Lerik1-3 x ICE106 appear to be less resistant than Lerik1-3. Ws-0 *eds1-1* is a positive infection control. **(B)** Two different *hr4* loss-of-function alleles (see Fig S3) are as resistant as Col-0 wild-type plants. *eds1-1* and *rpp7-15* are positive infection controls.

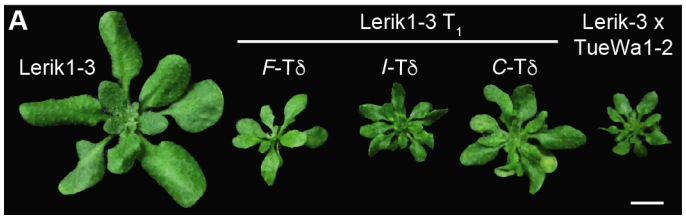


Fig S5. Hybrid necrosis by introduction of chimeras. Related to Fig 5.
Effects of chimeric *HR4* transgenes introduced into Lerik1-3, with negative and positive controls shown to the left and right. Scale bar represents 1cm. Five week-old plants are shown.

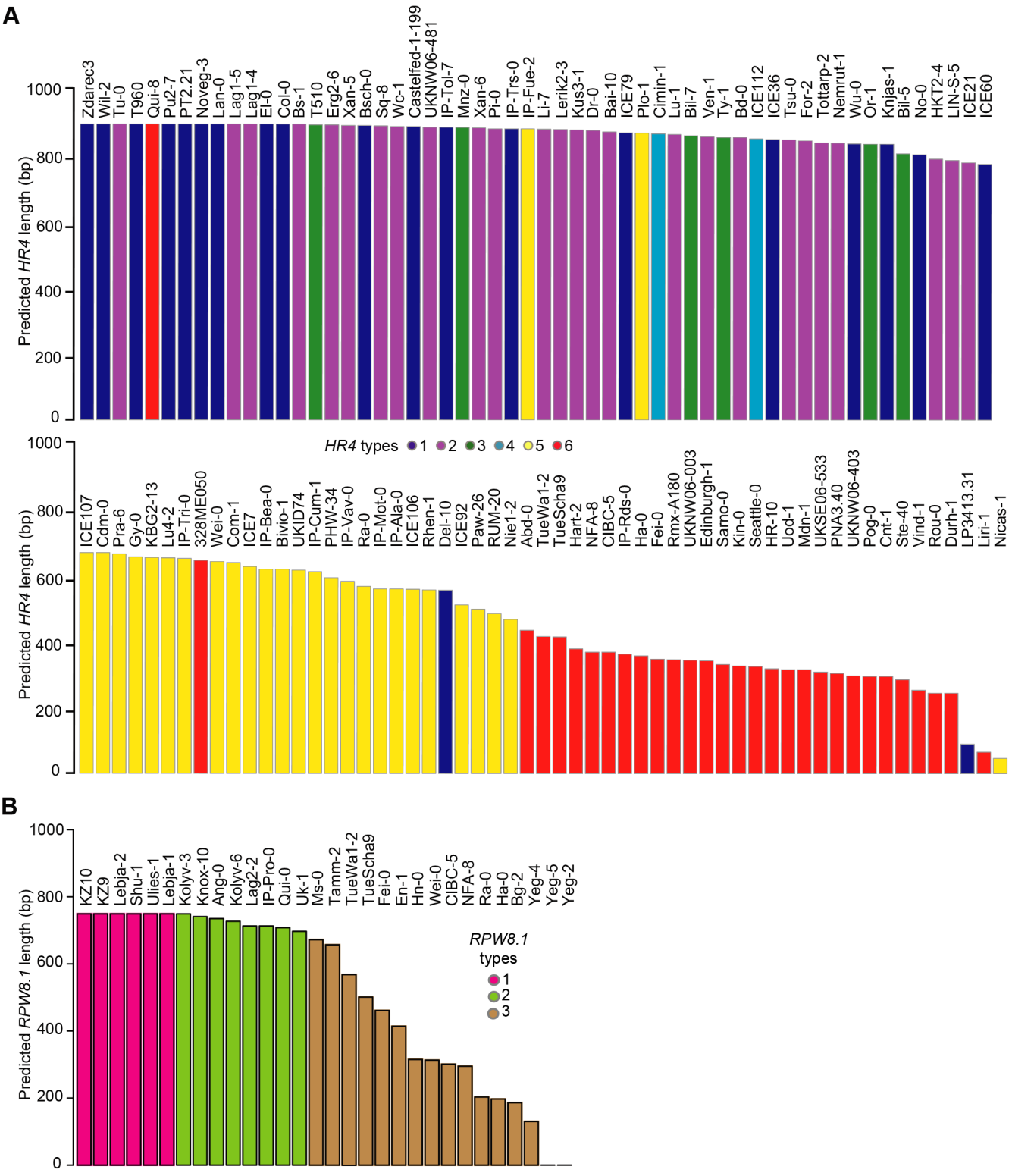


Fig S6. Predicted lengths of *HR4* and *RPW8.1* coding sequences from remapping of short reads from the 1001 Genomes Project. Related to Fig 6.

(A) *HR4* type assignments based on information from Sanger sequencing. **(B)** *RPW8.1* type based on information from Sanger sequencing.

Supplemental Tables

Table S1. Rescue of hybrid necrosis by amiRNAs against *RPP7* homologs. Related to Fig 1.

amiRNA	Sequence	Predicted Col-0 targets	Rescue			
			Mrk-0 x KZ10	Lerik1-3 x Fei-0	Lerik1-3 x ICE106	ICE79 x Don-0
EK19	TAAATGACCATATTCCTGCTC	*AT1G58602*, *AT1G59218*, *AT1G58807, AT1G58848*, *AT1G59124*	yes	yes	yes	yes
EK20	TTTTCCAGGTATTTCAGTCAA	*AT1G58602*, *AT1G59218*, *AT1G58848*, *AT1G58807*, *AT1G59124*	partial	yes	yes	yes
EK21	TCGAGGTATTTCAATCCGCTT	*AT1G58602*, *AT1G59218*, *AT1G58807*, *AT1G58848*, *AT1G95124*	yes	yes	yes	yes
EK22	TAAAGTTAGTTCTTGCTCCCA	*AT1G58602*, *AT1G59218*, *AT1G58807, AT1G58848*, *AT1G59124*	yes	yes	not tested	not tested
EK26	TTAGATCACGTTTTAGCCCAG	*AT1G58390*	no	no	no	no
EK27	TATGTCTAGATAGATCGGCAA	*AT1G58400*, 3' portion	no	no	not tested	not tested
EK28	TAAGTTAGTTTTGTGATGCGC	*AT1G58400*, 5' portion	partial	no	not tested	not tested
EK29	TCTTAATTCATGCATCCGCAT	*AT1G58390*, 3' portion	no	no	no	no
EK30	TATATCAGACGCAAGTTCCCT	*AT1G58410*	no	no	not tested	not tested
EK31	TAAAGTCGCTTTCGTAGCCGC	*AT1G58390*, 5' portion	no	no	not tested	not tested

AmiRNAs were designed based on NLR sequences of the *RPP7* cluster in Col-0 (Table S1) using WMD3 (http://wmd3.weigelworld.org/). Constructs were introduced into Mrk-0, Lerik1-3 or ICE79, and T_1 lines were crossed to incompatible parents. Hybrid necrosis was scored at 16°C. Examples of F_1 plants are shown in Fig S1.

Table S2. GWAS hits on chromosome 3 from Lerik1-3 x 80 accessions panel and tagging SNPs present in accessions carrying different HR4 types. Related to Fig 2.

		HR4 type			
		1	2	5	6
Position	-log(P-value)	No-0	ICE21	ICE106	TueWa1-2
18,743,818	26.95	A	A	***G***	***G***
18,745,470	11.22	C	***T***	***T***	***T***
18,741,733	8.67	C	C	C	C
18,742,020	8.67	G	G	G	G
18,742,285	8.67	G	G	G	G
18,742,408	8.67	C	C	C	C
18,742,741	8.67	G	G	G	G
18,742,864	8.67	T	T	T	T
18,741,725	8.08	A	A	A	A
18,742,108	8.08	A	*NA*	A	A
18,742,821	8.08	C	C	C	C

Location of *HR4* (*At3g50480*) is 18,733,287 to 18,734,180 bp on chromosome 3 of the reference Col-0 genome. The next protein-coding gene is *At3g50500* (18,741,805 to 18,743,904 bp), with *At3g50490* (18,738,630 to 18,739,261 bp) encoding a transposable element (see Fig 4A). SNPs in bold italics differ from the Col-0 reference.

Table S3. Rescue effects of amiRNAs targeting *RPW8* homologs. Related to Fig 1 and Fig 3.

amiRNA	Sequence	Predicted targets	Predicted non-targets	Rescue		
				Lerik1-3 x Fei-0	Lerik1-3 x ICE106	Lerik1-3 x TueWa1-2
MZ110	TTCAAGGAAACA CGTGAGACG	*RPW8.1a*, *RPW8.1b*, *HR4*	*RPW8.2*	partial	partial	not tested
MZ137	TGATACTAATGA TTGTAGCGC	*RPW8.1b*	*RPW8.1a*, *HR4*	no	not tested	not tested
MZ141	TCAGAACGTAAA TCGGATCGC	*RPW8.2* homolog	*RPW8.1*, *HR4*	no	not tested	not tested
KB amiR-*RPW8.1 b*	TATGATTGTAGC GCAGAGACG	*RPW8.1a*, *RPW8.1b*	*HR4*	no	not tested	not tested
ACB HR4.2	TCTTAATTCATGC ATCCGCAT	*HR4*	*RPW8.1a*, *RPW8.1b*	yes	yes	yes

AmiRNAs were designed based on sequence information of *RPW8*/*HR* clusters from Col-0, Ms-0 and KZ10. Constructs were introduced into Fei-0 or ICE106, and T_1 lines were crossed to the incompatible accession Lerik1-3. Hybrid necrosis was scored at 16°C. Parental genotypes and the presence of amiRNA constructs were confirmed with by PCR genotyping (see Fig 3A).

Table S4. Resistance to the *H. arabidopsidis* isolate Hiks1. Related to Fig S4.

	Accession / F_1 cross	**Hiks1 resistance***
Accessions	Col-0	weak resistance
	Lerik1-3	strong resistance
	ICE106	no resistance
	Fei-0	no resistance
	Ws-0	no resistance
	Ws-0 *eds1-1*	no resistance
	Oy-0	no resistance
	Col-0 *rpp7-15*	no resistance
Col-0 *HR4 CRISPR/Cas9 KOs*	Col-0 *hr4* #6/8	strong resistance
	Col-0 *hr4* #8/18	strong resistance
Lerik1-3 amiR-*RPP7* transgenics†	*EK29*	strong resistance
	EK28	strong resistance
	EK27	strong resistance
	EK26	strong resistance
	EK22	strong resistance
	EK21	strong resistance
	EK19	strong resistance
F_1 hybrids	Lerik1-3 x Fei-0	strong resistance
	Lerik1-3 x ICE106	weak resistance

*strong resistance: no conidiophores; weak resistance: 1-5 conidiophores/cotyledon, with some sporulation; very weak resistance: 6-19 conidiophores/cotyledon, with low to medium sporulation; no resistance: >20 conidiophores/cotyledon, heavy sporulation.

†See Table S1 for amiRNA key.

Table S5. Accessions for *HR4* survey. Related to Fig 6.

Accession	1001 Genomes Project ID	Stock center ID	*HR4* type	*HR4* haplotype	Covered region (bp)
IP-Tol-7	9588	CS77371	1	XV	885
IP-Trs-0	9590	CS77387	1	XV	880
El-0	7117	CS76479	1	XIV	894
Wu-0	7415	CS78858	1	XIV	835
Castelfed-1-199	9683	CS76748	1	XIII	887
ICE79/Voeran-1	9979	CS76352	1	XIII	868
T960	6148	CS77325	1	XII	894
ICE60/Stepn-2	9955	CS76377	1	XI	773
ICE36/Dobra-1	10018	CS76369	1	X	848
No-0	7273	CS77128	1	X	802
Knjas-1	9749	CS76971	1	X	834
LP3413.31	8464	CS79030	1	IX	894
Zdarec3	403	CS78873	1	IX	894
Del-10	10016	CS76397	1	IX	554
Lan-0	7208	CS76539	1	IX	894
Col-0	6909	CS76539	1	IX	894
Noveg-3	9638	CS77133	1	IX	894
Pu2-7	6956	CS76580	1	IX	894
Wil-2	7413	CS78856	1	IX	894
PT2.21	8077	CS77191	1	IX	894
Bsch-0	7031	CS76457	1	IX	890
Wc-1	7404	CS76627	2	V	887
Ven-1	7384	CS76624	2	V	856
UKNW06-481	5644	CS78798	2	V	885
Tu-0	7375	CS76617	2	V	894
ICE21/Petro-1	10017	CS76370	2	V	778
Lu-1	8334	CS77056	2	V	863
LIN-S-5	915	CS77040	2	V	785
Tsu-0	7373	CS77389	2	V	847
Bd-0	7013	CS76445	2	V	854
Bai-10	9779	CS76682	2	V	871
Kus3-1	9802	CS76991	2	V	877
Lag1-4	9102	CS76999	2	IV	894
Lag1-5	9103	CS77000	2	IV	894
HKT2-4	9995	CS76404	2	V	789
Dr-0	7106	CS78897	2	V	875
Pi-0	7298	CS76572	2	V	880
For-2	5741	CS78783	2	V	844
Erg2-6	9784	CS76845	2	V	892
Bay-0	6899	CS22633	2	V	-
Sq-8	6967	CS76604	2	V	889

Li-7	7231	CS77035	2	VI	879
Bs-1	7003	CS78888	2	V	894
Tottarp-2	6243	CS77381	2	V	838
Xan-6	9070	CS78862	2	VII	883
Nemrut-1	9993	CS76398	2	VII	837
Lerik2-3	9081	CS77026	2	VII	878
Xan-5	9069	CS78861	2	VII	890
Ty-1	7351	CS78790	3	III	854
Or-1	6074	CS77150	3	III	834
Bil-5	6900	CS76709	3	III	805
T510	6109	CS77301	3	III	892
Mnz-0	7244	CS76552	3	III	884
Bil-7	6901	CS76710	3	III	859
Cimin-1	9661	CS76771	4	VIII	865
ICE112/Moran-1	9967	CS76427	4	VIII	850
Bivio-1	9649	CS76713	5	II	618
Cdm-0	9943	CS76410	5	II	668
IP-Mot-0	9560	CS77109	5	II	558
Nicas-1	9658	CS77127	5	II	46
Wei-0	6979	CS76628	5	II	641
Rhen-1	7316	CS78916	5	II	555
ICE106/Mammo-1	9964	CS76365	5	II	557
ICE92/Angit-1	9981	CS76366	5	II	510
IP-Bea-0	9522	CS76695	5	II	618
IP-Ala-0	9515	CS76650	5	II	558
IP-Cum-1	9537	CS76787	5	II	610
Paw-26	2171	CS77164	5	II	497
IP-Vav-0	9511	CS78835	5	II	581
KBG2-13	9770	CS76966	5	II	653
ICE107/Mammo-2	9965	CS76364	5	II	668
Com-1	7092	CS76469	5	II	638
UKID74	5779	CS78789	5	II	615
PHW-34	8244	CS77174	5	II	592
PLO-1	9923	CS77180	5	II	867
IP-Tri-0	9900	CS77386	5	II	650
ICE7/Lecho-1	9987	CS76371	5	II	626
RUM-20	9925	CS77226	5	II	483
Lu4-2	9792	CS77058	5	II	652
Gy-0	8214	CS78901	5	II	655
Ra-0	6958	CS76582	5	II	566
IP-Fue-2	9541	CS76871	5	II	880
Nie1-2	9996	CS76402	5	II	466
Pra-6	9948	CS76416	5	II	664
328ME059	8584	CS76641	6	I	644
Abd-0	6986	CS76429	6	I	433

CIBC-5	6908	CS78894	6	I	367
Cnt-1	5726	CS78782	6	I	294
Durh-1	7107	CS76477	6	I	243
Edinburgh-1	9298	CS76832	6	I	341
Fei-0	9941	CS76412	6	I	346
Ha-0	7163	CS76500	6	I	356
Hart-2	9799	CS76913	6	I	377
HR-10	6923	CS76940	6	I	317
IP-Rds-0	9573	CS77206	6	I	361
Kin-0	6926	CS76527	6	I	325
Liri-1	9654	CS77041	6	I	65
Mdn-1	1829	CS77077	6	I	314
NFA-8	6944	CS78913	6	I	367
PNA3.40	7947	CS77184	6	I	303
Pog-0	7306	CS76576	6	I	294
QUI-8	9934	CS77199	6	I	894
Rmx-A180	7525	CS77218	6	I	344
Rou-0	7320	CS76591	6	I	243
Sarno-0	9660	CS77236	6	I	330
Seattle-0	7332	CS76598	6	I	324
Ste-40	2317	CS77278	6	I	284
TueScha-9	10000	CS76401	6	I	413
TueWa1-2	10002	CS76405	6	I	414
UKNW06-003	5353	CS78792	6	I	343
UKNW06-403	5577	CS78797	6	I	296
UKSE06-533	5276	CS78806	6	I	307
Uod-1	6975	CS76621	6	I	314
Vind-1	7387	CS76625	6	I	252

Covered region indicates the length of *HR4*$^{Col-0}$ (894 bp) covered by reads from the 1001 Genomes Project (http://1001genomes.org), allowing for five mismatches. *HR4* types are categorized according to the number of *RPW8*/*HR* repeats, and the haplotype is based on the entire *HR4* coding sequence.

Table S6. Accessions for *RPW8.1* survey. Related to Fig 6.

Accession	1001 Genomes Project ID	Stock center ID	*RPW8.1* type	Covered region (bp)
KZ-10	10019	CS28435	1	749
KZ-9	6931	CS76537	1	749
Lebja-2	9632	CS77016	1	749
Shu-1	14318	CS78930	1	749
Ulies-1	9737	CS78815	1	749
Lebja-1	9631	CS77015	1	749
Kolyv-3	9626	CS76978	1	749
Knox-10	6927	CS76973	2	741
Ang-0	6992	CS76436	2	735
Kolyv-6	9628	CS76980	2	727
Lag2-2	9990	CS76390	2	713
IP-Pro-0	9571	CS78914	2	713
Qui-0	9949	CS76417	2	708
Uk-1	7378	CS76620	2	697
Ms-0	6938	CS76555	3	672
Tamm-2	6968	CS76610	3	657
TueWa1-2	10002	CS76405	3	568
TueScha-9	10000	CS76401	3	501
Fei-0	9941	CS76412	3	461
En-1	8290	CS76841	3	414
Hn-0	7165	CS76513	3	315
CIBC-5	6908	CS78894	3	301
NFA-8	6944	CS78913	3	295
Ha-0	7163	CS76500	3	197
Bg-2	6709	CS28069	3	186
Yeg-4	9130	CS78865	3	130
Yeg-5	9131	CS78866	3	0
Yeg-2	9128	CS78864	3	0

Table S7. Hybrid necrosis in F_1 plants of Mrk-0 crossed to other accessions. Related to Fig 6.

Cross	*RPW8.1* type	Hybrid necrosis
Mrk-0 x KZ-10	1	strong
Mrk-0 x KZ-9	1	strong
Mrk-0 x Lebja-2	1	strong
Mrk-0 x Shu-1	1	intermediate
Mrk-0 x Ulies-1	1	strong
Mrk-0 x Lebja-1	1	intermediate
Mrk-0 x Kolyv-3	1	strong
Mrk-0 x Knox-10	2	none
Mrk-0 x Ang-0	2	none
Mrk-0 x Kolyv-6	2	none
Mrk-0 x Lag2-2	2	none
Mrk-0 x IP-Pro-0	2	none
Mrk-0 x Qui-0	2	none
Mrk-0 x Uk-1	2	none
Mrk-0 x Ms-0	3	none
Mrk-0 x Tamm-2	3	none
TueWa1-2 x Mrk-0	3	none
TueScha-9 x Mrk-0	3	none
Fei-0 x Mrk-0	3	none
Mrk-0 x En-1	3	none
Mrk-0 x Hn-0	3	none
Mrk-0 x Wei-0	3	none
Mrk-0 x CIBC-5	3	none
Mrk-0 x NFA-8	3	none
Mrk-0 x Ra-0	3	none
Mrk-0 x Ha-0	3	none
Mrk-0 x Bg-2	3	none
Yeg-4 x Mrk-0	3	none
Mrk-0 x Yeg-5	3	none
Mrk-0 x Yeg-2	3	none

Strong hybrid necrosis equals what is observed in KZ10 x Mrk-0 hybrids.

Table S8. Hybrid necrosis in F_1 plants of Lerik1-3 crossed to other accessions. Related to Fig 6.

Cross	*HR4* type	Hybrid necrosis
Lerik1-3 x Col-0	1	none
Lerik1-3 x Noveg-3	1	none
Lerik1-3 x Lerik2-3	2	none
Lerik1-3 x Ven-1	2	none
Ty-1 x Lerik1-3	3	none
Lerik1-3 x Bil-5	3	none
Lerik1-3 x Bil-7	3	none
Lerik1-3 x Or-1	3	none
Lerik1-3 x ICE112	4	none
Lerik1-3 x Cimin-1	4	none
Lu4-2 x Lerik1-3	5	mild
Lerik1-3 x Lu4-2	5	mild
Lerik1-3 x Tri-0	5	mild
Nicas-1 x Lerik1-3	5	mild
Lerik1-3 x IP-Cum	5	mild
Lerik1-3 x RUM20	5	mild
Lerik1-3 x PAW26	5	mild
Uod-1 x Lerik1-3	6	intermediate
Lerik1-3 x Liri-1	6	intermediate
Lerik1-3 x Vind-1	6	strong
Ste-40 x Lerik1-3	6	strong
Pog-0 x Lerik1-3	6	strong
Lerik1-3 x RmxA180	6	strong
Lerik1-3 x Edinburgh-1	6	strong
Lerik1-3 x PNA3.40	6	strong

Strong hybrid necrosis equals what is observed in Lerik1-3 x Fei-0 F_1 hybrids.

Table S9. Accessions and hybrids in which growth was analyzed with the automated phenotyping platform RAPA. Related to Fig 6.

Population	*HR4* type
Lerik1-3	-
Lerik1-3 x TueWa1-2	6
TueWa1-2	
TueScha-9 x Lerik1-3	
TueScha-9	
Lerik1-3 x Fei-0	
Fei-0	
Lerik1-3 x ICE106	5
ICE106	
Nie1-2 x Lerik1-3	
Nie1-2	
Lerik1-3 x Cdm-0	
Cdm-0	
Lerik1-3 x HKT2-4	2
HKT2-4	
Lerik1-3 x ICE21	
ICE21	
Nemrut-1 x Lerik1-3	
Nemrut-1	
ICE36 x Lerik1-3	1
ICE36	
ICE60 x Lerik1-3	
ICE60	
ICE79 x Lerik1-3	
ICE79	

Table S10. Constructs.

Construct	Description	Backbone	Promoter	CDS
ACB066	genomic *HR4*	pMLBart	*HR4*$^{Fei-0}$	*HR4*$^{Fei-0}$
ACB067	genomic *HR4*	pMLBart	*HR4*^{ICE106}	*HR4*^{ICE106}
ACB011	genomic *RPW8.1a*	pMLBart	*RPW8.1a*$^{Fei-0}$	*RPW8.1a*$^{Fei-0}$
ACB012	genomic *RPW8.1b*	pMLBart	*RPW8.1b*$^{Fei-0}$	*RPW8.1b*$^{Fei-0}$
ACB068	amiR-*RPW8.1*	pFK210	CaMV 35S	see Table S3
MZ110	amiR-*RPW8.1*	pFK210	CaMV 35S	see Table S3
MZ137	amiR-*RPW8.1*	pFK210	CaMV 35S	see Table S3
MZ141	amiR-*RPW8.1*	pFK210	CaMV 35S	see Table S3
KB amiR-RPW8.1 b	amiR-*RPW8.1*	pFK210	CaMV 35S	see Table S3
ACB074	amiR-*RPP7 EK19*	pFK210	CaMV 35S	see Table S1
ACB075	amiR-*RPP7 EK20*	pFK210	CaMV 35S	see Table S1
ACB076	amiR-*RPP7 EK21*	pFK210	CaMV 35S	see Table S1
ACB077	amiR-*RPP7 EK22*	pFK210	CaMV 35S	see Table S1
ACB078	amiR-*RPP7 EK26*	pFK210	CaMV 35S	see Table S1
ACB079	amiR-*RPP7 EK27*	pFK210	CaMV 35S	see Table S1
ACB080	amiR-*RPP7 EK28*	pFK210	CaMV 35S	see Table S1
ACB081	amiR-*RPP7 EK29*	pFK210	CaMV 35S	see Table S1
ACB082	amiR-*RPP7 EK30*	pFK210	CaMV 35S	see Table S1
ACB083	amiR-*RPP7 EK31*	pFK210	CaMV 35S	see Table S1
ACB042	*HR4* CRISPR/Cas9 Col-0	pRW006	CaMV 35S	see Fig S3
pWX031	*RPW8.1* CRISPR/Cas9 KZ10	pGGZ001	MAS	see Fig S3
pRW016	*K*-0	pMCY2	*RPW8.1*^{KZ10}	*RPW8.1*^{KZ10}
pRW017	*M*-0	pMCY2	*RPW8.1*^{KZ10}	*RPW8.1*$^{Ms-0}$
pRW020	*M*-Aα	pMCY2	*RPW8.1*^{KZ10}	*RPW8.1*$^{Ms-0}$
pRW021	*K*-Aα	pMCY2	*RPW8.1*^{KZ10}	*RPW8.1*^{KZ10}
pRW018	*K*-Bβ	pMCY2	*RPW8.1*^{KZ10}	*RPW8.1*^{KZ10}
pRW019	*K*-Bcβ	pMCY2	*RPW8.1*^{KZ10}	*RPW8.1*^{KZ10}
ACB085	*K*-BBβ	pMCY2	*RPW8.1*^{KZ10}	*RPW8.1*^{KZ10}
ACB086	*K*-BBcβ	pMCY2	*RPW8.1*^{KZ10}	*RPW8.1*^{KZ10}
ACB087	*M*-BBcβ	pMCY2	*RPW8.1*^{KZ10}	*RPW8.1*$^{Ms-0}$
ACB088	*K*-BBBβ	pMCY2	*RPW8.1*^{KZ10}	*RPW8.1*^{KZ10}
ACB089	*K*-BBBcβ	pMCY2	*RPW8.1*^{KZ10}	*RPW8.1*^{KZ10}
ACB045	*C*-O	pMCY2	*HR4*$^{Fei-0}$	*HR4*$^{Col-0}$
ACB044	*I*-0	pMCY2	*HR4*$^{Fei-0}$	*HR4*^{ICE106}
ACB050	*F*-0	pMCY2	*HR4*$^{Fei-0}$	*HR4*$^{Fei-0}$
ACB046	*C*-QQRSRγ	pMCY2	*HR4*$^{Fei-0}$	*HR4*$^{Col-0}$
ACB048	*F*-QQRSRγ	pMCY2	*HR4*$^{Fei-0}$	*HR4*$^{Fei-0}$
ACB058	*F*-QQSRγ	pMCY2	*HR4*$^{Fei-0}$	*HR4*$^{Fei-0}$
ACB069	*F*-QQRSRδ	pMCY2	*HR4*$^{Fei-0}$	*HR4*$^{Fei-0}$
ACB070	*F*-QQRSδ	pMCY2	*HR4*$^{Fei-0}$	*HR4*$^{Fei-0}$
ACB051	*I*-RT	pMCY2	*HR4*$^{Fei-0}$	*HR4*^{ICE106}

ACB054	*F*-RTδ	pMCY2	*HR4*$^{Fei-0}$	*HR4*$^{Fei-0}$
ACB047	*F*-RRTδ	pMCY2	*HR4*$^{Fei-0}$	*HR4*$^{Fei-0}$
ACB071	*F*-RRTγ	pMCY2	*HR4*$^{Fei-0}$	*HR4*$^{Fei-0}$
ACB053	*C*-Tδ	pMCY2	*HR4*$^{Fei-0}$	*HR4*$^{Col-0}$
ACB052	*I*-Tδ	pMCY2	*HR4*$^{Fei-0}$	*HR4*^{ICE106}
ACB065	*F*-Tδ	pMCY2	*HR4*$^{Fei-0}$	*HR4*$^{Fei-0}$

Table S11. Oligonucleotides used for amplifying *RPW8.1/HR4* genomic fragments and swap constructs. Related to Fig 3 and Fig 5.

Primer	Sequence	Purpose
G-41108	GCCACATTGGTCTCTCAATTTGT	PCR *HR4* genomic fragments
G-40245	CTCCATTAATCTGCAAATTTGCTAA	
G-40185	TCTGGGCTAATTCAAATTTCATAT	
G-12714	GACCCGTACAGTACTAAGTCTA	
G-41573	CAATCATTGTTGGGAAGAAGAAAGA	
G-41558	gctcttcaATGCCGATTGGTGAGCTTGC	PCR *RPW8.1* genomic fragments
G-39847	TAGATCATTGTCAGTAAA	
G-37615	AGATAAGCCATAGAACCTCAGTGATAC	Lerik1-3 ClaI CAPS
G-37616	GGTTTGCTGCTTCCTTCTAAATACATT	
G-37354	TAAGTCTTGCATATCAGGCATTTCATC	ICE79 ClaI CAPS
G-37355	TATTTGTAGCTTTAGAAGTTGAGGCTG	
G-36173	CCAATGAACTCTATTTCAGGAATCTGG	Don-0 XhoI CAPS
G-36174	AGAGCGGGAAGAAATATCAGATTAAGA	
G-37109	CTGATGATACTTTGTGATTCCAGGATG	Fei-0 XhoI CAPS
G-37110	AGTTCAATTTCACAGGTCTACCATAGA	
G-36261	ATCAAAGTAAATCACAGGAGCATCATC	ICE106 XbaI CAPS
G-36262	GAGAGCTTTGAAACTGAACAAGAAGTA	
G-41539	gctcttcaagtTGTCCCAATTGTGTCAAACGACTC	*RPW8.1*^{KZ10} promoter
G-41556	gctcttcacatTTTTTTAAAGTAGTTGTTTAGCTCTCGAGG	
G-41558	see above	*RPW8.1*$^{Ms-0}$ CDS truncation
G-41925	gctcttctATCAACTTGAAAATCCACAACTATTATGC	
G-41558	see above	*RPW8.1*^{KZ10} CDS truncation
G-41924	gctcttctATCAACTTGAAAATCCACAACTATTATCC	
G-41558	see above	*RPW8.1*$^{Ms-0}$ CDS truncation +TAG
G-42151	gctcttcaccaCTAATCAACTTGAAAATCCACAACTATTAT	
G-41558	see above	*RPW8.1*^{KZ10} CDS truncation +TAG
G-42150	gctcttcaccaCTAATCAACTTGAAAATCCACAACTATTATCC	
G-41927	gctcttcaagtGCCACATTGGTCTCTCAATTTGT	*HR4*$^{Fei-0}$ promoter
G-41928	gctcttcacatTTTTTTTAAGTAGTTCTTTAGCTCTCGA	
G-41929	gctcttcaATGCCGCTTCTTGAGCTTGC	*HR4*$^{Fei-0}$ CDS truncation
G-41930	gctcttcaCTCAAGTACTAGCCTTACTAATTCAAGTT	
G-41929	see above	*HR4*^{ICE106} CDS truncation
G-41931	gctcttcaCTCAAGTACTACCCTTACTAATTCAAGTT	
G-41932	gctcttcaATGCCGATTGCTGAGCTTGC	*HR4*$^{Col-0}$ CDS truncation
G-41933	gctcttcaCTCACGTGCTACCCTTACTAATTCAAGTT	
G-41929	see above	*HR4*$^{Fei-0}$ CDS truncation + TAG
G-42152	gctcttcaCCACTACTCAAGTACTAGCCTTACTAATTCAAGTT	
G-41929	see above	*HR4*^{ICE106} CDS truncation + TAG
G-42153	gctcttcaCCACTACTCAAGTACTACCCTTACTAATTCAAGTT	
G-41932	gctcttcaATGCCGATTGCTGAGCTTGC	*HR4*$^{Col-0}$ CDS truncation + TAG
G-42154	gctcttcaCCACTACTCACGTGCTACCCTTACTAATTCAAGTT	

Supplemental Experimental Procedures

RPP7 phylogeny

The NB domain was predicted using SMART (http://smart.embl-heidelberg.de/). NB amino acid sequences were aligned using MUSCLE [1]. A maximum-likelihood tree was generated using the BLOSUM62 model in RaxML [2]. Topological robustness was assessed by bootstrapping 1,000 replicates.

RAPA phenotyping

Images were acquired daily in top view using two cameras per tray. Cameras were equipped with OmniVision OV5647 sensors with a resolution of 5 megapixels. Each camera was attached to a Raspberry Pi computer (Revision 1.2, Raspberry Pi Foundation, UK) [3]. Images of individual plants were extracted using a predefined mask for each plant. Segmentation of plant leaves and background was then performed by removing the background voxels then a GrabCut-based automatic postprocessing was applied [4]. Lastly, unsatisfactory segmentations were manually corrected. The leaf area of each plant was then calculated based on the segmented plant images.

Pathology

The *Hyaloperonospora arabidopsidis* isolate Hiks1 was maintained by weekly subculturing on susceptible Ws-0 *eds1-1* plants [5]. To assay resistance of susceptibility, 12- to 13-day old seedlings were inoculated with 5 x 10^4 spores/ml. Sporangiophores were counted 5 days after infection.

Constructs and transgenic lines

Genomic fragments were PCR amplified, cloned into pGEM®-T Easy (Promega, Madison, WI, USA), and either directly transferred to binary vector pMLBart or Gateway vectors pJLBlue and pFK210. amiRNAs [6] against members of the *RPP7* and *RPW8/HR* clusters were designed using the WMD3 online tool (http://wmd3.weigelworld.org/), and placed under the CaMV 35S promoter in the binary vector pFK210 derived from pGreen [7]. amiRNA constructs were introduced into plants using *Agrobacterium*-mediated transformation [8]. T_1 transformants were selected on BASTA, and crossed to incompatible accessions. For the chimeras, promoters and 5' coding sequences were PCR amplified from genomic DNA , repeat and tail sequences were synthesized using Invitrogen's GeneArt gene synthesis service, all were cloned into pBlueScript. The three parts, promoter, 5' and 3' coding sequences, were assembled using Greengate cloning [9] in the backbone vector pMCY2 [10]. Quality control was done by Sanger sequencing. Transgenic T_1 plants were selected based on mCherry seed fluorescence. For CRISPR/Cas9 constructs, sgRNAs targeting *HR4* or *RPW8.1* were designed on the Chopchop website (http://chopchop.cbu.uib.no/), and assembled using a Greengate reaction into supervector pRW006 (pEF005-sgRNA-shuffle-in [11] Addgene plasmid #104441).

mCherry positive T_2 transformants were screened for CRISPR/Cas9-induced mutations by Illumina MiSeq based sequencing of barcoded 250-bp amplicons. Non-transgenic homozygous T_3 lines were selected based on absence of fluorescence in seed coats.

Supplemental References

1. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic acids research, 2004. **32**(5): p. 1792-1797.
2. Stamatakis, A., *RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.* Bioinformatics, 2014. **30**(9): p. 1312-1313.
3. Vasseur, F., et al., *Image-based methods for phenotyping growth dynamics and fitness components in Arabidopsis thaliana.* Plant Methods, 2018. **14**(1): p. 63.
4. Cheng, M.M., et al., *DenseCut: Densely Connected CRFs for Realtime GrabCut.* Computer Graphics Forum, 2015. **34**(7): p. 193-201.
5. Holub, E.B.B., J.L., *Symbiology of Mouse-Ear Cress (Arabidopsis thaliana) and Oomycetes.* Advances in Botanical Research, 1997. **24**: p. 227-273.
6. Schwab, R., et al., *Highly specific gene silencing by artificial microRNAs in Arabidopsis.* Plant Cell, 2006. **18**(5): p. 1121-33.
7. Hellens, R.P., et al., *pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation.* Plant Mol Biol, 2000. **42**(6): p. 819-32.
8. Weigel, D. and J. Glazebrook, *Arabidopsis : a laboratory manual*. 2002: Cold Spring Harbor (N.Y.) : Cold Spring Harbor laboratory press.
9. Lampropoulos, A., et al., *GreenGate - A Novel, Versatile, and Efficient Cloning System for Plant Transgenesis.* PLOS ONE, 2013. **8**(12): p. e83043.
10. Emami, S., M.-c. Yee, and J.R. Dinneny, *A robust family of Golden Gate Agrobacterium vectors for plant synthetic biology.* Frontiers in Plant Science, 2013. **4**: p. 339.
11. Wu, R., et al., *An efficient CRISPR vector toolbox for engineering large deletions in Arabidopsis thaliana.* Plant Methods, 2018. **14**(1): p. 65.

A truncated singleton NLR causes hybrid necrosis in *Arabidopsis thaliana*

A. Cristina Barragan[1], Maximilian Collenberg[1], Jinge Wang[2], Rachelle R.Q. Lee[2], Wei Yuan Cher[2], Fernando A. Rabanal[1], Haim Ashkenazy[1], Detlef Weigel[1]*, Eunyoung Chae[1,2]*

*Corresponding authors: weigel@tue.mpg.de (D.W.), dbsce@nus.edu.sg (E.C.)

[1]Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076, Tübingen, Germany

[2]Department of Biological Sciences, National University of Singapore, 117558, Singapore

Keywords

Hybrid incompatibility, autoimmunity, singleton NLR, *DM10*, LRR-PL region, interchromosomal relocation

Abstract

Hybrid necrosis in plants arises from conflict between divergent alleles of immunity genes contributed by different parents, resulting in autoimmunity. We investigate a severe hybrid necrosis case in *Arabidopsis thaliana*, where the hybrid does not develop past the cotyledon stage and dies three weeks after sowing. Massive transcriptional changes take place in the hybrid, including the upregulation of most NLR disease resistance genes. This is due to an incompatible interaction between the singleton TIR-NLR gene *DANGEROUS MIX 10* (*DM10*), which was recently relocated from a larger NLR cluster, and an unlinked locus, *DANGEROUS MIX 11 (DM11)*. There are multiple *DM10* allelic variants in the global *A. thaliana* population, several of which have premature stop codons. One of these, which has a truncated LRR-PL region, corresponds to the *DM10* risk allele. The *DM10* locus and the adjacent genomic region in the risk allele carriers are highly differentiated from those in the non-risk carriers in the global *A. thaliana* population, suggesting that this allele became geographically widespread only relatively recently. The *DM11* risk allele is much rarer and found only in two accessions from southwestern Spain – a region from which the *DM10* risk haplotype is absent – indicating that the ranges of *DM10* and *DM11* risk alleles may be non-overlapping.

Introduction

Hybrid necrosis, a common form of hybrid incompatibility in plants, is caused by conflicting elements of the plant immune system originating from different parental accessions. These pairwise deleterious epistatic interactions usually involve at least one nucleotide binding site-leucine-rich repeat (NLR) protein (Bomblies et al. 2007; Alcázar et al. 2009; Yamamoto et al. 2010; Chae et al. 2014; Sicard et al. 2015; Deng et al. 2019). NLRs function as intracellular immune receptors, similarly to NOD/CARD proteins in animals, and play a major role in plant innate immunity (Maekawa, Kufer, and Schulze-Lefert 2011; Jones, Vance, and Dangl 2016). The constant co-evolutionary arms-race between plants and their pathogens has led to a high diversification of many elements of the plant immune system, including NLRs (Jones and Dangl 2006; Dodds and Rathjen 2010). Hybrid necrosis can be viewed as collateral damage resulting from, sometimes relatively minor, sequence differences between NLR alleles. This phenomenon can limit the possible NLR allele combinations found in an individual plant (Chae et al. 2014).

Plant NLRs are multidomain proteins usually composed of N-terminal Toll/Interleukin-1 receptor/Resistance protein (TIR), coiled-coil (CC) or RESISTANCE TO POWDERY MILDEW 8 (RPW8) domains, a central nucleotide-binding site (NBS) and C-terminal leucine-rich repeats (LRRs) (Meyers et al. 2003; Shao et al. 2016). The N-terminal domain is usually thought to be involved in signal transduction, while the NBS domain can act as a molecular ON/OFF switch (Bentham et al. 2017). The LRR domain is highly variable and consists of multiple repeats of 20-30 amino acid stretches that are often responsible for direct or indirect pathogen effector recognition as well as NLR auto-inhibition (Ade et al. 2007; Krasileva, Dahlbeck, and Staskawicz 2010; Steinbrenner, Goritschnig, and Staskawicz 2015). In addition, many TIR-NLRs carry a post-LRR (PL) domain, which is involved in pathogen effector recognition (Van Ghelder and Esmenjaud 2016; Martin et al. 2020).

Approximately half of all NLR genes in a given *A. thaliana* accession are found in multi-gene clusters, which are unevenly distributed across the genome (Meyers et al. 2003; Van de Weyer et al. 2019). Tandem duplication events are common in NLR clusters, and duplicate genes are a major source of genetic variation, since they often experience relaxed selection and enable neofunctionalization (Ohno 1970; Force et al. 1999; Lynch and Conery 2000; Conant and Wolfe 2008). Sequence homogenization through intergenic exchange among cluster members is greatly reduced when an NLR gene is translocated away from its original cluster to an unlinked genomic region, thereby preserving its original function or potentially developing a new one (Baumgarten et al. 2003; Leister 2004). For NLRs, neofunctionalization of duplicated or translocated genes can expand the repertoire of pathogen effectors an individual plant is able to recognize (Botella et al. 1998; Michelmore and Meyers 1998; Holub 2001; Kim et al. 2017).

Genome-wide analysis of structural variation across eight high-quality *A. thaliana* genomes identified rearrangement hot spots coinciding with numerous multi-gene NLR clusters (Jiao and Schneeberger 2020), including the previously described *DANGEROUS MIX* (*DM*) loci, which are causal for hybrid necrosis (Bomblies et al. 2007; Chae et al. 2014). This raises the possibility that accelerated evolution associated with genomic rearrangements contribute to the generation of incompatibility alleles, pointing to genomic architecture as a driver of hybrid incompatibility. So far, over a dozen NLR loci with hybrid necrosis alleles are known from multiple plant species. Curiously, even though singletons account for about a quarter of NLRs in different species (Jacob, Vernaldi, and Maekawa 2013), none of the causal NLR loci identified so far is a singleton NLR, here defined as a physical single-gene NLR and not to be confused with a functional singleton (Adachi, Derevnina, and Kamoun 2019). Most, but not all, well-characterized singleton NLRs, such as *RPM1* and *RPS2* in *A. thaliana*, show ancient balanced polymorphisms that maintain active and inactive alleles at intermediate frequencies in natural metapopulations (Stahl et al. 1999; Caicedo, Schaal, and Kunkel 1999; Mauricio et al. 2003; Allen et al. 2004; MacQueen, Sun, and Bergelson 2016). Thus, with less functional diversity, and beneficial alleles often being relatively common, one would indeed expect that singleton NLRs are underrepresented among hybrid necrosis loci.

Here, we are investigating a case of severe hybrid necrosis, where hybrid plants do not develop past the cotyledon stage, become necrotic, and die three weeks after sowing. Extensive transcriptional changes occur in the hybrid, including the induction of most NLR genes. Through a combination of QTL analysis and GWAS, we identified two new incompatibility loci, *DANGEROUS MIX 10* (*DM10*), a TIR-NLR on chromosome 5, and *DM11*, an unlinked locus on chromosome 1, as causal for incompatibility. *DM10* is an unusual hybrid incompatibility locus because it is a singleton NLR that arose after *A. thaliana* speciation through interchromosomal transposition from the *RLM1* cluster, which confers resistance to *Leptosphaeria maculans*, that causes blackleg disease in *Brassica* species (Staal et al. 2006; Guo et al. 2011). The causal allele has a premature stop codon that removes the C-terminal quarter of the protein, which includes part of the LRR-PL region, indicating that substantial NLR truncations can lead to hybrid incompatibility.

Results

A particularly severe case of hybrid necrosis: Cdm-0 x TueScha-9

Eighty *A. thaliana* accessions have previously been intercrossed with the goal of identifying hybrid incompatibility hot spots (Chae et al. 2014). A particularly severe case was observed in the crosses between Cdm-0 and five other accessions: TueScha-9, Yeg-1, Bak-2, ICE21 and Leo-1. The F_1 progeny of these two parents did not develop past the cotyledon stage, even at temperatures that suppress hybrid necrosis in most other cases

(Chae et al. 2014), and severe necrosis developed during the second week after sowing, followed by complete withering in the third week (**Fig 1A**).

To obtain insights into the transcriptional changes in the hybrid, we performed RNA-seq on the parental accessions Cdm-0 and TueScha-9, as well as in F_1 hybrid plants 10 days after germination, when the hybrid was already slightly stunted, but before there were visible signs of necrosis (**Fig 1B**). We observed massive transcriptional changes, in which around half of all 20,000 detectable genes (**Fig S1A**) were differentially expressed in the hybrid when compared to either parent (**Fig 1C**, **S1B**, **Table S1**). This represents one third of the entire *A. thaliana* transcriptome (Klepikova et al. 2016). A principal component analysis (PCA) showed that most of the variance in gene expression is driven by the difference between the parents and the hybrid (PC1: 83 %) (**Fig 1D**). In addition, we generated in silico hybrids (see Methods) and compared these with the biological F_1 hybrids through a PCA. This confirmed that gene expression in the F_1 hybrid is not an additive result of expression in the two parental accessions (**Fig S1C**). Next, we carried out a Gene Ontology (GO) analysis using the top 1,000 differentially expressed genes (DEGs) from a comparison of the F_1 hybrids and the mid-parental values (MPV) (**Table S2**). "Defense response" and "salicylic acid biosynthesis" were the categories with the highest number of DEGs in the hybrid versus MPV comparison (**Fig 1E**, **Table S2**).

Since the F_1 hybrid displayed signs of an increased pathogen defense response, we analyzed the expression of a set of marker genes for defense-associated phytohormones such as jasmonic acid (JA), salicylic acid (SA) and ethylene (ET) (Papadopoulou et al. 2018), as well as early pathogen response genes induced by both cell surface receptors and NLRs (Ding et al. 2020) (**Table S3**). Genes involved in SA biosynthesis and signaling, such as *EDS1, ICS1*, *EDS5*, *PAD4*, *PBS3, CBP60* and *FMO1*, were strongly overexpressed in F_1 hybrid plants, in concordance with the GO analysis, as was the SA-induced camalexin biosynthesis gene *CYP71A13*. The expression of genes encoding transcription factors *WRKY46* and *WRKY51* and of the late immune response gene *PR1* was also increased in the hybrid (**Fig 1F**, **Table S3**). In contrast, the expression of genes required for JA-mediated resistance, such as *MYC2* or *DDE2*, or genes involved in ET signaling, such as *ETR1, ERF1* and *EIN3*, changed to a lesser extent in the F_1 hybrid, similar to control genes *ACT7* and *TUB2* (**Fig 1F**, **Table S3**).

Since an increase in NLR expression has been linked to autoimmunity (Stokes, Kunkel, and Richards 2002; Mackey et al. 2003; Palma et al. 2010; Lai and Eulgem 2018), and since some NLRs are upregulated when SA levels rise (Shirano et al. 2002; Yang and Hua 2004; Tan et al. 2007; Mohr et al. 2010; MacQueen and Bergelson 2016), we set out to investigate NLR expression levels in the hybrid. Out of a set of 166 NLRs found in the Col-0 genome, 150 were expressed in at least one of the three genotypes studied, and 128 were significantly ($|\log_2 \text{FoldChange}| > 1$, padj value < 0.01) differentially expressed in at least one genotype comparison (**Fig S2D**, **Table S4**). From these 128 NLRs, all but one were differentially expressed when comparing the hybrid

with either parental accession (**Fig S1D**). NLRs were mostly upregulated in the F_1 hybrid: of the 95 NLRs with significant expression changes in the hybrid versus the MPV, all but three were overexpressed (**Fig 1G**, **S1E**, **Table S1**). When the F_1 hybrid was compared to the parents, the expression of individual NLRs largely followed the same pattern (**Fig S1E-G**), this was not the case when comparing the two parents (**Fig S1H**). The fraction of genes overexpressed in the F_1 hybrid was similar for the different NLR classes as well as between singleton and clustered NLRs (**Fig 1G**, **S1F-H**, **Table S3**).

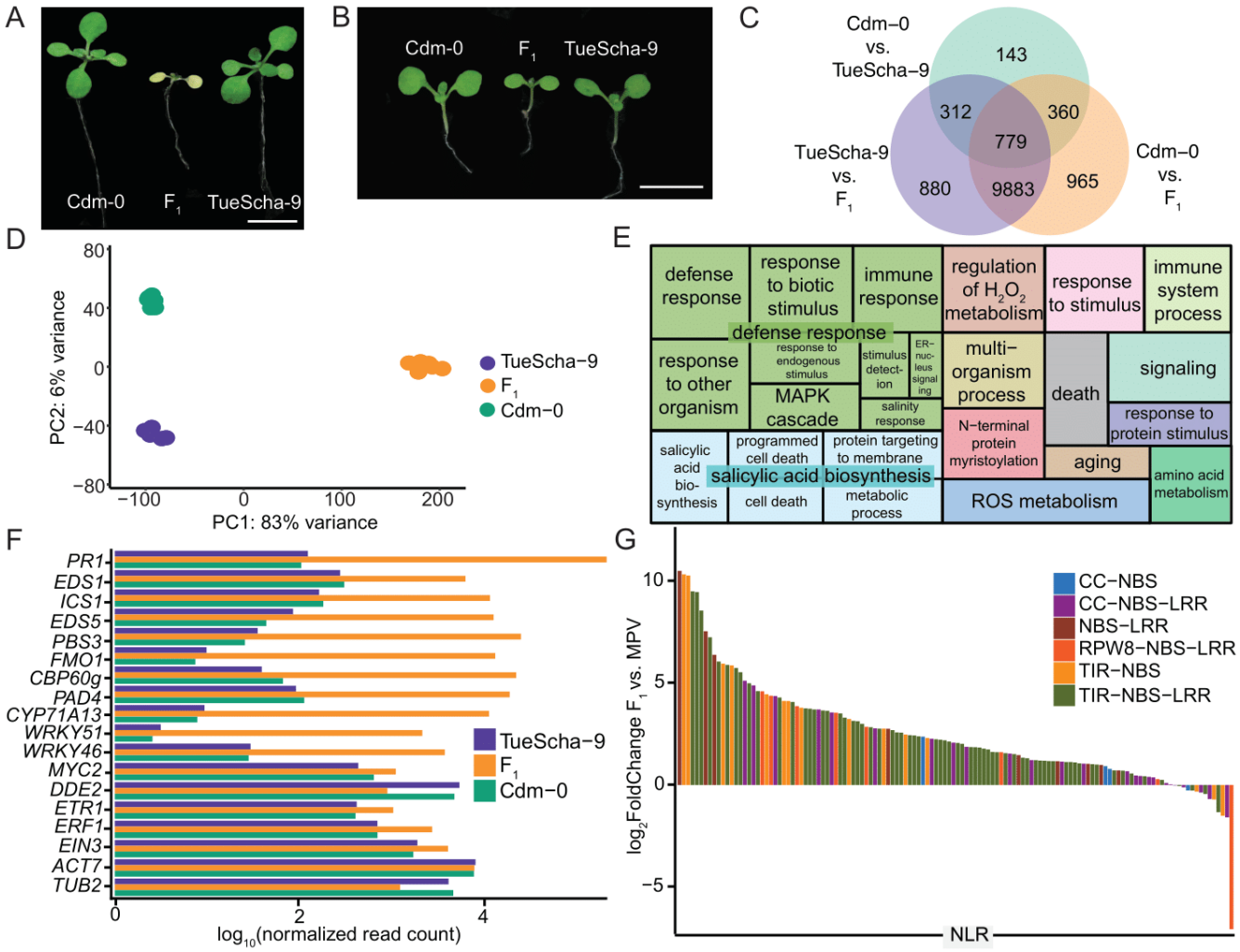


Fig 1. RNA-seq analysis of Cdm-0 x TueScha-9 hybrids. A. At 21 days, the Cdm-0 x TueScha-9 F_1 hybrid is necrotic. Plants were grown at 16°C. Scale bar represents 1 cm. **B.** Examples of a 10-day old Cdm-0 x TueScha-9 F_1 hybrid and parental accessions harvested for RNA-seq. Plants were grown at 23°C. Scale bar represents 1 cm. **C.** Intersection of DEGs between the F_1 hybrid and parents. **D.** PCA of gene expression values. The main variance is between the F_1 hybrid and parents. Each dot indicates one biological replicate, with six per genotype. **E.** REVIGO Gene Ontology treemap. Size of the square represents -$\log_{10}$(*p* value) of each GO term. **F.** $\log_{10}$(normalized read count) of defense-related marker genes of the hybrid and the parents. **G.** Differences in expression between the F_1 hybrids and the mid-parental values (MPV) of NLR genes, with 128 significantly (|$\log_2$FoldChange| >1, padj value < 0.01) differentially expressed in at least one genotype comparison.

QTL mapping of *DM10* and *DM11* in a triple-hybrid cross

Having found that a very large fraction of NLR genes is upregulated in the Cdm-0 x TueScha-9 hybrid, we wondered whether hybrid necrosis in this case was due to global NLR regulators (Li, Pennington, and Hua 2009; Zhai et al. 2011; Shivaprasad et al. 2012; Gloggnitzer et al. 2014; Sicard et al. 2015), or to NLRs, as in other hybrid necrosis cases. We therefore proceeded to map the underlying causal loci via quantitative trait locus (QTL) analysis. Since the F_1 hybrid seedlings died very young, we could not directly generate a segregating F_2 mapping population (Bomblies et al. 2007; Chae et al. 2014; Barragan et al. 2019). Instead, we designed a triple-hybrid cross (Cooper et al. 2019) and first generated two sets of heterozygous plants by crossing Cdm-0 and TueScha-9 separately to a third, innocuous background, the Col-0 reference accession. We then intercrossed these Cdm-0/Col-0 and TueScha-9/Col-0 plants (**Fig 2A**). In the resulting pseudo-F_2 generation, we collected both normal and necrotic plants and individually genotyped them by RAD-seq (Rowan et al. 2017). For QTL mapping, we focused on polymorphic markers between Cdm-0 and TueScha-9, including markers overlapping with the Col-0 reference (**Fig 2B**), and also analyzed polymorphic markers for each accession independently (**Fig S2A**, **B**). We identified two genomic regions that interact epistatically to cause the severe hybrid necrosis phenotype. We called the QTL on chromosome 5 (23.35 to 24.45 Mb) *DM10*, and the QTL on chromosome 1 (21.55 to 22.18 Mb) *DM11*. Both intervals contain NLRs but no clear candidates for global NLR regulators, so we chose to focus on NLR genes. In the *DM10* mapping interval, one NLR is present, At5g58120, while the *DM11* interval was NLR-rich and encompassed 10 NLRs in Col-0 (**Table S5**). Loci in the interval include the highly polymorphic *RPP7* cluster of CC-NLR genes (McDowell et al. 2000; Guo et al. 2011; Li et al. 2020), as well as two CC-NLR singleton genes, *CW9* (At1g59620) and At1g59780 (Meyers et al. 2003). To identify potential differences between Col-0 and Cdm-0 in the *DM11* interval, we generated a PacBio long-read-based genome assembly of this accession (**Table S6**). Notably, most chromosome arms were assembled in single contigs, including the long arm of chromosome 1, where the *DM11* mapping interval is located (**Fig S3**). Since at the time the full Cdm-0 annotation was not yet available, we manually annotated homologs of NLR genes corresponding to the genomic region that spans from At1g56510 to At1g64070 in Col-0, which includes the *DM11* mapping interval as well as neighboring NLRs. Like Col-0, Cdm-0 carries groups of both clustered and singleton NLRs, adding up to a total of 21 NLRs, compared to 28 NLRs in Col-0 (**Fig 2C**, **Table S5**).

To pinpoint *DM11* candidate genes, we sought to identify additional accessions that had similar alleles as Cdm-0 at *DM11* candidate loci by creating Neighbor-Joining (NJ) trees (**Fig S2C**) and PCA plots (**Fig S2D**, **Table S7**), using sequences from the 1001 Genomes Project (1001 Genomes Consortium 2016). IP-Cum-1 was the accession most similar to Cdm-0 for the whole *DM11* mapping interval, and when we crossed it to TueScha-9, Cdm-0 x TueScha-9-like hybrid necrosis was observed (**Fig 2D**). Eleven other accessions that were less

closely related to Cdm-0 in this genomic interval did not produce necrotic F_1 hybrids (**Table S7**). Because accessions Istisu-1 and ICE134, like Cdm-0, lack a transposable element that is present in most *RPP7* (At1g58602) alleles (Tsuchiya and Eulgem 2013), we also crossed these two accessions to TueScha-9, but no hybrid necrosis was observed (**Table S7**). Artificial miRNAs (amiRNAs) (Schwab et al. 2006) targeting different members of the *RPP7* cluster were previously designed to perform rescue experiments for other cases of hybrid necrosis (Chae et al. 2014; Barragan et al. 2019); although predicted to target all members of the Cdm-0 *RPP7* cluster, neither these nor amiRNAs targeting *CW9*$^{Cdm-0}$ or At1g59780^{Cdm-0} suppressed hybrid necrosis (**Table S8**). Lastly, a genomic *CW9*$^{Cdm-0}$ fragment was unable to induce hybrid necrosis when introduced into TueScha-9 (**Table S18**).

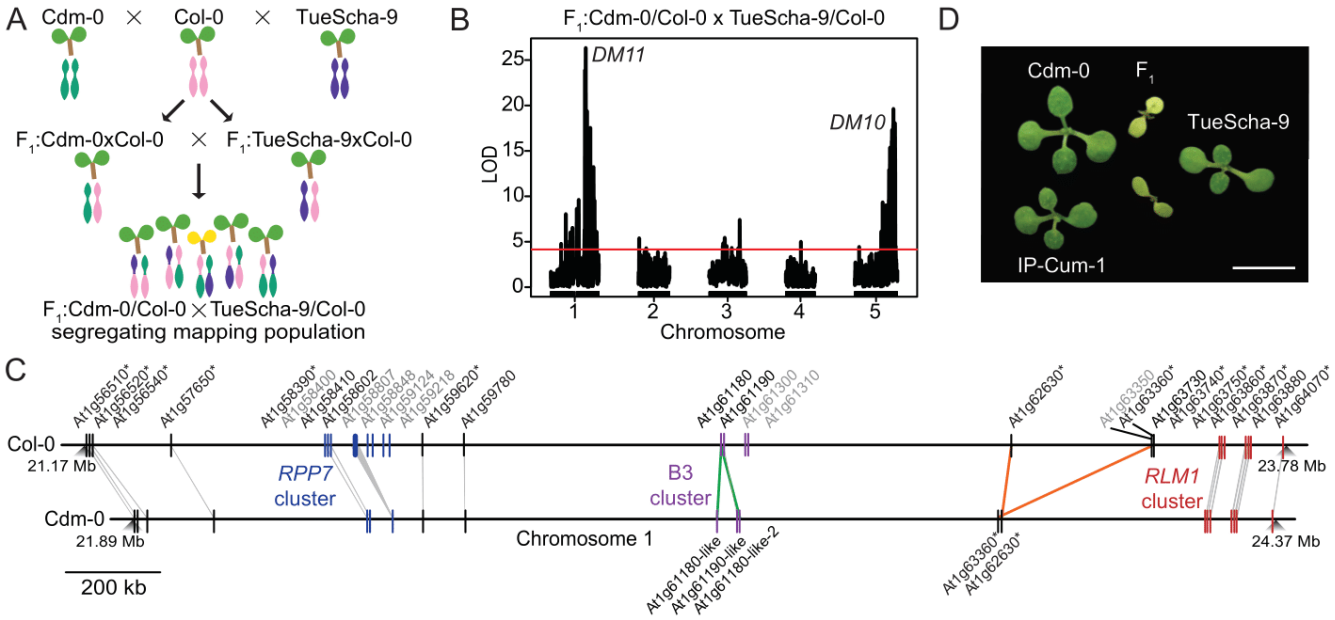


Fig 2. QTL mapping with a triple-hybrid cross. A. Creation of a Cdm-0 x TueScha-9 mapping population. **B.** QTL analysis from polymorphic Cdm-0 and TueScha-9 markers. QTL peaks are found on chromosome 5 (23.35-24.45 Mb), *DM10*, and chromosome 1 (21.55-22.18 Mb), *DM11*. The horizontal lines indicate 0.05 significance threshold established with 1,000 permutations. **C.** Comparison and distribution of candidate *DM11* NLR genes between At1g56510 and At1g64070 on chromosome 1. Gene IDs in grey are present in Col-0 but not in Cdm-0, gene duplications are marked in green and inversion events in orange. Asterisks indicate significant ($|\log_2 FoldChange|>1$, padj value < 0.01) gene expression changes in the F_1 hybrid when compared to the MPV. **D.** Cdm-0 x TueScha-9 and IP-Cum-1 x TueScha-9 hybrids. Plants are two weeks old and were grown at 16°C. Scale bar represents 1 cm.

Being aware that the precision of QTL mapping in NLR-rich regions can be affected by structural variation, we also tested NLRs adjacent to the *DM11* mapping interval. The *RLM1* cluster is highly similar among Cdm-0 and IP-Cum-1, both of which carry the causal *DM11* allele in addition, some cluster members show an increased expression in the F_1 hybrid, which is sometimes the case for causal NLRs (Bomblies et al. 2007) (**Fig S2E**). We therefore tested six of the seven *RLM1* cluster members via *Nicotiana benthamiana* co-expression with *DM10*$^{TueScha-9}$ (see Fig 4 for cloning of causal *DM10* allele), but none induced a hypersensitive response (HR)

(**Table S5**). Six accessions with a similar *RLM1* locus to that of Cdm-0 and IP-Cum-1 were crossed with TueScha-9, but no necrosis was observed (**Table S7**). Finally, because At1g57650 was strongly upregulated among *DM11* NLR candidate genes, we tested it with *DM10*$^{\text{TueScha-9}}$ in *N. benthamiana*, but again no HR was observed (**Fig S2E**, **Table S4**, **S5**). This may indicate that *DM11* is either an NLR that was not tested, or not an NLR at all. Other *DM11* candidates may include any of the genes in this interval that encode proteins that are not annotated as NLRs but have a TIR or LRR domain (**Tables S9-S11**). Note that some Col-0 NLRs that had no homologs in the interval from At1g56510 to At1g64070 in Cdm-0 attracted non-specific RNA-seq reads, most likely because there are homologs elsewhere in the Cdm-0 genome (**Table S5**).

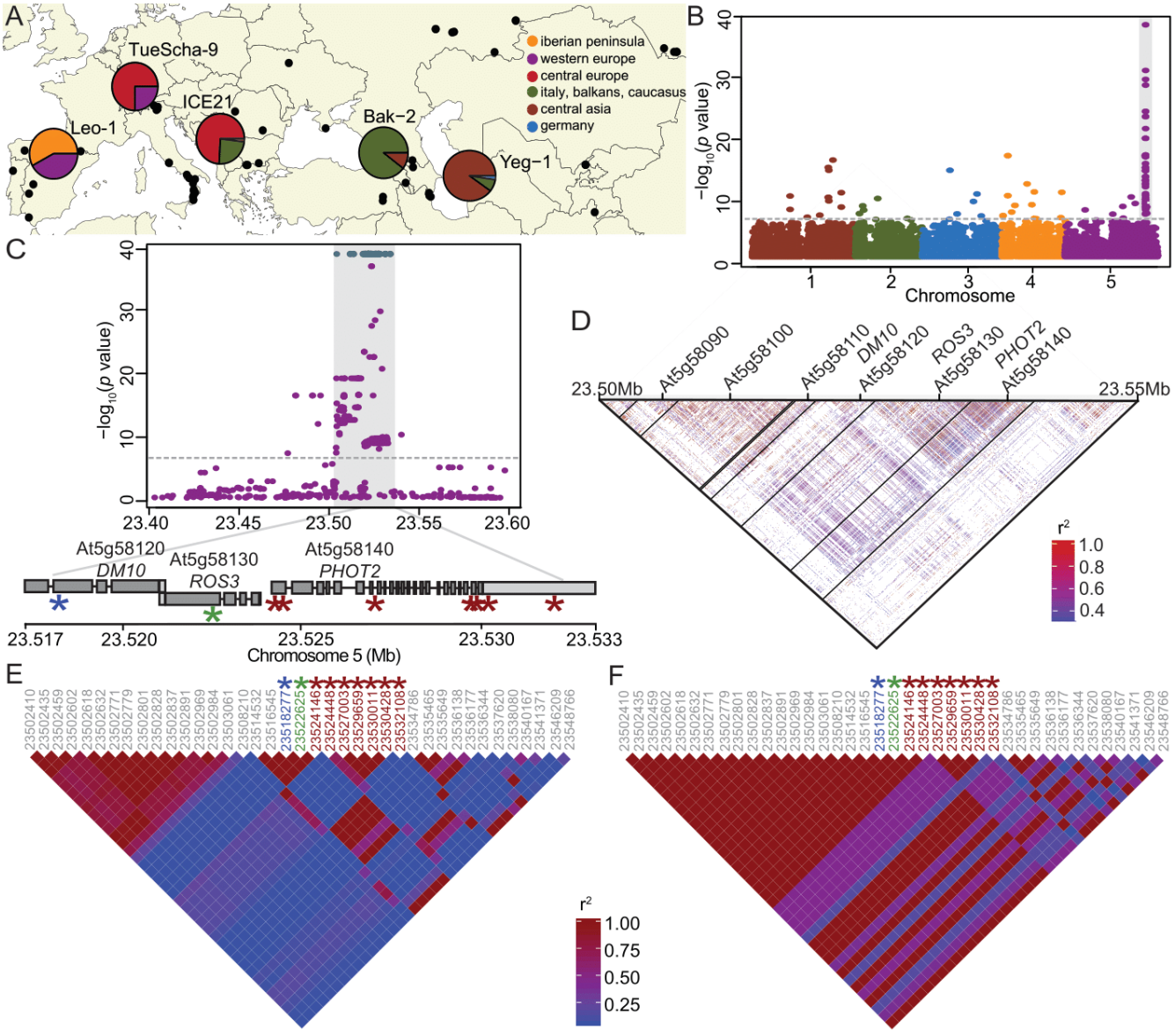


Fig 3. GWAS of hybrid necrosis in 80 accessions. A. Map of 80 accessions (black dots), with the five risk accessions colored according to 1001 Genomes admixture groups (1001 Genomes Consortium 2016). **B.** Manhattan plot for association of necrosis in Cdm-0 hybrid progeny when selfed and crossed to 79 other

accessions. The significance threshold (Bonferroni correction, 5% familywise error) is indicated as a horizontal dotted line (same in C). **C.** Close-up of the region highly associated with hybrid necrosis; SNPs with a 1:1 association marked in teal. Asterisks indicate such 1:1 associations in At5g58120, *ROS3* and *PHOT2*; see also E and F. SNP positions are given in Table S13. **D.** Linkage disequilibrium (LD) across a 50 kb region in chromosome 5. Strong linkage is observed from At5g58090 to At5g58140. **E.** LD across the same 50 kb region as in D, with a subset of markers from 80 accessions crossed to Cdm-0. Asterisks indicate markers highlighted in C. **F.** LD across a 50 kb region with the same markers as in E, but for the five risk accessions only. Higher LD is seen here than in E.

Fine-mapping of *DM10* using genome-wide association studies

In the original collection of 6,409 crosses among 80 accessions (Chae et al. 2014), four accessions in addition to TueScha-9 produced severe hybrid necrosis when crossed to Cdm-0: Yeg-1, Bak-2, ICE21 and Leo-1. Together with TueScha-9, these represent much of the Eurasian range of the species, both geographically and genetically; six of the nine previously identified admixture groups (1001 Genomes Consortium 2016) are present in these five risk accessions (**Fig 3A**, **Table S12**). Given the diversity of the five incompatible accessions, and knowing that most, but not all, causal genes for hybrid incompatibility are NLRs, we attempted to narrow down causal *DM10* candidate genes by GWAS, with Cdm-0-dependent F_1 necrosis as a binary trait (Grimm et al. 2016). We discovered a remarkably high association between this phenotype and several closely linked markers on the bottom of chromosome 5, with corrected *p* values as low as 10^{-38}. In addition, 79 SNPs showed a one-to-one association with the necrotic phenotype, resulting in $-\log_{10} p$ values of 0 (**Fig 3B**, **Table S13**). The markers with the strongest associations tagged three loci: At5g58120, encoding a TIR-NLR without known function, *ROS3* (At5g58130), encoding an enzyme involved in DNA demethylation (Zheng et al. 2008), and *PHOT2* (At5g58140), encoding a blue light receptor that mediates phototropism (Harada, Sakai, and Okada 2003) (**Fig 3C**, **Table S13**). These three loci are genetically similar among the five risk accessions, yet differentiated from the other 75 accessions used for GWAS (**Fig S4A-C**). Looking at linkage among loci in this genomic region, we could see that, when taking all 80 accessions into account, six loci (At5g58090-Atg58140) belong to one large linkage block, in which *ROS3* and *PHOT2* are under tight linkage and the TIR-NLR At5g58120 constitutes a separate linkage block (**Fig 3D**). Notably, in the five accessions causing hybrid incompatibility, stronger linkage is observed in this region than that seen among the same markers from all 80 accessions (**Fig 3E**, **F**). In the risk accessions, At5g58120, *ROS3* and the proximal part of *PHOT2* form one linkage block, while SNPs located in the distal half of *PHOT2* are found in a separate linkage block, rendering At5g58120 and *ROS3* as primary candidates for causality in hybrid necrosis (**Fig 3F**).

DM10, a singleton TIR-NLR, as cause of severe hybrid necrosis

Having candidate genes for *DM10*, we next sought to experimentally test their causality for severe hybrid necrosis. Genomic fragments of the TIR-NLR At5g58120 and *ROS3*, from both Col-0 and TueScha-9, were

introduced into Cdm-0 plants. A 4.8 kb genomic fragment containing At5g58120$^{TueScha-9}$ recapitulated the Cdm-0 x TueScha-9 hybrid necrosis phenotype (**Fig 4A**, **Table S14**). At5g58120 is henceforth called *DM10*. When *DM10*$^{TueScha-9}$ was introduced into a Col-0 background and the resulting T_1 plants were subsequently crossed to Cdm-0, we also observed the hybrid incompatibility phenotype in the F_1 progeny (**Fig 4B**). *DM10*$^{Col-0}$, *ROS3*$^{TueScha-9}$ and *ROS3*$^{Col-0}$ did not produce any necrosis when introduced into a Cdm-0 background (**Table S14**). We also observed that, when infiltrated in *N. benthamiana* leaves and overexpressed under an EtOH-inducible promoter, both *DM10*$^{Col-0}$ and *DM10*$^{TueScha-9}$ were able to trigger cell death, which was not the case when *DM10*$^{Col-0}$ and *DM10*$^{TueScha-9}$ were expressed under the control of their native promoters (**Fig 4C**). The cell death-triggering activities under forced overexpression in the heterologous *N. benthamiana* system indicate that these NLRs are competent in immune signaling and, in the case of DM10$^{TueScha-9}$, this is not abolished by its substantial truncation.

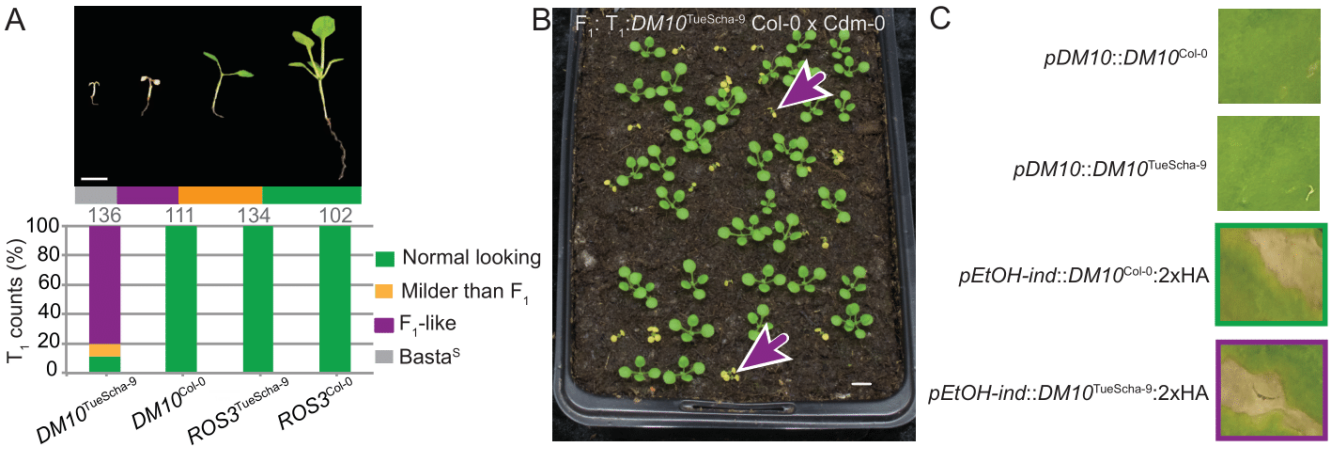


Fig 4. Experimental identification of *DM10*. A. Recapitulation of hybrid necrosis in 25-day old Cdm-0 T_1 plants transformed with the indicated genomic fragment from TueScha-9 or Col-0. Representative phenotype and total number of T_1 plants examined given on top. Plants were grown at 16°C. Scale bar represents 1 cm. **B.** The same *DM10*$^{TueScha-9}$ genomic fragment as in A was introduced into Col-0, and T_1 plants were crossed to Cdm-0. The F_1 hybrid phenotype was recapitulated (magenta arrows). Plants were 18 days old and grown at 16°C. Scale bar represents 1 cm. **C.** Infiltration of *N. benthamiana* leaves with the indicated constructs. Overexpression of either *DM10*$^{TueScha-9}$ or *DM10*$^{Col-0}$ under an EtOH inducible promoter (*pEtOH-ind*) triggered cell death, while expression from their native promoter (*pDM10*) did not. Images were taken 7 days afetr infection.

Prevalence and genetic differentiation of the *DM10* risk allele in the global *A. thaliana* population

To study natural variation across different *DM10* alleles, 73 alleles belonging to accessions originating from across *A. thaliana*'s native range were extracted from preliminary short- and long-read genome assemblies available in-house (**Fig S5A**, **Table S15**). A Maximum-Likelihood (ML) tree of these alleles showed that there are multiple well-supported DM10 clades (**Fig S5B**), and that variation between DM10 proteins was

most prevalent at the C-terminal end (**Fig 5A**, **Fig S5C**, **Table S16**). Ten alleles were predicted to produce proteins truncated at three different points. Four accessions, including TueScha-9, the original *DM10* risk allele carrier, share the same stop codon (**Fig 5B**, **S5B**), removing three LRRs and the post-LRR domain (PL). Three short motifs have been previously recognized as being conserved in PLs of different NLRs (Van Ghelder and Esmenjaud 2016) (**Fig 5B**, **Fig S5C**). The DM10 PL has a variant of the first of these motifs, in a degenerate form. Five accessions had shorter, 335 amino acid long DM10 proteins; in these, the NBS domain was truncated, lacking motifs which are important regulators of NLR activity (Bendahmane et al. 2002; Sueldo et al. 2015; Bentham et al. 2017) (**Fig 5B**, **Fig S5C**). These five accessions carrying short *DM10* alleles included Cdm-0 and IP-Cum-1, which also carry *DM11* risk alleles. This implies that the short Cdm-0-like *DM10* variants do not interact with *DM11* to produce hybrid necrosis. The shortest predicted DM10 protein, found in the Sha accession, is only 90 amino acids long and is truncated midway through the TIR-2 motif (**Fig 5B**, **Fig S5C**). There are full-length DM10 alleles without any non-synonymous differences to $DM10^{Cdm\text{-}0}$ and $DM10^{Sha}$, and that are distinguished from these only by truncation. Similarly, the full-length $DM10^{Col\text{-}0}$ and the truncated $DM10^{TueScha\text{-}9}$ proteins differ only at 3% of shared sites, which is low for within-species variation among NLR alleles (Van de Weyer et al. 2019). Furthermore, not only the coding sequence, but also the sequence after the premature stop codon of the short *DM10* alleles, is highly similar to that of full-length alleles. This lack of signs of pseudogenization suggests that the truncations occurred relatively recently. As is typical for NLRs (Mondragón-Palomino et al. 2002; Ruggieri, Nunziata, and Barone 2014), Ka/Ks values above 1 were found in the LRR domain when comparing $DM10^{Col\text{-}0}$ and $DM10^{TueScha\text{-}9}$ (**Fig 5C**). In contrast, $DM10^{Col\text{-}0}$ and $DM10^{Lerik1\text{-}3}$, which are both full-length DM10 proteins but from different clades, are more differentiated in TIR and NB-ARC domains, although Ka/Ks values above 1 are also restricted to the LRR domain (**Fig S5D**).

Next, to assess how common the *DM10* risk allele is in the global *A. thaliana* population, we again turned to the 1001 Genomes collection (1001 Genomes Consortium 2016). Since *DM10*, *ROS3* and the proximal part of *PHOT2* were strongly linked in accessions carrying the *DM10* risk allele, we focused on this region, which contained 785 SNPs. In a NJ tree, all five confirmed *DM10* risk allele carriers were found in the same branch, which included 95 other accessions (**Fig 5D**, **Table S17**). In a PCA of this region, these 100 accessions were clearly separated from the rest (**Fig 5E**), which was not the case in a whole-genome PCA (**Fig S5E**), indicating that population structure is not the main driving force separating risk from non-risk allele carriers. To experimentally confirm that sequence was predictive of interaction with the *DM11* risk allele, 25 of the 100 accessions were crossed to Cdm-0 (**Fig S5B**, **Table S17**). All but IP-Alm-0 produced hybrid necrosis. Notably, while DM10 from IP-Alm-0 is 99.2% identical with $DM10^{TueScha\text{-}9}$, it does not have the LRR truncation (**Fig S5B**). This implies that the truncation in *DM10* risk alleles is likely responsible for incompatibility, and not individual amino acid changes. Ten random accessions not predicted to carry the *DM10* risk allele were

crossed to Cdm-0 as a control; as expected, none produced hybrid necrosis (**Table S17**). Similarly, we investigated how common the other two DM10 truncations are in the global *A. thaliana* population. The shortest Sha-like *DM10* allele was found in 29 accessions, while the Cdm-0-like truncation is more common, although not as common as the *DM10* risk allele, and was found in 67 accessions (**Table S17**).

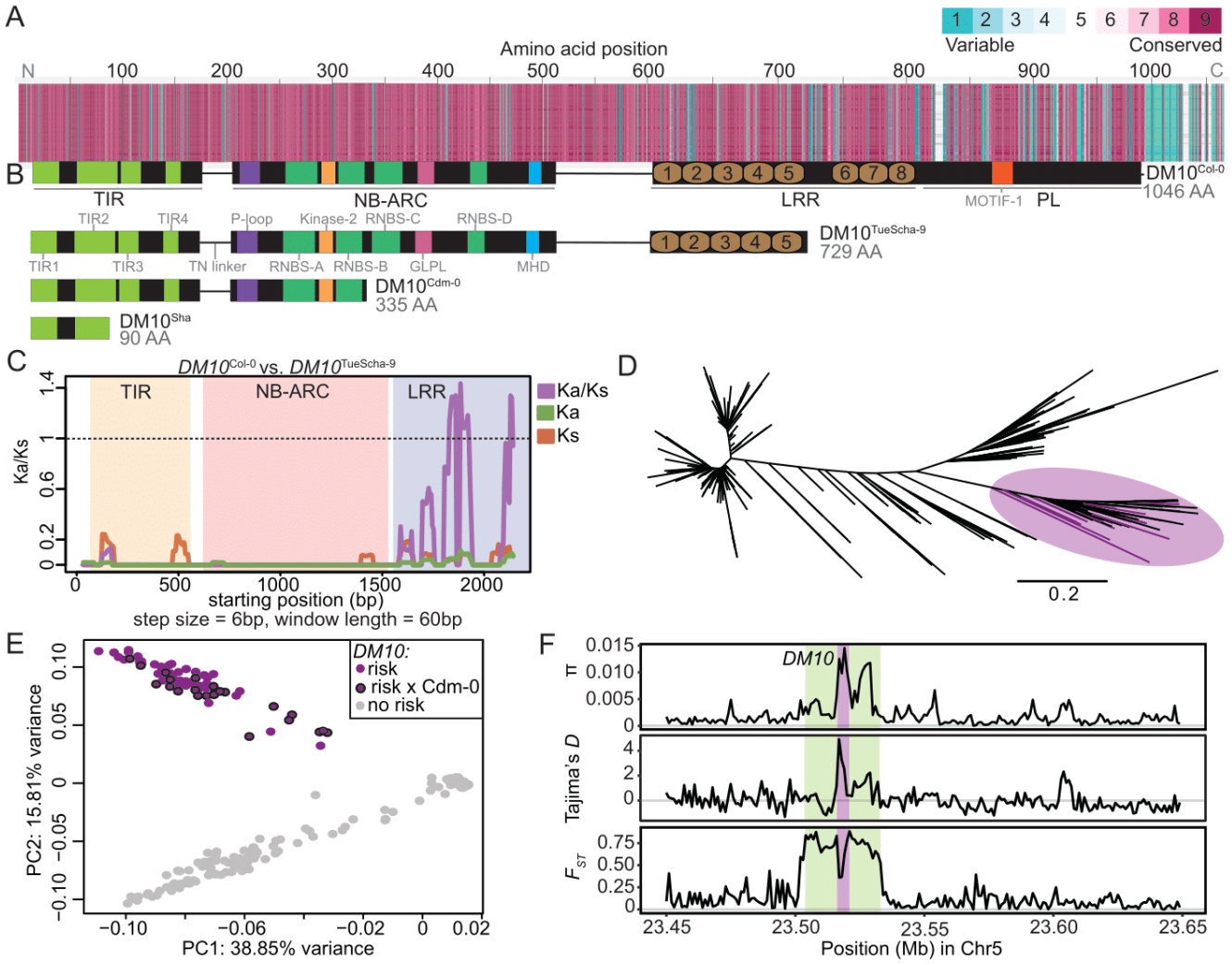


Fig 5. *DM10* natural variation. A. Amino acid alignment of 73 DM10 proteins color-coded by its conservation score (Armon, Graur, and Ben-Tal 2001). **B.** Comparison between different DM10 proteins (aligned with A). **C.** Ka/Ks ratios between *DM10*$^{Col-0}$ and *DM10*$^{TueScha-9}$. **D.** NJ tree of a region including *DM10*, *ROS3* and *PHOT2* sequences from the 1001 Genomes Project (1001 Genomes Consortium 2016). Branch lengths in nucleotide substitutions are indicated. Accessions carrying the *DM10* risk alleles group together in a branch (magenta), risk accessions crossed to Cdm-0 are highlighted. **E.** PCA. Accessions carrying the predicted *DM10* risk (magenta) versus non-risk (grey) alleles are clearly separated in PC2. Risk accessions crossed to Cdm-0 are outlined in black. **F.** π, Tajima's *D* and F_{ST} for *DM10* (magenta), the *DM10* linkage block comprising At5g58090-Atg58140 (green) and surrounding genomic regions.

In a 200 kb region around *DM10*, nucleotide diversity (π) was highest, up to 0.015, in the distal half of *DM10*, encoding the more polymorphic LRR domain (**Fig 5F**). However, in comparison with other TIR-NLRs present

in most or all accessions, overall *DM10* nucleotide diversity was not uncommon (Van de Weyer et al. 2019). Values for Tajima's *D* reached 4.8 in the proximal half of *DM10,* hinting at multiple *DM10* alleles being prevalent at intermediate frequencies in the global *A. thaliana* population, as is often the case for NLRs (Stahl et al. 1999; Caicedo, Schaal, and Kunkel 1999; Bakker et al. 2006; Karasov et al. 2014). Lastly, the fixation index (F_{ST}) between 98 accessions with predicted *DM10* risk alleles (excluding IP-Alm-0 and RAD-21, which did not have truncated LRR domains) and 1,037 non-risk allele carrying accessions, peaked at 0.88 across the *DM10* linkage block (**Fig 3D-F**, **Fig 5F**). This was the only peak detected both across the entire chromosome 5 (**Fig S5F**) and the whole genome. Inside this block, a drop in F_{ST} is seen over the proximal half of *DM10*, which is consistent with this region being similar between risk and some non-risk alleles (**Fig 5C**).

Taken together, these results show that there are multiple *DM10* alleles in the global *A. thaliana* population, three of which are predicted to result in truncated proteins due to the presence of premature stop codons, one of which is the *DM10* risk allele. Notably, the *DM10* risk allele is not only relatively common and genetically differentiated in our GWAS population, but also in the global *A. thaliana* population.

No documented co-occurrence of *DM10* and *DM11* risk alleles in the global *A. thaliana* population

Looking at the geographical distribution of accessions carrying different *DM10* alleles with premature stop codons, we observed that both the Cdm-0-like *DM10* allele as well as the risk *DM10* allele were found at similar densities throughout *A. thaliana*'s native range, while the Sha-like *DM10* allele was mainly found towards the eastern part of the species' distribution (**Fig 6A**, **Table S17**). In the case of the *DM10* risk allele, the one exception to where this allele was found, was the southwestern part of Spain and Portugal, even though *A. thaliana* has been heavily sampled in this region (1001 Genomes Consortium 2016). The fact that the only two *DM11* risk carriers identified so far, Cdm-0 and IP-Cum-1, are found in southwestern Spain may indicate that the *DM10* and *DM11* risk alleles do not geographically co-occur (**Fig 6B**).

To provide additional support for this assertion, we first attempted to identify more *DM11* risk carriers in Spain and Portugal. We crossed TueScha-9, a *DM10* risk allele carrier, to 24 accessions from these two countries, which were found at different geographical distances from the two *DM11* risk carriers Cdm-0 and IP-Cum-1, as well as from accessions carrying the *DM10* risk allele (**Table S7**). No hybrid necrosis was observed in any of the resulting F_1 progeny (**Fig 6B**, **Table S7**). This, together with our aforementioned attempts to find additional *DM11* carriers among accessions that are closely related in the *DM11* genomic region to Cdm-0 and IP-Cum-1, indicates that the *DM11* risk allele is rare and potentially only found in southwestern Spain, a region where the *DM10* risk allele appears to be absent.

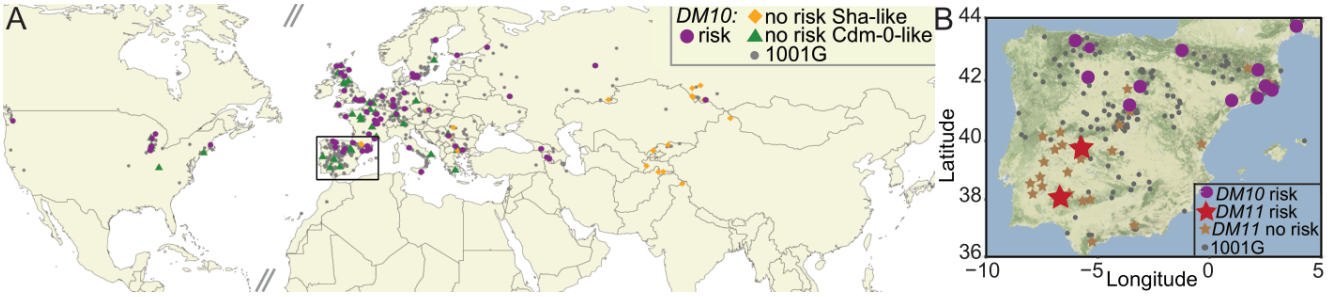


Fig 6. Geographical distribution of *DM10* and *DM11* alleles. A. Geographical distribution of 1001 Genomes Project accessions (1001 Genomes Consortium 2016) carrying the *DM10* risk (magenta), non-risk (grey) alleles, Sha-like non-risk (orange) and Cdm-0-like non-risk (green) alleles. Rectangle zooms into the region shown in B. **B.** Distribution of 1001 Genomes Project accessions (grey) in Spain and Portugal, carrying the *DM10* (magenta) and *DM11* (red) risk alleles, as well as accessions carrying *DM11* non-risk alleles (brown) which were crossed to TueScha-9.

Origin of the *DM10* NLR singleton locus through a recent interchromosomal relocation event out of the *RLM1* cluster

In the *A. thaliana* Col-0 reference genome, we identified nine NLR genes closely related to *DM10*. Seven of these make up the *RLM1* cluster on chromosome 1, and two others, At2g16870, At4g14370 are dispersed singletons. In the related species *Arabidopsis lyrata* and *Brassica rapa*, we identified a further 20 *DM10/RLM1* homologs (**Fig 7A**). As in *A. thaliana*, the cluster homologous to the *A. thaliana RLM1* cluster in these two species (not to be confused with the *RLM1* locus that provides resistance to blackleg disease in *Brassica* (Delourme et al. 2008; Fu et al. 2019)) underwent within-species duplication and inversion events (**Fig 7B**). Most *RLM1* members from *A. thaliana* have a clear one-to-one homolog in *A. lyrata*, so the expansion of the *RLM1* cluster must have occurred before the two species diverged (Beilstein et al. 2010). The *A. lyrata* homologs of At2g16870 and At4g14370, 480565 and 493465, are found in a different chromosome than the main *RLM1* cluster (**Fig 7C**). This is not the case for the *DM10* homolog from *A. lyrata*, 875509, which is located inside the main *RLM1* cluster (**Fig 7C**). This indicates that *DM10* was relocated away from the main *RLM1* cluster to another chromosome and that this occurred after *A. lyrata* and *A. thaliana* diverged.

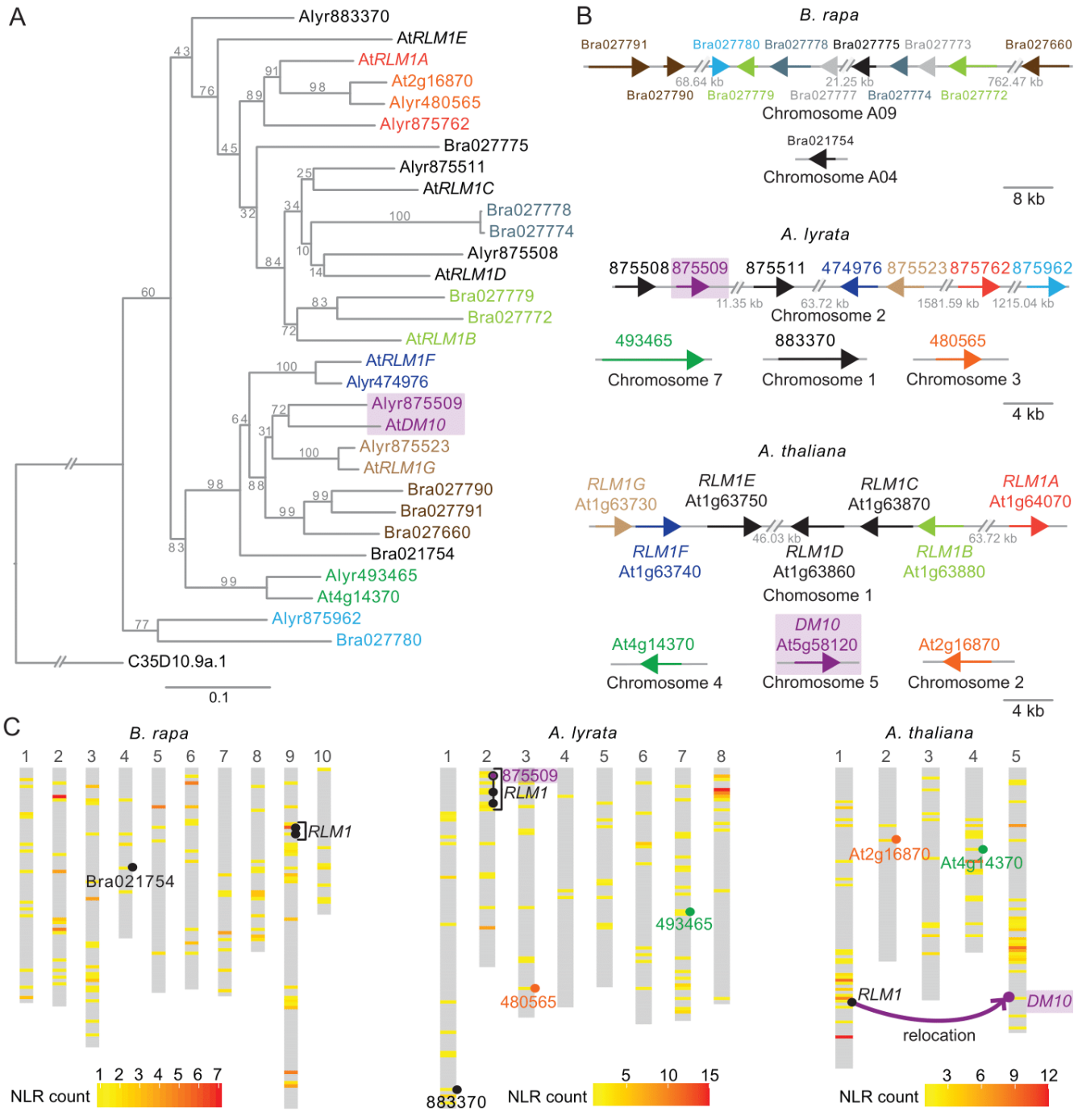


Fig 7. *RLM1* locus in *B. rapa*, *A. lyrata* and *A. thaliana* reference genomes. A. ML tree of the NBS domain (CDS) of *RLM1* cluster members from *B. rapa*, *A. lyrata* and *A. thaliana*, with the NBS domain of *C. elegans* CED-4 as outgroup (C35D10.9a.1). 1,000 bootstrap replicates were performed, values are shown on each branch. Branch lengths in nucleotide substitutions are indicated. The same color was chosen for genes in neighboring branches with bootstrapping values above 70. Diagonal lines indicate a gap in the tree branches. **B.** *RLM1* cluster members and their homologs in *B. rapa*, *A. lyrata* and *A. thaliana*. Color-coding the same as in A, genes in grey are truncated, arrows represent size of NLR loci. Diagonal lines indicate a positional gap along the chromosome, the length of the gap is indicated. **C.** Heatmaps of NLR densities across the three genomes. Window sizes were calculated by dividing the length of the longest chromosome by 100. *RLM1* cluster and

closely related singletons indicated. *DM10* in *A. thaliana* and its homolog in *A. lyrata*, 875509 are highlighted in magenta.

Discussion

Over ten causal genes for hybrid necrosis have been identified in *A. thaliana* and other plants (Krüger et al. 2002; Bomblies et al. 2007; Alcázar et al. 2009; Jeuken et al. 2009; Yamamoto et al. 2010; Chen et al. 2014; Chae et al. 2014; Todesco et al. 2014; Sicard et al. 2015; Deng et al. 2019; Sandstedt, Wu, and Sweigart 2020). In many instances, at least one of the two causal genes is an NLR, which is also the case for the Cdm-0 x TueScha-9 incompatibility. What makes this case particularly interesting is the extreme severity of hybrid necrosis, the transcriptional hyper-induction of NLR genes, and the causality of a truncated singleton NLR, *DM10*, which was recently relocated from a larger NLR cluster.

As with normal immune responses and autoimmune syndromes, the expression of hybrid necrosis is typically temperature-dependent, and hybrid necrosis in *A. thaliana* can usually be completely suppressed when grown above 23°C (Bomblies et al. 2007; Alcázar et al. 2009; Chae et al. 2014; Todesco et al. 2014; Świadek et al. 2017). In contrast, the extreme autoimmune response in Cdm-0 x TueScha-9 F_1 seedlings cannot be rescued even by growing these hybrids at 28°C (Chae et al. 2014). In other necrotic hybrids, non-causal NLRs have been reported to be differentially expressed between hybrids and their parents (Bomblies et al. 2007; Atanasov et al. 2018), but the NLR induction seen in Cdm-0 x TueScha-9 is clearly the most extreme. For example, in the F_1 progeny, 128 of 150 expressed NLRs are differentially expressed in at least one genotype comparison, with almost all being overexpressed. When we reanalyzed published data from another, relatively strong hybrid necrosis case, we found 104 out of 166 NLR genes to be differentially expressed, yet both extreme as well as the mean of overexpression was lower than in Cdm-0 x TueScha-9 hybrids (**Fig S6** and **Tables S1**-**S4**). In addition, the specific NLR genes that are overexpressed differ in both cases, indicating that specific NLRs respond differently depending on their genetic background (**Fig S6** and **Table S4**). Simultaneous upregulation of several NLR genes has been observed after exposure to biotic (Zipfel et al. 2004; Tan et al. 2007; Ribot et al. 2008; Mohr et al. 2010; Yu et al. 2013; Sohn et al. 2014; Chen et al. 2015; Mine et al. 2018; Steuernagel et al. 2020) and abiotic stresses (MacQueen and Bergelson 2016), but not to the extent seen in Cdm-0 x TueScha-9 hybrids. Given that elevated NLR expression levels can trigger cell death (Stokes, Kunkel, and Richards 2002; Mackey et al. 2003; Palma et al. 2010; Lai and Eulgem 2018), we expect that widespread NLR hyper-induction is a significant contributor to the strongly necrotic phenotype of Cdm-0 x TueScha-9 F_1 hybrids.

NLR transcript levels are tightly controlled through a variety of regulatory mechanisms (Lai and Eulgem 2018), and large-scale upregulation of NLRs could possibly require multiple pathways. We found WRKY transcription factors to be overexpressed in the hybrids; these proteins bind to W box motifs enriched in the promoters of

multiple members of the plant immune system, including NLRs, and can induce widespread NLR expression, enhancing basal immunity (Eulgem and Somssich 2007; Pandey and Somssich 2009; Mohr et al. 2010). Two other mechanisms known to affect a broad set of NLRs are the miRNA-dependent phasiRNA production (Zhai et al. 2011; Li et al. 2012; Shivaprasad et al. 2012; Xia et al. 2015) as well as nonsense-mediated decay (NMD) (Gloggnitzer et al. 2014), both of which help to dampen NLR gene expression in the absence of pathogen threats. Repression is attenuated after an incoming pathogen is detected by the plant, enabling global NLR levels to increase (Lai and Eulgem 2018). While we have no direct evidence for transcription factors, small RNAs or NMD as contributors to aberrant NLR expression in the Cdm-0 x TueScha-9 hybrid, this exceptional hybrid necrosis case may present a good tool for comparing NLR regulation under pathogen attack with strong autoimmunity.

We found 17% of DM10 proteins encoded in a global set of *A. thaliana* accessions to be truncated in either their TIR, NBS or LRR domain. Similar to several full-length variants, the alleles for all three truncated proteins have intermediate frequencies and are relatively wide-spread, suggesting that they are actively maintained in the global population by balancing selection. The most common of the three truncation alleles is the *DM10* risk version, which lacks three of the eight LRRs and the PL domain, and which shows evidence for its LRR domain being under diversifying selection. While the TIR domain alone can induce cell death (Swiderski, Birker, and Jones 2009; Bernoux et al. 2011), a complete NBS domain is essential in many instances (Dodds, Lawrence, and Ellis 2001; Dinesh-Kumar, Tham, and Baker 2000; Bendahmane et al. 2002; Tameling et al. 2002; Williams et al. 2011; Steinbrenner, Goritschnig, and Staskawicz 2015; Wang et al. 2015; Sueldo et al. 2015; Bernoux et al. 2016). NLRs lacking the NBS or LRR domain are not only known to retain the ability to cause cell death, but there are cases where truncated NLRs are bona fide resistance genes (Roth et al. 2017; Nishimura et al. 2017; Marchal et al. 2018). Conversely, other proteins, including at least one full-length NLR, can induce cell death through activation of naturally occurring truncated NLRs (Zhao et al. 2015; Y. Zhang et al. 2017). In the case of *DM10*, we do not know whether only the full-length variants or the truncated variants, or both, are functional and if they confer resistance to unknown pathogens even though their prevalence and geographical distribution suggest so. Alternatively the "less is more" hypothesis (Olson 1999) may explain the wide prevalence of truncated *DM10* alleles even if these are non-functional. Minor mutations in these alleles could readily remove the premature stop codons, making them "nearly functional" alleles that could act as easily activable functional reservoirs, as previously discussed for *RPM1* (Rose et al. 2012). The particular length of the risk DM10$^{\text{TueScha-9}}$ protein combines the autoactive tendencies associated with the partial loss of the LRR-PL domain (Qi et al. 2018) with what appear to be functional TIR and NBS domains.

Because NLR allelic diversity is often not easily captured by short-read based resequencing (Van de Weyer et al. 2019), we still do not have a good grasp on whether NLR alleles in general, and specifically beneficial alleles,

spread through the population more quickly than other adaptive alleles. The Iberian peninsula is a center of *A. thaliana* genetic diversity, with strong geographical structure across a north-south latitudinal gradient (Picó et al. 2008; Brennan et al. 2014). We observed a lack of co-occurrence between *DM11* risk alleles, restricted to southwestern Spain, and *DM10* risk alleles, restricted to the northern half of Spain (1001 Genomes Consortium 2016). Absence of co-occurrence between risk alleles may partly be the result of population structure: two geographical barriers potentially reducing gene flow, the Tagus river and the Central System mountains, divide populations carrying either *DM10* or *DM11* risk alleles. In any case, more definitive proof of the mutual exclusion of *DM10* and *DM11* risk alleles will require more extensive sampling of natural populations across the Iberian peninsula. Co-occurrence of hybrid incompatibility alleles in a single population has been observed before, where different alleles are maintained at intermediate frequencies, but in this case, the hybrids show a milder necrosis phenotype in the lab than Cdm-0 x TueScha-9, and no obvious phenotype in the wild (Todesco et al. 2014). The extreme necrotic phenotype caused by the *DM10-DM11* interaction, which appears to be largely independent of growth conditions, makes it unlikely that the hybrid phenotype would be suppressed in the wild. In addition, since outcrossing rates of *A. thaliana* in the wild can be substantial (Bomblies et al. 2010), it is conceivable that in some areas these rates are high enough for lethal hybrids to exact a noticeable fitness cost on risk allele carriers.

An interchromosomal relocalization event of the *RLM1* cluster gave rise to *DM10* after *A. thaliana* speciation. Which evolutionary forces might have helped *DM10* to become established on a separate chromosome, if any? NLR genes in clusters are likely to be more mutable than singletons because of illegitimate recombination (Michelmore and Meyers 1998; Baumgarten et al. 2003; Meyers, Kaushik, and Nandety 2005; Wong and Wolfe 2005; Wicker, Yahiaoui, and Keller 2007). If *DM10* underwent beneficial neofunctionalization after duplication, its relocation away from the cluster might have stabilized the locus. Another possibility could be conflicts among gene cluster members. Cluster members are sometimes transcriptionally co-regulated (Yi and Richards 2007; Deng et al. 2017), so translocation away from the cluster would allow for evolution of new expression patterns for *DM10*. More generally, genomic relocation would enable *DM10* to be subjected to different selection regimes than its cluster homologs. Either way, the fact that the genomic region surrounding *DM10* – different from some other *RLM1* cluster members – is a recombination cold spot (Choi et al. 2016) is consistent with our finding of high LD around the *DM10* locus, especially in accessions carrying the *DM10* risk allele. Together with our phylogenetic results and Tajima's *D* measurements, this would seem to support the idea of stable *DM10* haplotypes being particularly advantageous.

While our triple-hybrid cross enabled the identification of the *DM10* and *DM11* QTLs, fine-mapping was complicated by three sets of markers and two loci being involved. Genotyping around the *DM11* locus to differentiate alleles from each of the three grandparents in the mapping cross was further confounded by

structural variants, which are typical for NLR-rich regions. Members of the *DM10*-related *RLM1* cluster near the inferred *DM11* QTL are in principle good hybrid necrosis candidates, because TIR domains tend to form homomeric complexes (Zhang et al. 2017; Dong et al. 2018; Martin et al. 2020); the similar TIR domains between DM10 and RLM1 members make it particularly likely that they oligomerize, which is often an important step in NLR activation. We cannot entirely exclude members of the *RLM1* cluster, because we tested most of them only by co-expression with *DM10* in the heterologous *N. benthamiana* system, and not by genetic inactivation or recapitulation in *A. thaliana*.

In conclusion, we have presented a severe case of hybrid necrosis in *A. thaliana*, where the hybrids show global NLR hyper-induction triggered by the interaction of *DM10*, a relocated singleton NLR gene, and *DM11*, an unlinked locus in chromosome 1. Comparative structure-function analysis of the truncated $DM10^{TueScha-9}$ hybrid necrosis risk allele and the closely related full-length $DM10^{Col-0}$ allele, which does not cause hybrid necrosis, should reveal the exact contributions of LRR and PL subdomains to NLR activity. In addition, the *DM10*/*DM11* case provides a good tool to investigate the consequences of simultaneous activation of a large fraction of NLRs. In the future, by identifying the role of different *DM10* and *RLM1* alleles in response to natural pathogens, one could test whether chromosomal relocation affects how evolution is acting on this group of highly related NLR genes.

Materials and Methods

Constructs are listed in Table S18 and primers in Table S19.

Plant material

Stock numbers of accessions used are listed in Supplementary Material. All plants were stratified in the dark at 4°C for 4-6 days prior to planting on soil. Late flowering accessions were vernalized six weeks under short day conditions (8 h light) at 4°C. All plants were grown in long days (16 h of light) at 16°C or 23°C at 65% relative humidity under 110 to 140 µmol m^{-2} s^{-1} light provided by Philips GreenPower TLED modules (Philips Lighting GmbH, Hamburg, Germany).

RNA sequencing

Six biological replicates of 10 day-old shoots of Cdm-0 x TueScha-9 hybrids and their parental accessions were collected. RNA was extracted as described in (Yaffe et al. 2012). The NEBNext magnetic isolation module (New England Biolabs), was used for mRNA enrichment. Sequencing libraries were prepared using NEBNext Ultra II directional RNA library kit and paired-end sequenced (150bp) in an Illumina HiSeq3000 (Illumina Inc.,

San Diego, USA) instrument. Reads were mapped against the *A. thaliana* reference TAIR10 using bowtie2 (v2.2.6) (Langmead and Salzberg 2012). Default parameters were chosen unless mentioned otherwise. Transcript abundance was calculated with RSEM (v1.2.31) (Li and Dewey 2011). In silico hybrids were generated to enable mid-parent value calculations: parental read files were normalized according to sequencing depth and were subsampled by randomly drawing 50 % of the reads with seqtk (v2.0-r82-dirty; https://github.com/lh3/seqtk). Differential gene expression analyses were performed using DESeq2 (v1.18.1) (Love, Huber, and Anders 2014). Genes with less than ten counts over all 18 samples were removed from downstream analyses. Significant changes in gene expression between two genotypes were determined by filtering for genes with a $|\log_2\text{FoldChange}| > 1$ and padj value < 0.01. One read was added to all normalized read counts in Fig 1G, S2E and S6E to avoid plotting -INF values in non-expressed genes ($\log_{10}(0+1)=0$). Non-additive gene expression between Cdm-0 x TueScha-9 F_1 hybrids in silico hybrids was analyzed by computing principal components based on the normalized read counts of the top 500 most variable genes across all 18 samples. Plots were generated using the R package ggplot2 (v3.2.0) (Wickham 2009) and heatmaps were plotted using pheatmap (v1.0.8) (Kolde 2012). Gene Ontology (GO) analyses were performed using AgriGO (Tian et al. 2017) using the SEA method. The GO results were visualized with REVIGO treemap (Supek et al. 2011), for clearer visualization only the top 13 and GO categories with the lowest *p* values were plotted in Fig 1G, the complete list of GO terms is found in Table S2.

Genotyping-by-sequencing and QTL mapping

F_1 progeny from bi-directional crosses of F_1 (TueScha-9/Col-0) x (Cdm-0/Col-0) was used as a mapping population. The seedlings showing the hybrid necrosis phenotype vs. those that did not, were genotyped individually in a 1:1 ratio. Plants were 10 days old when collected. Genomic DNA was extracted with CTAB (cetyl trimethyl ammonium bromide) buffer (Doyle and Doyle 1987) and then purified through chloroform extraction and isopropanol precipitation (Ashktorab and Cohen 1992). Genotyping-by-Sequencing (GBS) using RAD-seq was used to genotype individuals in the mapping populations with KpnI tags (Rowan et al. 2017). Briefly, libraries were single-end sequenced on a HiSeq 3000 instrument with 150 bp reads. Reads were processed with Stacks (v1.35) (Catchen et al. 2013) and mapped to TAIR10 with bwa-mem (v0.7.15) (H. Li 2013), variant calling was performed with GATK (v3.5) (McKenna et al. 2010). QTL was performed using R/qtl (Broman et al. 2003) with the information from 348 F_2 individuals from 4 independent lines of this segregating population and 6179 markers.

De novo genome assembly and annotation

The Cdm-0 accession (ID 9943; CS76410) was grown as described above. To reduce starch accumulation, 3-week-old plants were put into darkness for 30 h before harvesting. Sixteen grams of flash frozen leaf tissue were ground in liquid nitrogen and nuclei isolation was performed according to (Workman et al. 2018) with the following modifications for *A. thaliana*: eight independent reactions of two grams each were carried out, and the filtered cellular homogenate was centrifuged at 7,000 x g. High-molecular-weight DNA was recovered with the Nanobind Plant Nuclei Kit (Circulomics; SKU NB-900-801-01), and needle-sheared 1x (FINE-JECT® 26Gx1" 0.45x25mm, LOT 14-13651). A 35-kb template library was prepared with the SMRTbell® Express Template Preparation Kit 2.0, and size-selected with the BluePippin system according to the manufacturer's instructions (P/N 101-693-800-01, Pacific Biosciences, California, USA). In addition, a PCR-free library was prepared with the NxSeq® AmpFREE Low DNA Library Kit from Lucigen® according to the manufacturer's instructions (Cat No. 14000-2). The final library was sequenced on a Pacific Biosciences Sequel instrument with Binding Kit 3.0. PacBio long-reads were assembled with Canu (v1.71) (Koren et al. 2017). The resulting contigs were first polished using the long-reads with the Arrow algorithm (v2.3.2; https://github.com/PacificBiosciences/GenomicConsensus), followed by a second polishing step with PCR-free short-reads using the Pilon algorithm (v1.22) (Walker et al. 2014). Lastly, the resulting contigs were scaffolded based on TAIR10 assembly by REVEAL (v0.2.1) (Linthorst et al. 2015). The previously generated Cdm-0 transcriptome sequencing data were mapped against the scaffolded genome assembly using HISAT (v2.0.5) (Kim, Langmead, and Salzberg 2015). Subsequently, the mapping results were used as extrinsic RNA sequencing evidence when annotating the genome using AUGUSTUS (v3.2.3) (Stanke et al. 2006). Transposable elements and repetitive regions were identified and masked prior to gene annotation using RepeatModeler2 (v2.01) (Flynn et al. 2020). Orthologous genes shared between Cdm-0 and the current *A. thaliana* reference annotation from Araport11 were identified using Orthofinder (v2.4.0) (Emms and Kelly 2019).

Manual NLR annotation of the *DM11* mapping interval

The 20-25 Mb region of chromosome 1 was extracted from the Cdm-0 assembly. The assembly was used as a query against a subject FASTA file containing 167 NLR genes from the Col-0 reference accession using blastn (Altschul et al. 1990). Hits were binned in 20 kb intervals and the percentage identity between the queries and the subject was visualized across all bins. NLRs between At1g56510 to At1g64070 in Col-0 found in this interval were manually annotated based on the percentage identity plotted and on AUGUSTUS gene predictions (v2.5.5) (Stanke et al. 2006).

GWAS

Cdm-0-dependent hybrid necrosis in the F_1 progeny from crosses with 80 accessions (Chae et al. 2014) was scored as 0 or 1. The binary trait with accession information was submitted to the easyGWAS platform (Grimm et al. 2016) using the FaSTLMM algorithm. A -$\log_{10}$(*p* value) was calculated for every SNP along the five *A. thaliana* chromosomes.

Constructs and transgenic lines

Genomic fragments were PCR amplified, cloned into pGEM®-T Easy (Promega, Madison, WI, USA) and then transferred to the binary vectors pMLBart, pCambia1300 or pFK210. Constructs were introduced into plants using *Agrobacterium*-mediated transformation (Weigel and Glazebrook 2002). T_1 transformants were selected on BASTA (pMLBart and pFK210) and crossed to incompatible accessions. Ethanol-inducible constructs were PCR amplified, cloned into pGEM®-T Easy, as part of a separate experiment, 2xHA tags were added via PCR and the whole fragment, which was then transferred to the pCR8® entry vector (ThermoFisher Scientific). Next, the genomic fragment was moved to the destination vector pZZ006 (Caddick et al. 1998) through the Gateway® LR reaction (ThermoFisher Scientific). Quality control for all constructs was done by Sanger sequencing. For transient expression in *N. benthamiana, A. tumefaciens* strains ASE (*RLM1*) or GV3101 (*DM10* and At1g57650) were grown to an OD_{600} of 1.2-1.8 and incubated in induction medium (10 mM MES (pH 5.6), 10 mM $MgCl_2$, and 150 μM acetosyringone) overnight. The cell suspensions were normalized to an OD_{600} of 0.8 and co-infiltrations suspensions were mixed 1:1. Suspensions were then infiltrated into the abaxial side of *N. benthamiana* leaves. In the case of EtOH inducible constructs, infiltrated *N. benthamiana* were induced at 18 h post-infiltration (hpi) by irrigation with 1% ethanol and kept within a transparent plastic dome for another 18 h. *DM10 N. benthamiana* constructs shown in Fig 4 were co-expressed with a *35S::GFP* construct as part of a larger experiment to test for candidate *DM11* loci.

Population genetic analyses

Amino acid sequence conservation scores were calculated with ConSurf (Ashkenazy et al. 2016; Armon, Graur, and Ben-Tal 2001). SNPs occurring in repetitive regions and only present in one of the 73 extracted *DM10* alleles were considered sequencing errors and were manually curated. Protein domains were predicted using InterProScan (Jones et al. 2014). LRR domains were predicted with LRRsearch and the score threshold was set at 7 (Bej et al. 2014). NLR motifs were defined based on previous studies (Meyers et al. 2003; Shao et al. 2016). Nonsynonymous to synonymous substitution rates were calculated using KaKs_Calculator (v2.0) (Wang et al. 2010) with the NG method (Nei and Gojobori 1986); a window length of 60 bp and a step size of 6 bp were chosen. Genomic regions of interest were subsetted from a 1135 genomes VCF file (1001 Genomes Consortium 2016) using VCFtools (v0.1.14) (Danecek et al. 2011). The resulting VCF file was filtered by MAF=0.01 and a maximum percent of missing data per SNP of 30%. Sequences were converted to FASTA,

aligned with MUSCLE (v3.8.31) (Edgar 2004) and then visualized with Aliview (v1.18.1) (Larsson 2014). Neighbor-Joining trees were calculated with Fastphylo (v1.0.1) (Khan et al. 2013) and visualized with iTol (Letunic and Bork 2007) (https://itol.embl.de/shared/cbarragan). Maximum-likelihood trees were calculated with RaxML (v0.6.0) using the GTR+G4 model (Stamatakis 2014). Linkage disequilibrium (r^2), was calculated with PLINK (v1.90) (Purcell et al. 2007). Principal component analyses were calculated with smartPCA (Patterson, Price, and Reich 2006). Tajima's *D*, F_{ST}, and nucleotide diversity (π) were also calculated with VCFtools. Maps were created with the R-packages maps (v3.3) and ggmap (v3.0) (Kahle and Wickham 2013). Admixture groups were assigned to each accession in accordance with the 1001 Genomes project (1001 Genomes Consortium 2016); since TueScha-9 had not been part of that study, admixture group assignments for it were estimated based on the genetic make-up of neighboring accessions. *RLM1* homologs in *A. lyrata* and *B. rapa* were identified using the Ensembl Plants portal (Bolser et al. 2016). Sequences from the genome assemblies TAIR10 (*A. thaliana*), *B. rapa* (v1.5) and *A. lyrata* (v1.0) were used for phylogenetic analyses.

Data Availability

Sequencing data can be found at the European Nucleotide Archive (ENA) under project numbers PRJEB38267 (RNA-seq experiment) and PRJEB40125 (Cdm-0 assembly) and in the GenBank under accession numbers MT488482 to MT488554 (*DM10* alleles).

Author contributions

Conceptualization: ACB, DW, EC.

Formal analysis: ACB, MC, RL, FR, HA, EC.

Funding acquisition: DW, EC.

Investigation: ACB, JW, WYC, EC.

Methodology: ACB, EC.

Project administration: DW.

Supervision: DW.

Writing – original draft: ACB.

Writing – review & editing: ACB, DW, EC.

Competing interests

The authors have declared that no competing interests exist.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft through the Collaborative Research Center (CRC1101), the Max Planck Society (to D.W.) and the Academic Research Fund (MOE2019-T2-1-134) from the Ministry of Education, Singapore, Intramural Research Fund (R-154-000-B33-114) from the National University of Singapore (to E.C.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We thank Sang-Tae Kim and members of the Weigel lab for critical reading of the manuscript, Adrian Streit for suggesting the triple-hybrid cross experiment, Christa Lanz for preparing the Cdm-0 PacBio library, Lei Li for performing the *RLM1 N. benthamiana* transient expression experiment, Rebecca Schwab for generating some crosses, Gautam Shirsekar and Sergio Latorre for discussion, Joe Win for suggesting to look into the DM10 PL domain and Sophien Kamoun, Jiorgos Kourelis and Ksenia Krasileva for pointers to examples of truncated NLRs. We also thank the 1001G+ team for providing access to preliminary whole genome *A. thaliana* assemblies. Lastly, we thank the anonymous reviewers for their valuable suggestions.

References

1001 Genomes Consortium. 2016. "1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana." *Cell* 166 (2): 481–91.

Adachi, Hiroaki, Lida Derevnina, and Sophien Kamoun. 2019. "NLR Singletons, Pairs, and Networks: Evolution, Assembly, and Regulation of the Intracellular Immunoreceptor Circuitry of Plants." *Current Opinion in Plant Biology* 50 (August): 121–31.

Ade, Jules, Brody J. DeYoung, Catherine Golstein, and Roger W. Innes. 2007. "Indirect Activation of a Plant Nucleotide Binding Site-Leucine-Rich Repeat Protein by a Bacterial Protease." *Proceedings of the National Academy of Sciences of the United States of America* 104 (7): 2531–36.

Alcázar, Rubén, Ana V. García, Jane E. Parker, and Matthieu Reymond. 2009. "Incremental Steps toward Incompatibility Revealed by Arabidopsis Epistatic Interactions Modulating Salicylic Acid Pathway Activation." *Proceedings of the National Academy of Sciences* 106 (1): 334–39.

Allen, R. L., P. D. Bittner-Eddy, L. J. Grenville-Briggs, J. C. Meitz, A. P. Rehmany, L. E. Rose, and J. L. Beynon. 2004. "Host-Parasite Coevolutionary Conflict between Arabidopsis and Downy Mildew." *Science* 306 (5703): 1957–60.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10.

Armon, A., D. Graur, and N. Ben-Tal. 2001. "ConSurf: An Algorithmic Tool for the Identification of Functional Regions in Proteins by Surface Mapping of Phylogenetic Information." *Journal of Molecular Biology* 307 (1): 447–63.

Ashkenazy, Haim, Shiran Abadi, Eric Martz, Ofer Chay, Itay Mayrose, Tal Pupko, and Nir Ben-Tal. 2016. "ConSurf 2016: An Improved Methodology to Estimate and Visualize Evolutionary Conservation in Macromolecules." *Nucleic Acids Research* 44 (W1): W344–50.

Ashktorab, H., and R. J. Cohen. 1992. "Facile Isolation of Genomic DNA from Filamentous Fungi." *BioTechniques* 13 (2): 198–200.

Atanasov, Kostadin E., Changxin Liu, Alexander Erban, Joachim Kopka, Jane E. Parker, and Rubén Alcázar. 2018. "NLR Mutations Suppressing Immune Hybrid Incompatibility and Their Effects on Disease Resistance." *Plant Physiology* 177 (3): 1152–69.

Bakker, Erica G., Christopher Toomajian, Martin Kreitman, and Joy Bergelson. 2006. "A Genome-Wide Survey of R Gene Polymorphisms in Arabidopsis." *The Plant Cell* 18 (8): 1803–18.

Barragan, Cristina A., Rui Wu, Sang-Tae Kim, Wanyan Xi, Anette Habring, Jörg Hagmann, Anna-Lena Van de Weyer, et al. 2019. "RPW8/HR Repeats Control NLR Activation in Arabidopsis thaliana." *PLoS Genetics* 15 (7): e1008313.

Baumgarten, Andrew, Steven Cannon, Russ Spangler, and Georgiana May. 2003. "Genome-Level Evolution of Resistance Genes in Arabidopsis thaliana." *Genetics* 165 (1): 309–19.

Beilstein, M. A., N. S. Nagalingum, M. D. Clements, S. R. Manchester, and S. Mathews. 2010. "Dated Molecular Phylogenies Indicate a Miocene Origin for Arabidopsis thaliana." *Proceedings of the National Academy of Sciences of the United States of America* 107 (43): 18724–28.

Bej, Aritra, Bikash Ranjan Sahoo, Banikalyan Swain, Madhubanti Basu, Pallipuram Jayasankar, and Mrinal Samanta. 2014. "LRRsearch: An Asynchronous Server-Based Application for the Prediction of Leucine-Rich Repeat Motifs and an Integrative Database of NOD-like Receptors." *Computers in Biology and Medicine* 53 (October): 164–70.

Bendahmane, Abdelhafid, Garry Farnham, Peter Moffett, and David C. Baulcombe. 2002. "Constitutive Gain-of-Function Mutants in a Nucleotide Binding Site-Leucine Rich Repeat Protein Encoded at the Rx Locus of Potato." *The Plant Journal: For Cell and Molecular Biology* 32 (2): 195–204.

Bentham, Adam, Hayden Burdett, Peter A. Anderson, Simon J. Williams, and Bostjan Kobe. 2017. "Animal NLRs Provide Structural Insights into Plant NLR Function." *Annals of Botany* 119 (5): 827–702.

Bernoux, Maud, Hayden Burdett, Simon J. Williams, Xiaoxiao Zhang, Chunhong Chen, Kim Newell, Gregory J. Lawrence, et al. 2016. "Comparative Analysis of the Flax Immune Receptors L6 and L7 Suggests an Equilibrium-Based Switch Activation Model." *The Plant Cell* 28 (1): 146–59.

Bernoux, Maud, Thomas Ve, Simon Williams, Christopher Warren, Danny Hatters, Eugene Valkov, Xiaoxiao Zhang, Jeffrey G. Ellis, Bostjan Kobe, and Peter N. Dodds. 2011. "Structural and Functional Analysis of a Plant Resistance Protein TIR Domain Reveals Interfaces for Self-Association, Signaling, and Autoregulation." *Cell Host & Microbe* 9 (3): 200–211.

Bolser, Dan, Daniel M. Staines, Emily Pritchard, and Paul Kersey. 2016. "Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomics Data." *Methods in Molecular Biology* 1374: 115–40.

Bomblies, Kirsten, Janne Lempe, Petra Epple, Norman Warthmann, Christa Lanz, Jeffery L. Dangl, and Detlef Weigel. 2007. "Autoimmune Response as a Mechanism for a Dobzhansky-Muller-Type Incompatibility Syndrome in Plants." *PLoS Biology* 5 (9): e236.

Bomblies, Kirsten, Levi Yant, Roosa A. Laitinen, Sang-Tae Kim, Jesse D. Hollister, Norman Warthmann, Joffrey Fitz, and Detlef Weigel. 2010. "Local-Scale Patterns of Genetic Variability, Outcrossing, and Spatial Structure in Natural Stands of Arabidopsis thaliana." *PLoS Genetics* 6 (3): e1000890.

Botella, M. A., J. E. Parker, L. N. Frost, P. D. Bittner-Eddy, J. L. Beynon, M. J. Daniels, E. B. Holub, and J. D. Jones. 1998. "Three Genes of the Arabidopsis RPP1 Complex Resistance Locus Recognize Distinct Peronospora Parasitica Avirulence Determinants." *The Plant Cell* 10 (11): 1847–60.

Brennan, Adrian C., Belén Méndez-Vigo, Abdelmajid Haddioui, José M. Martínez-Zapater, F. Xavier Picó, and Carlos Alonso-Blanco. 2014. "The Genetic Structure of Arabidopsis thaliana in the South-Western Mediterranean Range Reveals a Shared History between North Africa and Southern Europe." *BMC Plant Biology* 14 (January): 17.

Broman, Karl W., Hao Wu, Saunak Sen, and Gary A. Churchill. 2003. "R/qtl: QTL Mapping in Experimental Crosses." *Bioinformatics* 19 (7): 889–90.

Caddick, M. X., A. J. Greenland, I. Jepson, K. P. Krause, N. Qu, K. V. Riddell, M. G. Salter, W. Schuch, U. Sonnewald, and A. B. Tomsett. 1998. "An Ethanol Inducible Gene Switch for Plants Used to Manipulate Carbon Metabolism." *Nature Biotechnology* 16 (2): 177–80.

Caicedo, A. L., B. A. Schaal, and B. N. Kunkel. 1999. "Diversity and Molecular Evolution of the RPS2 Resistance Gene in Arabidopsis thaliana." *Proceedings of the National Academy of Sciences of the United States of America*

96 (1): 302–6.
Catchen, Julian, Paul A. Hohenlohe, Susan Bassham, Angel Amores, and William A. Cresko. 2013. "Stacks: An Analysis Tool Set for Population Genomics." *Molecular Ecology* 22 (11): 3124–40.
Chae, Eunyoung, Kirsten Bomblies, Sang-Tae Kim, Darya Karelina, Maricris Zaidem, Stephan Ossowski, Carmen Martín-Pizarro, et al. 2014. "Species-Wide Genetic Incompatibility Analysis Identifies Immune Genes as Hot Spots of Deleterious Epistasis." *Cell* 159 (6): 1341–51.
Chen, Chen, Hao Chen, You-Shun Lin, Jin-Bo Shen, Jun-Xiang Shan, Peng Qi, Min Shi, et al. 2014. "A Two-Locus Interaction Causes Interspecific Hybrid Weakness in Rice." *Nature Communications* 5: 3357.
Chen, Jingjing, Wenxing Pang, Bing Chen, Chunyu Zhang, and Zhongyun Piao. 2015. "Transcriptome Analysis of Brassica Rapa Near-Isogenic Lines Carrying Clubroot-Resistant and -Susceptible Alleles in Response to Plasmodiophora Brassicae during Early Infection." *Frontiers in Plant Science* 6: 1183.
Choi, Kyuha, Carsten Reinhard, Heïdi Serra, Piotr A. Ziolkowski, Charles J. Underwood, Xiaohui Zhao, Thomas J. Hardcastle, et al. 2016. "Recombination Rate Heterogeneity within Arabidopsis Disease Resistance Genes." *PLoS Genetics* 12 (7): e1006179.
Conant, Gavin C., and Kenneth H. Wolfe. 2008. "Turning a Hobby into a Job: How Duplicated Genes Find New Functions." *Nature Reviews. Genetics* 9 (12): 938–50.
Cooper, Jacob C., Ping Guo, Jackson Bladen, and Nitin Phadnis. 2019. "A Triple-Hybrid Cross Reveals a New Hybrid Incompatibility Locus between D. Melanogaster and D. Sechellia." *bioRxiv*. https://doi.org/10.1101/590588.
Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, et al. 2011. "The Variant Call Format and VCFtools." *Bioinformatics* 27 (15): 2156–58.
Delourme, R., N. Piel, R. Horvais, N. Pouilly, C. Domin, P. Vallee, C. Falentin, M. J. Manzanares-Dauleux, and M. Renard. 2008. "Molecular and Phenotypic Characterization of near Isogenic Lines at QTL for Quantitative Resistance to Leptosphaeria Maculans in Oilseed Rape (Brassica Napus L.)." *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 117 (7): 1055–67.
Deng, Jieqiong, Lei Fang, Xiefei Zhu, Baoliang Zhou, and Tianzhen Zhang. 2019. "A CC-NBS-LRR Gene Induces Hybrid Lethality in Cotton." *Journal of Experimental Botany* 70 (19): 5145–56.
Deng, Yiwen, Keran Zhai, Zhen Xie, Dongyong Yang, Xudong Zhu, Junzhong Liu, Xin Wang, et al. 2017. "Epigenetic Regulation of Antagonistic Receptors Confers Rice Blast Resistance with Yield Balance." *Science* 355 (6328): 962–65.
Dinesh-Kumar, S. P., W. H. Tham, and B. J. Baker. 2000. "Structure-Function Analysis of the Tobacco Mosaic Virus Resistance Gene N." *Proceedings of the National Academy of Sciences of the United States of America* 97 (26): 14789–94.
Ding, Pingtao, Bruno Pok Man Ngou, Oliver J. Furzer, Toshiyuki Sakai, Ram Krishna Shrestha, Dan MacLean, and Jonathan D. G. Jones. 2020. "High-resolution Expression Profiling of Selected Gene Sets during Plant Immune Activation." *Plant Biotechnology Journal*, January. https://doi.org/10.1111/pbi.13327.
Dodds, Peter N., and John P. Rathjen. 2010. "Plant Immunity: Towards an Integrated View of Plant-Pathogen Interactions." *Nature Reviews. Genetics* 11 (8): 539–48.
Dodds, P. N., G. J. Lawrence, and J. G. Ellis. 2001. "Six Amino Acid Changes Confined to the Leucine-Rich Repeat Beta-Strand/beta-Turn Motif Determine the Difference between the P and P2 Rust Resistance Specificities in Flax." *The Plant Cell* 13 (1): 163–78.
Dong, Oliver Xiaoou, Kevin Ao, Fang Xu, Kaeli C. M. Johnson, Yuxiang Wu, Lin Li, Shitou Xia, et al. 2018. "Individual Components of Paired Typical NLR Immune Receptors Are Regulated by Distinct E3 Ligases." *Nature Plants* 4 (9): 699–710.
Doyle, Jeff J., and Jan L. Doyle. 1987. "A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue." https://worldveg.tind.io/record/33886/.
Edgar, Robert C. 2004. "MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput." *Nucleic Acids Research* 32 (5): 1792–97.
Emms, David M., and Steven Kelly. 2019. "OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics." *Genome Biology* 20 (1): 238.
Eulgem, Thomas, and Imre E. Somssich. 2007. "Networks of WRKY Transcription Factors in Defense

Signaling." *Current Opinion in Plant Biology* 10 (4): 366–71.
Flynn, Jullien M., Robert Hubley, Clément Goubert, Jeb Rosen, Andrew G. Clark, Cédric Feschotte, and Arian F. Smit. 2020. "RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families." *Proceedings of the National Academy of Sciences of the United States of America* 117 (17): 9451–57.
Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait. 1999. "Preservation of Duplicate Genes by Complementary, Degenerative Mutations." *Genetics* 151 (4): 1531–45.
Fu, Fuyou, Xunjia Liu, Rui Wang, Chun Zhai, Gary Peng, Fengqun Yu, and W. G. Dilantha Fernando. 2019. "Fine Mapping of Brassica Napus Blackleg Resistance Gene Rlm1 through Bulked Segregant RNA Sequencing." *Scientific Reports* 9 (1): 14600.
Gloggnitzer, Jiradet, Svetlana Akimcheva, Arunkumar Srinivasan, Branislav Kusenda, Nina Riehs, Hansjörg Stampfl, Jaqueline Bautor, et al. 2014. "Nonsense-Mediated mRNA Decay Modulates Immune Receptor Levels to Regulate Plant Antibacterial Defense." *Cell Host & Microbe* 16 (3): 376–90.
Grimm, Dominik G., Damian Roqueiro, Patrice Salome, Stefan Kleeberger, Bastian Greshake, Wangsheng Zhu, Chang Liu, et al. 2016. "easyGWAS: A Cloud-Based Platform for Comparing the Results of Genome-Wide Association Studies." *The Plant Cell*, December. https://doi.org/10.1105/tpc.16.00551.
Guo, Ya-Long, Joffrey Fitz, Korbinian Schneeberger, Stephan Ossowski, Jun Cao, and Detlef Weigel. 2011. "Genome-Wide Comparison of Nucleotide-Binding Site-Leucine-Rich Repeat-Encoding Genes in Arabidopsis." *Plant Physiology* 157 (2): 757–69.
Harada, Akiko, Tatsuya Sakai, and Kiyotaka Okada. 2003. "Phot1 and phot2 Mediate Blue Light-Induced Transient Increases in Cytosolic Ca2+ Differently in Arabidopsis Leaves." *Proceedings of the National Academy of Sciences of the United States of America* 100 (14): 8583–88.
Holub, E. B. 2001. "The Arms Race Is Ancient History in Arabidopsis, the Wildflower." *Nature Reviews. Genetics* 2 (7): 516–27.
Jacob, Florence, Saskia Vernaldi, and Takaki Maekawa. 2013. "Evolution and Conservation of Plant NLR Functions." *Frontiers in Immunology* 4 (September): 297.
Jeuken, Marieke J. W., Ningwen W. Zhang, Leah K. McHale, Koen Pelgrom, Erik den Boer, Pim Lindhout, Richard W. Michelmore, Richard G. F. Visser, and Rients E. Niks. 2009. "Rin4 Causes Hybrid Necrosis and Race-Specific Resistance in an Interspecific Lettuce Hybrid." *The Plant Cell* 21 (10): 3368–78.
Jiao, Wen-Biao, and Korbinian Schneeberger. 2020. "Chromosome-Level Assemblies of Multiple Arabidopsis Genomes Reveal Hotspots of Rearrangements with Altered Evolutionary Dynamics." *Nature Communications* 11 (1): 989.
Jones, Jonathan D. G., and Jeffery L. Dangl. 2006. "The Plant Immune System." *Nature* 444 (7117): 323–29.
Jones, Jonathan D. G., Russell E. Vance, and Jeffery L. Dangl. 2016. "Intracellular Innate Immune Surveillance Devices in Plants and Animals." *Science* 354 (6316): aaf6395.
Jones, Philip, David Binns, Hsin-Yu Chang, Matthew Fraser, Weizhong Li, Craig McAnulla, Hamish McWilliam, et al. 2014. "InterProScan 5: Genome-Scale Protein Function Classification." *Bioinformatics* 30 (9): 1236–40.
Kahle, David, and Hadley Wickham. 2013. "Ggmap: Spatial Visualization with ggplot2." *The R Journal* 5 (1): 144–61.
Karasov, Talia L., Joel M. Kniskern, Liping Gao, Brody J. DeYoung, Jing Ding, Ullrich Dubiella, Ruben O. Lastra, et al. 2014. "The Long-Term Maintenance of a Resistance Polymorphism through Diffuse Interactions." *Nature* 512 (7515): 436–40.
Khan, Mehmood Alam, Isaac Elias, Erik Sjölund, Kristina Nylander, Roman Valls Guimera, Richard Schobesberger, Peter Schmitzberger, Jens Lagergren, and Lars Arvestad. 2013. "Fastphylo: Fast Tools for Phylogenetics." *BMC Bioinformatics* 14 (November): 334.
Kim, Daehwan, Ben Langmead, and Steven L. Salzberg. 2015. "HISAT: A Fast Spliced Aligner with Low Memory Requirements." *Nature Methods* 12 (4): 357–60.
Kim, Seungill, Jieun Park, Seon-In Yeom, Yong-Min Kim, Eunyoung Seo, Ki-Tae Kim, Myung-Shin Kim, et al. 2017. "New Reference Genome Sequences of Hot Pepper Reveal the Massive Evolution of Plant Disease-Resistance Genes by Retroduplication." *Genome Biology* 18 (1): 210.
Klepikova, Anna V., Artem S. Kasianov, Evgeny S. Gerasimov, Maria D. Logacheva, and Aleksey A. Penin. 2016. "A High Resolution Map of the Arabidopsis thaliana Developmental Transcriptome Based on RNA-Seq

Profiling." *The Plant Journal: For Cell and Molecular Biology* 88 (6): 1058–70.
Kolde, Raivo. 2012. "Pheatmap: Pretty Heatmaps." *R Package Version* 61: 617.
Koren, Sergey, Brian P. Walenz, Konstantin Berlin, Jason R. Miller, Nicholas H. Bergman, and Adam M. Phillippy. 2017. "Canu: Scalable and Accurate Long-Read Assembly via Adaptive K-Mer Weighting and Repeat Separation." *Genome Research* 27 (5): 722–36.
Krasileva, Ksenia V., Douglas Dahlbeck, and Brian J. Staskawicz. 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." *The Plant Cell* 22 (7): 2444–58.
Krüger, Julia, Colwyn M. Thomas, Catherine Golstein, Mark S. Dixon, Matthew Smoker, Saijun Tang, Lonneke Mulder, and Jonathan D. G. Jones. 2002. "A Tomato Cysteine Protease Required for Cf-2-Dependent Disease Resistance and Suppression of Autonecrosis." *Science* 296 (5568): 744–47.
Lai, Yan, and Thomas Eulgem. 2018. "Transcript-Level Expression Control of Plant NLR Genes." *Molecular Plant Pathology* 19 (5): 1267–81.
Langmead, Ben, and Steven L. Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59.
Larsson, Anders. 2014. "AliView: A Fast and Lightweight Alignment Viewer and Editor for Large Datasets." *Bioinformatics* 30 (22): 3276–78.
Leister, Dario. 2004. "Tandem and Segmental Gene Duplication and Recombination in the Evolution of Plant Disease Resistance Gene." *Trends in Genetics: TIG* 20 (3): 116–22.
Letunic, Ivica, and Peer Bork. 2007. "Interactive Tree Of Life (iTOL): An Online Tool for Phylogenetic Tree Display and Annotation." *Bioinformatics* 23 (1): 127–28.
Li, Bo, and Colin N. Dewey. 2011. "RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome." *BMC Bioinformatics* 12 (August): 323.
Li, Feng, Daniela Pignatta, Claire Bendix, Jacob O. Brunkard, Megan M. Cohn, Jeffery Tung, Haoyu Sun, Pavan Kumar, and Barbara Baker. 2012. "MicroRNA Regulation of Plant Innate Immune Receptors." *Proceedings of the National Academy of Sciences of the United States of America* 109 (5): 1790–95.
Li, Heng. 2013. "Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM." *arXiv* http://arxiv.org/abs/1303.3997.
Li, Lei, Anette Habring, Kai Wang, and Detlef Weigel. 2020. "Atypical Resistance Protein RPW8/HR Triggers Oligomerization of the NLR Immune Receptor RPP7 and Autoimmunity." *Cell Host & Microbe*, February. https://doi.org/10.1016/j.chom.2020.01.012.
Linthorst, Jasper, Marc Hulsman, Henne Holstege, and Marcel Reinders. 2015. "Scalable Multi Whole-Genome Alignment Using Recursive Exact Matching." *bioRxiv*. https://doi.org/10.1101/022715.
Li, Yongqing, Britney O. Pennington, and Jian Hua. 2009. "Multiple R-like Genes Are Negatively Regulated by BON1 and BON3 in Arabidopsis." *Molecular Plant-Microbe Interactions: MPMI* 22 (7): 840–48.
Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12): 550.
Lynch, M., and J. S. Conery. 2000. "The Evolutionary Fate and Consequences of Duplicate Genes." *Science*.
Mackey, David, Youssef Belkhadir, Jose M. Alonso, Joseph R. Ecker, and Jeffery L. Dangl. 2003. "Arabidopsis RIN4 Is a Target of the Type III Virulence Effector AvrRpt2 and Modulates RPS2-Mediated Resistance." *Cell* 112 (3): 379–89.
MacQueen, Alice, and Joy Bergelson. 2016. "Modulation of R-Gene Expression across Environments." *Journal of Experimental Botany* 67 (7): 2093–2105.
MacQueen, Alice, Xiaoqin Sun, and Joy Bergelson. 2016. "Genetic Architecture and Pleiotropy Shape Costs of Rps2-Mediated Resistance in Arabidopsis thaliana." *Nature Plants* 2 (July): 16110.
Maekawa, Takaki, Thomas A. Kufer, and Paul Schulze-Lefert. 2011. "NLR Functions in Plant and Animal Immune Systems: So Far and yet so Close." *Nature Immunology* 12 (9): 817–26.
Marchal, Clemence, Jianping Zhang, Peng Zhang, Paul Fenwick, Burkhard Steuernagel, Nikolai M. Adamski, Lesley Boyd, et al. 2018. "BED-Domain-Containing Immune Receptors Confer Diverse Resistance Spectra to Yellow Rust." *Nature Plants* 4 (9): 662–68.
Martin, Raoul, Tiancong Qi, Haibo Zhang, Furong Liu, Miles King, Claire Toth, Eva Nogales, and Brian J.

Staskawicz. 2020. "Structure of the Activated Roq1 Resistosome Directly Recognizing the Pathogen Effector XopQ." https://doi.org/10.1101/2020.08.13.246413.
Mauricio, Rodney, Eli A. Stahl, Tonia Korves, Dacheng Tian, Martin Kreitman, and Joy Bergelson. 2003. "Natural Selection for Polymorphism in the Disease Resistance Gene Rps2 of Arabidopsis thaliana." *Genetics* 163 (2): 735–46.
McDowell, J. M., A. Cuzick, C. Can, J. Beynon, J. L. Dangl, and E. B. Holub. 2000. "Downy Mildew (Peronospora Parasitica) Resistance Genes in Arabidopsis Vary in Functional Requirements for NDR1, EDS1, NPR1 and Salicylic Acid Accumulation." *The Plant Journal: For Cell and Molecular Biology* 22 (6): 523–29.
McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, et al. 2010. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data." *Genome Research* 20 (9): 1297–1303.
Meyers, Blake C., Shail Kaushik, and Raja Sekhar Nandety. 2005. "Evolving Disease Resistance Genes." *Current Opinion in Plant Biology* 8 (2): 129–34.
Meyers, Blake C., Alexander Kozik, Alyssa Griego, Hanhui Kuang, and Richard W. Michelmore. 2003. "Genome-Wide Analysis of NBS-LRR-Encoding Genes in Arabidopsis." *The Plant Cell* 15 (4): 809–34.
Michelmore, R. W., and B. C. Meyers. 1998. "Clusters of Resistance Genes in Plants Evolve by Divergent Selection and a Birth-and-Death Process." *Genome Research* 8 (11): 1113–30.
Mine, Akira, Carolin Seyfferth, Barbara Kracher, Matthias L. Berens, Dieter Becker, and Kenichi Tsuda. 2018. "The Defense Phytohormone Signaling Network Enables Rapid, High-Amplitude Transcriptional Reprogramming during Effector-Triggered Immunity." *The Plant Cell* 30 (6): 1199–1219.
Mohr, Toni J., Nicole D. Mammarella, Troy Hoff, Bonnie J. Woffenden, John G. Jelesko, and John M. McDowell. 2010. "The Arabidopsis Downy Mildew Resistance Gene RPP8 Is Induced by Pathogens and Salicylic Acid and Is Regulated by W Box Cis Elements." *Molecular Plant-Microbe Interactions: MPMI* 23 (10): 1303–15.
Mondragón-Palomino, Mariana, Blake C. Meyers, Richard W. Michelmore, and Brandon S. Gaut. 2002. "Patterns of Positive Selection in the Complete NBS-LRR Gene Family of Arabidopsis thaliana." *Genome Research* 12 (9): 1305–15.
Nei, M., and T. Gojobori. 1986. "Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions." *Molecular Biology and Evolution* 3 (5): 418–26.
Nishimura, Marc T., Ryan G. Anderson, Karen A. Cherkis, Terry F. Law, Qingli L. Liu, Mischa Machius, Zachary L. Nimchuk, et al. 2017. "TIR-Only Protein RBA1 Recognizes a Pathogen Effector to Regulate Cell Death in Arabidopsis." *Proceedings of the National Academy of Sciences of the United States of America* 114 (10): E2053–62.
Ohno, Susumu. 1970. *Evolution by Gene Duplication*. Springer, Berlin, Heidelberg.
Olson, M. V. 1999. "When Less Is More: Gene Loss as an Engine of Evolutionary Change." *American Journal of Human Genetics* 64 (1): 18–23.
Palma, Kristoffer, Stephan Thorgrimsen, Frederikke Gro Malinovsky, Berthe Katrine Fiil, H. Bjørn Nielsen, Peter Brodersen, Daniel Hofius, Morten Petersen, and John Mundy. 2010. "Autoimmunity in Arabidopsis acd11 Is Mediated by Epigenetic Regulation of an Immune Receptor." *PLoS Pathogens* 6 (10): e1001137.
Pandey, Shree P., and Imre E. Somssich. 2009. "The Role of WRKY Transcription Factors in Plant Immunity." *Plant Physiology* 150 (4): 1648–55.
Papadopoulou, Galini V., Anne Maedicke, Katharina Grosser, Nicole M. van Dam, and Ainhoa Martínez-Medina. 2018. "Defence Signalling Marker Gene Responses to Hormonal Elicitation Differ between Roots and Shoots." *AoB Plants* 10 (3): ly031.
Patterson, Nick, Alkes L. Price, and David Reich. 2006. "Population Structure and Eigenanalysis." *PLoS Genetics* 2 (12): e190.
Picó, F. Xavier, Belén Méndez-Vigo, José M. Martínez-Zapater, and Carlos Alonso-Blanco. 2008. "Natural Genetic Variation of Arabidopsis thaliana Is Geographically Structured in the Iberian Peninsula." *Genetics* 180 (2): 1009–21.
Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics* 81 (3): 559–75.

Qi, Tiancong, Kyungyong Seong, Daniela P. T. Thomazella, Joonyoung Ryan Kim, Julie Pham, Eunyoung Seo, Myeong-Je Cho, Alex Schultink, and Brian J. Staskawicz. 2018. "NRG1 Functions Downstream of EDS1 to Regulate TIR-NLR-Mediated Plant Immunity in Nicotiana Benthamiana." *Proceedings of the National Academy of Sciences of the United States of America* 115 (46): E10979–87.

Ribot, Cécile, Judith Hirsch, Sandrine Balzergue, Didier Tharreau, Jean-Loup Nottéghem, Marc-Henri Lebrun, and Jean-Benoit Morel. 2008. "Susceptibility of Rice to the Blast Fungus, Magnaporthe Grisea." *Journal of Plant Physiology* 165 (1): 114–24.

Rose, Laura, Susanna Atwell, Murray Grant, and Eric B. Holub. 2012. "Parallel Loss-of-Function at the RPM1 Bacterial Resistance Locus in Arabidopsis thaliana." *Frontiers in Plant Science* 3 (December): 287.

Roth, Charlotte, Daniel Lüdke, Melanie Klenke, Annalena Quathamer, Oliver Valerius, Gerhard H. Braus, and Marcel Wiermer. 2017. "The Truncated NLR Protein TIR-NBS13 Is a MOS6/IMPORTIN- α 3 Interaction Partner Required for Plant Immunity." *The Plant Journal: For Cell and Molecular Biology* 92 (5): 808–21.

Rowan, Beth A., Danelle K. Seymour, Eunyoung Chae, Derek S. Lundberg, and Detlef Weigel. 2017. "Methods for Genotyping-by-Sequencing." *Methods in Molecular Biology* 1492: 221–42.

Ruggieri, Valentino, Angelina Nunziata, and Amalia Barone. 2014. "Positive Selection in the Leucine-Rich Repeat Domain of Gro1 Genes in Solanum Species." *Journal of Genetics* 93 (3): 755–65.

Sandstedt, Gabrielle D., Carrie A. Wu, and Andrea L. Sweigart. 2020. "Evolution of Multiple Postzygotic Barriers between Species in the Mimulus Tilingii Species Complex." https://doi.org/10.1101/2020.08.07.241489.

Schwab, Rebecca, Stephan Ossowski, Markus Riester, Norman Warthmann, and Detlef Weigel. 2006. "Highly Specific Gene Silencing by Artificial microRNAs in Arabidopsis." *The Plant Cell* 18 (5): 1121–33.

Shao, Zhu-Qing, Jia-Yu Xue, Ping Wu, Yan-Mei Zhang, Yue Wu, Yue-Yu Hang, Bin Wang, and Jian-Qun Chen. 2016. "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.

Shirano, Yumiko, Pradeep Kachroo, Jyoti Shah, and Daniel F. Klessig. 2002. "A Gain-of-Function Mutation in an Arabidopsis Toll Interleukin1 Receptor-Nucleotide Binding Site-Leucine-Rich Repeat Type R Gene Triggers Defense Responses and Results in Enhanced Disease Resistance." *The Plant Cell* 14 (12): 3149–62.

Shivaprasad, Padubidri V., Ho-Ming Chen, Kanu Patel, Donna M. Bond, Bruno A. C. M. Santos, and David C. Baulcombe. 2012. "A microRNA Superfamily Regulates Nucleotide Binding Site-Leucine-Rich Repeats and Other mRNAs." *The Plant Cell* 24 (3): 859–74.

Sicard, Adrien, Christian Kappel, Emily B. Josephs, Young Wha Lee, Cindy Marona, John R. Stinchcombe, Stephen I. Wright, and Michael Lenhard. 2015. "Divergent Sorting of a Balanced Ancestral Polymorphism Underlies the Establishment of Gene-Flow Barriers in Capsella." *Nature Communications* 6 (August): 7960.

Sohn, Kee Hoon, Cécile Segonzac, Ghanasyam Rallapalli, Panagiotis F. Sarris, Joo Yong Woo, Simon J. Williams, Toby E. Newman, Kyung Hee Paek, Bostjan Kobe, and Jonathan D. G. Jones. 2014. "The Nuclear Immune Receptor RPS4 Is Required for RRS1SLH1-Dependent Constitutive Defense Activation in Arabidopsis thaliana." *PLoS Genetics* 10 (10): e1004655.

Staal, Jens, Maria Kaliff, Svante Bohman, and Christina Dixelius. 2006. "Transgressive Segregation Reveals Two Arabidopsis TIR-NB-LRR Resistance Genes Effective against Leptosphaeria Maculans, Causal Agent of Blackleg Disease." *The Plant Journal: For Cell and Molecular Biology* 46 (2): 218–30.

Stahl, E. A., G. Dwyer, R. Mauricio, M. Kreitman, and J. Bergelson. 1999. "Dynamics of Disease Resistance Polymorphism at the Rpm1 Locus of Arabidopsis." *Nature* 400 (6745): 667–71.

Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30 (9): 1312–13.

Stanke, Mario, Oliver Schöffmann, Burkhard Morgenstern, and Stephan Waack. 2006. "Gene Prediction in Eukaryotes with a Generalized Hidden Markov Model That Uses Hints from External Sources." *BMC Bioinformatics* 7 (February): 62.

Steinbrenner, Adam D., Sandra Goritschnig, and Brian J. Staskawicz. 2015. "Recognition and Activation Domains Contribute to Allele-Specific Responses of an Arabidopsis NLR Receptor to an Oomycete Effector Protein." *PLoS Pathogens* 11 (2): e1004665.

Steuernagel, Burkhard, Kamil Witek, Simon G. Krattinger, Ricardo H. Ramirez-Gonzalez, Henk-Jan Schoonbeek, Guotai Yu, Erin Baggs, et al. 2020. "The NLR-Annotator Tool Enables Annotation of the Intracellular Immune Receptor Repertoire." *Plant Physiology* 183 (2): 468–82.

Stokes, Trevor L., Barbara N. Kunkel, and Eric J. Richards. 2002. "Epigenetic Variation in Arabidopsis Disease Resistance." *Genes & Development* 16 (2): 171–82.

Sueldo, Daniela J., Mahdere Shimels, Laurentiu N. Spiridon, Octav Caldararu, Andrei-Jose Petrescu, Matthieu H. A. J. Joosten, and Wladimir I. L. Tameling. 2015. "Random Mutagenesis of the Nucleotide-Binding Domain of NRC1 (NB-LRR Required for Hypersensitive Response-Associated Cell Death-1), a Downstream Signalling Nucleotide-Binding, Leucine-Rich Repeat (NB-LRR) Protein, Identifies Gain-of-Function Mutations in the Nucleotide-Binding Pocket." *The New Phytologist* 208 (1): 210–23.

Supek, Fran, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. 2011. "REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms." *PloS One* 6 (7): e21800.

Świadek, Magdalena, Sebastian Proost, Daniela Sieh, Jing Yu, Marco Todesco, Christian Jorzig, Andrés Eduardo Rodriguez Cubillos, et al. 2017. "Novel Allelic Variants in ACD6 Cause Hybrid Necrosis in Local Collection of Arabidopsis thaliana." *The New Phytologist* 213 (2): 900–915.

Swiderski, Michal R., Doris Birker, and Jonathan D. G. Jones. 2009. "The TIR Domain of TIR-NB-LRR Resistance Proteins Is a Signaling Domain Involved in Cell Death Induction." *Molecular Plant-Microbe Interactions: MPMI* 22 (2): 157–65.

Tameling, Wladimir I. L., Sandra D. J. Elzinga, Patricia S. Darmin, Jack H. Vossen, Frank L. W. Takken, Michel A. Haring, and Ben J. C. Cornelissen. 2002. "The Tomato R Gene Products I-2 and MI-1 Are Functional ATP Binding Proteins with ATPase Activity." *The Plant Cell* 14 (11): 2929–39.

Tan, Xiaoping, Blake C. Meyers, Alexander Kozik, Marilyn A. L. West, Michele Morgante, Dina A. St Clair, Andrew F. Bent, and Richard W. Michelmore. 2007. "Global Expression Analysis of Nucleotide Binding Site-Leucine Rich Repeat-Encoding and Related Genes in Arabidopsis." *BMC Plant Biology* 7 (October): 56.

Tian, Tian, Yue Liu, Hengyu Yan, Qi You, Xin Yi, Zhou Du, Wenying Xu, and Zhen Su. 2017. "agriGO v2.0: A GO Analysis Toolkit for the Agricultural Community, 2017 Update." *Nucleic Acids Research* 45 (W1): W122–29.

Todesco, Marco, Sang-Tae Kim, Eunyoung Chae, Kirsten Bomblies, Maricris Zaidem, Lisa M. Smith, Detlef Weigel, and Roosa A. E. Laitinen. 2014. "Activation of the Arabidopsis thaliana Immune System by Combinations of Common ACD6 Alleles." *PLoS Genetics* 10 (7): e1004459.

Tsuchiya, Tokuji, and Thomas Eulgem. 2013. "An Alternative Polyadenylation Mechanism Coopted to the Arabidopsis RPP7 Gene through Intronic Retrotransposon Domestication." *Proceedings of the National Academy of Sciences of the United States of America* 110 (37): E3535–43.

Van de Weyer, Anna-Lena, Freddy Monteiro, Oliver J. Furzer, Marc T. Nishimura, Volkan Cevik, Kamil Witek, Jonathan D. G. Jones, Jeffery L. Dangl, Detlef Weigel, and Felix Bemm. 2019. "A Species-Wide Inventory of NLR Genes and Alleles in Arabidopsis thaliana." *Cell* 178 (5): 1260–72.e14.

Van Ghelder, Cyril, and Daniel Esmenjaud. 2016. "TNL Genes in Peach: Insights into the Post-LRR Domain." *BMC Genomics* 17 (April): 317.

Walker, Bruce J., Thomas Abeel, Terrance Shea, Margaret Priest, Amr Abouelliel, Sharadha Sakthikumar, Christina A. Cuomo, et al. 2014. "Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement." *PloS One* 9 (11): e112963.

Wang, Dapeng, Yubin Zhang, Zhang Zhang, Jiang Zhu, and Jun Yu. 2010. "KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies." *Genomics, Proteomics & Bioinformatics* 8 (1): 77–80.

Wang, Guan-Feng, Jiabing Ji, Farid El-Kasmi, Jeffery L. Dangl, Guri Johal, and Peter J. Balint-Kurti. 2015. "Molecular and Functional Analyses of a Maize Autoactive NB-LRR Protein Identify Precise Structural Requirements for Activity." *PLoS Pathogens* 11 (2): e1004674.

Weigel, Detlef, and Jane Glazebrook. 2002. *Arabidopsis: A Laboratory Manual*. CSHL Press.

Wicker, Thomas, Nabila Yahiaoui, and Beat Keller. 2007. "Illegitimate Recombination Is a Major Evolutionary Mechanism for Initiating Size Variation in Plant Resistance Genes." *The Plant Journal: For Cell and Molecular Biology* 51 (4): 631–41.

Wickham, Hadley. 2009. *Ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. Springer Publishing Company, Incorporated.

Williams, Simon J., Pradeep Sornaraj, Emma deCourcy-Ireland, R. Ian Menz, Bostjan Kobe, Jeffrey G. Ellis, Peter N. Dodds, and Peter A. Anderson. 2011. "An Autoactive Mutant of the M Flax Rust Resistance Protein Has a Preference for Binding ATP, Whereas Wild-Type M Protein Binds ADP." *Molecular Plant-Microbe Interactions: MPMI* 24 (8): 897–906.

Wong, Simon, and Kenneth H. Wolfe. 2005. "Birth of a Metabolic Gene Cluster in Yeast by Adaptive Gene Relocation." *Nature Genetics* 37 (7): 777–82.

Workman, R., W. Timp, R. Fedak, D. Kilburn, S. Hao, and K. Liu. 2018. "High Molecular Weight DNA Extraction from Recalcitrant Plant Species for Third Generation Sequencing." *Protocol Exchange*. https://doi.org/ 10.1038/protex.2018.059.

Xia, Rui, Jing Xu, Siwaret Arikit, and Blake C. Meyers. 2015. "Extensive Families of miRNAs and PHAS Loci in Norway Spruce Demonstrate the Origins of Complex phasiRNA Networks in Seed Plants." *Molecular Biology and Evolution* 32 (11): 2905–18.

Yaffe, Hila, Kobi Buxdorf, Illil Shapira, Shachaf Ein-Gedi, Michal Moyal-Ben Zvi, Eyal Fridman, Menachem Moshelion, and Maggie Levy. 2012. "LogSpin: A Simple, Economical and Fast Method for RNA Isolation from Infected or Healthy Plants and Other Eukaryotic Tissues." *BMC Research Notes* 5 (January): 45.

Yamamoto, Eiji, Tomonori Takashi, Yoichi Morinaka, Shaoyang Lin, Jianzhong Wu, Takashi Matsumoto, Hidemi Kitano, Makoto Matsuoka, and Motoyuki Ashikari. 2010. "Gain of Deleterious Function Causes an Autoimmune Response and Bateson-Dobzhansky-Muller Incompatibility in Rice." *Molecular Genetics and Genomics: MGG* 283 (4): 305–15.

Yang, Shuhua, and Jian Hua. 2004. "A Haplotype-Specific Resistance Gene Regulated by BONZAI1 Mediates Temperature-Dependent Growth Control in Arabidopsis." *The Plant Cell* 16 (4): 1060–71.

Yi, Hankuil, and Eric J. Richards. 2007. "A Cluster of Disease Resistance Genes in Arabidopsis Is Coordinately Regulated by Transcriptional Activation and RNA Silencing." *The Plant Cell* 19 (9): 2929–39.

Yu, Agnès, Gersende Lepère, Florence Jay, Jingyu Wang, Laure Bapaume, Yu Wang, Anne-Laure Abraham, et al. 2013. "Dynamics and Biological Relevance of DNA Demethylation in Arabidopsis Antibacterial Defense." *Proceedings of the National Academy of Sciences of the United States of America* 110 (6): 2389–94.

Zhai, Jixian, Dong-Hoon Jeong, Emanuele De Paoli, Sunhee Park, Benjamin D. Rosen, Yupeng Li, Alvaro J. González, et al. 2011. "MicroRNAs as Master Regulators of the Plant NB-LRR Defense Gene Family via the Production of Phased, Trans-Acting siRNAs." *Genes & Development* 25 (23): 2540–53.

Zhang, Xiaoxiao, Maud Bernoux, Adam R. Bentham, Toby E. Newman, Thomas Ve, Lachlan W. Casey, Tom M. Raaymakers, et al. 2017. "Multiple Functional Self-Association Interfaces in Plant TIR Domains." *Proceedings of the National Academy of Sciences of the United States of America* 114 (10): E2046–52.

Zhang, Yao, Yuancong Wang, Jingyan Liu, Yanglin Ding, Shanshan Wang, Xiaoyan Zhang, Yule Liu, and Shuhua Yang. 2017. "Temperature-Dependent Autoimmunity Mediated by chs1 Requires Its Neighboring TNL Gene SOC3." *The New Phytologist* 213 (3): 1330–45.

Zhao, Ting, Lu Rui, Juan Li, Marc T. Nishimura, John P. Vogel, Na Liu, Simu Liu, Yaofei Zhao, Jeffery L. Dangl, and Dingzhong Tang. 2015. "A Truncated NLR Protein, TIR-NBS2, Is Required for Activated Defense Responses in the exo70B1 Mutant." *PLoS Genetics* 11 (1): e1004945.

Zheng, Xianwu, Olga Pontes, Jianhua Zhu, Daisuke Miki, Fei Zhang, Wen-Xue Li, Kei Iida, Avnish Kapoor, Craig S. Pikaard, and Jian-Kang Zhu. 2008. "ROS3 Is an RNA-Binding Protein Required for DNA Demethylation in Arabidopsis." *Nature* 455 (7217): 1259–62.

Zipfel, Cyril, Silke Robatzek, Lionel Navarro, Edward J. Oakeley, Jonathan D. G. Jones, Georg Felix, and Thomas Boller. 2004. "Bacterial Disease Resistance in Arabidopsis through Flagellin Perception." *Nature* 428 (6984): 764–67.

Supplemental Material

A truncated singleton NLR causes hybrid necrosis in *Arabidopsis thaliana*

A. Cristina Barragan[1], Maximilian Collenberg[1], Jinge Wang[2], Rachelle R.Q. Lee[2], Wei Yuan Cher[2], Fernando A. Rabanal[1], Haim Ashkenazy[1], Detlef Weigel[1]*, Eunyoung Chae[1,2]*

Supplemental Figures

Fig S1. RNA-seq analysis of Cdm-0 x TueScha-9 hybrid plants. Related to Fig 1.

Fig S2. Identification of *DM10* and *DM11*. Related to Fig 2.

Fig S3. De novo Cdm-0 genome assembly. Related to Fig 2.

Fig S4. Pairwise genetic distances for three *DM10* candidate genes across 80 accessions. Related to Fig 3.

Fig S5. *DM10* natural variation. Related to Fig 5.

Fig S6. RNA-seq analysis of Ler/Kas-2 NIL vs. Kas-2 plants. Related to Fig 1.

Supplemental Methods

Supplemental References

Supplemental Figures

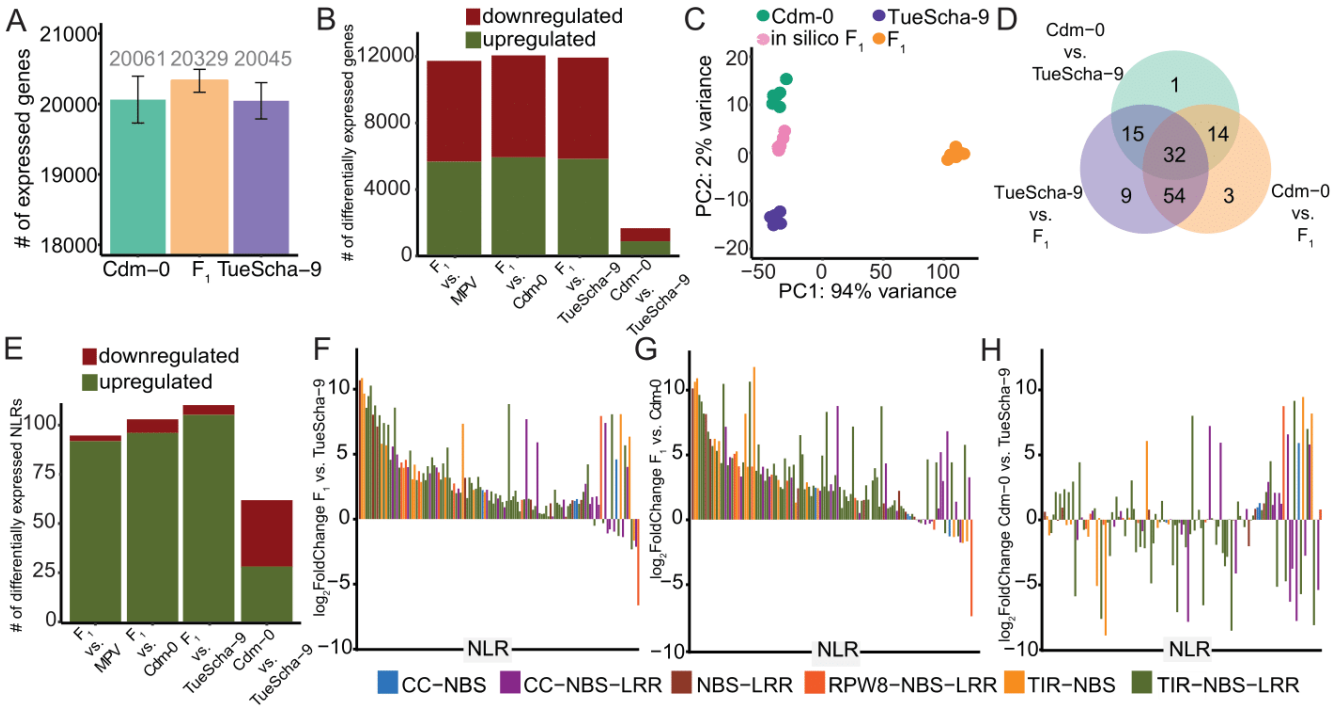


Fig S1. RNA-seq analysis of Cdm-0 x TueScha-9 hybrid plants. A. Total number of expressed genes in both the F_1 hybrid and parents. **B.** Significantly ($|\log_2 FoldChange| > 1$, padj value < 0.01) up- and downregulated genes across different genotype comparisons. **C.** PCA of gene expression variance separating the F_1 hybrids, parents and in silico hybrids. **D.** Intersection of differentially expressed NLRs between the F_1 hybrid and parents. **E.** Significantly up- and downregulated NLR genes across different genotype comparisons. **F-H.** NLR expression changes between the F_1 hybrid and TueScha-9 (F), F_1 hybrid and Cdm-0 (G), Cdm-0 and TueScha-9 (H). The NLR gene order follows Fig 1G.

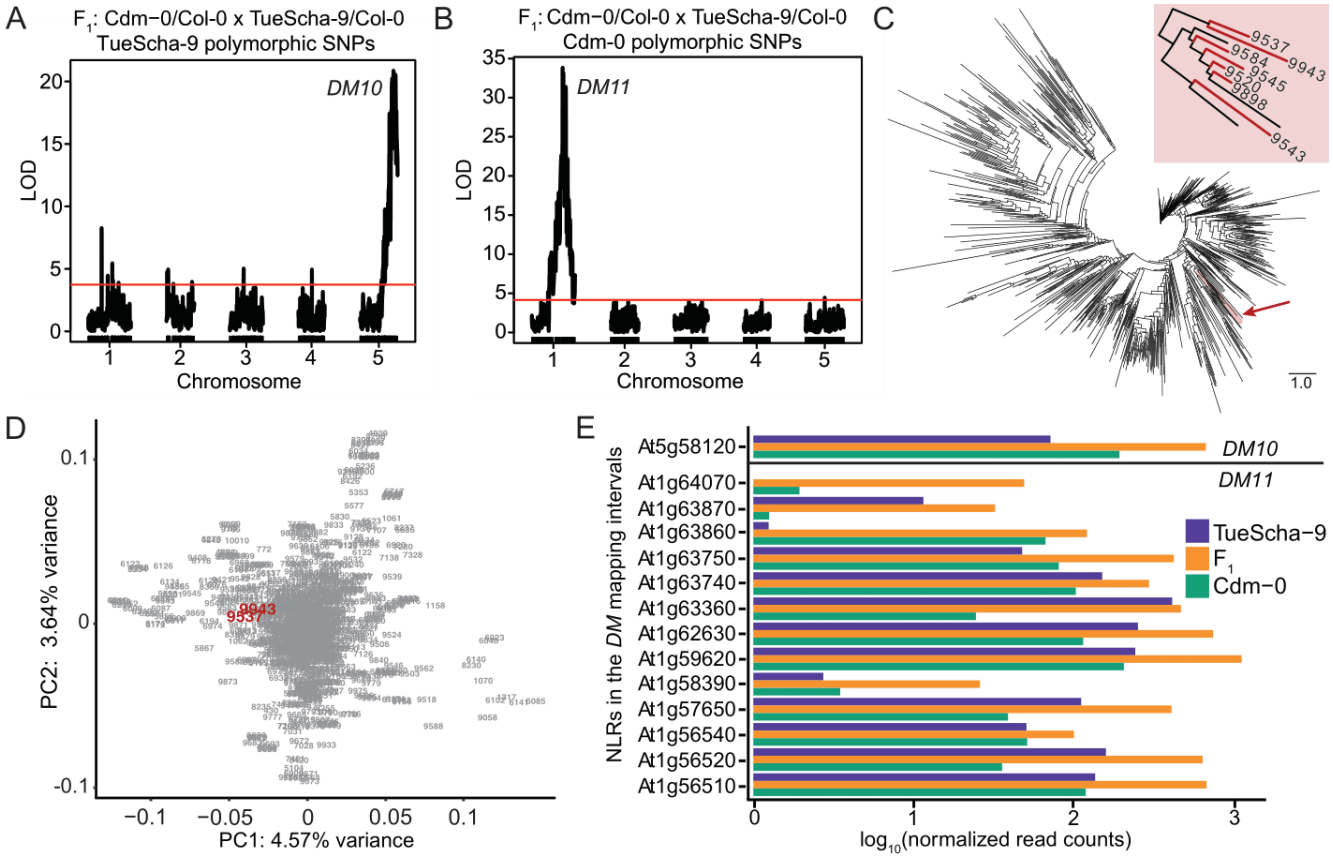


Fig S2. Identification of *DM10* and *DM11*. A. Polymorphic SNPs from TueScha-9, *DM10* on chromosome 5 (22.35-24.45 Mb). **B.** Polymorphic SNPs from Cdm-0, *DM11* on chromosome 1 (21.55-22.18 Mb). Horizontal lines indicate 0.05 significance threshold established with 1,000 permutations. **C.** Example of an NJ tree from one of the *DM11* candidate loci: At1g59780. Region where Cdm-0 (9943) and IP-Cum-1 (9537) are found is highlighted in red. Inset shows a close-up of Cdm-0-like accessions, accessions crossed to TueScha-9 are marked in red. **D.** Example of a PCA plot using VCF information for the entire *DM11* mapping interval from the 1001 Genomes Project (1001 Genomes Consortium 2016). Accession IDs from the 1001 Genomes Project in grey, with Cdm-0 (9943) and IP-Cum-1 (9537) in red. **E.** Normalized RNA-seq read counts for the hybrid and parents. Shown are the only NLR in the *DM10* mapping interval, At5g58120, as well as NLRs found between At1g56510 and At1g64070 on chromosome 1, which includes the *DM11* interval. Missing bars mean the gene was not expressed in that genotype.

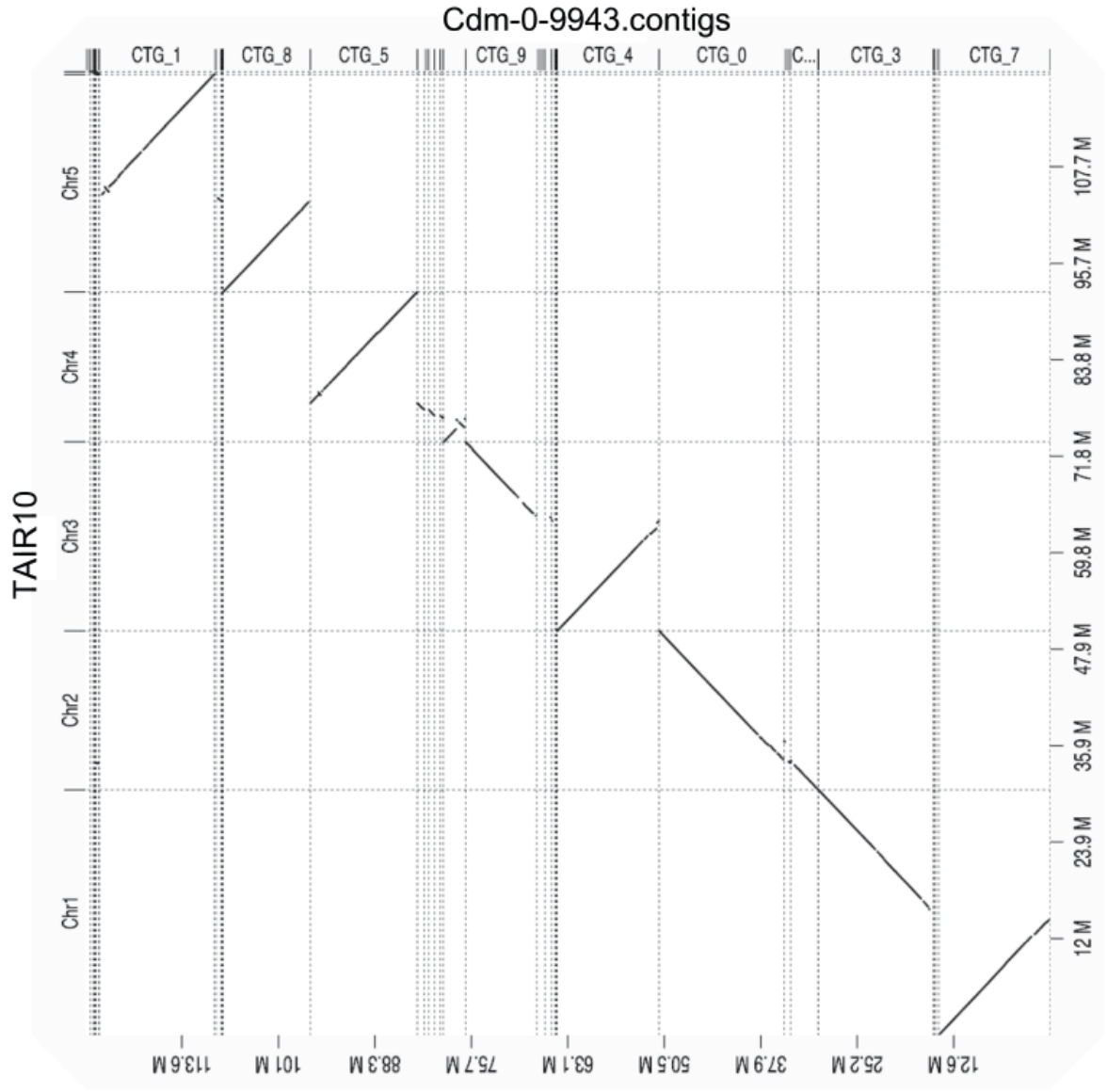


Fig S3. De novo Cdm-0 genome assembly. Dot plot based on minimap2 (Li 2018) alignment between the Cdm-0 contigs and the reference genome (TAIR10) using D-GENIES (Cabanettes and Klopp 2018).

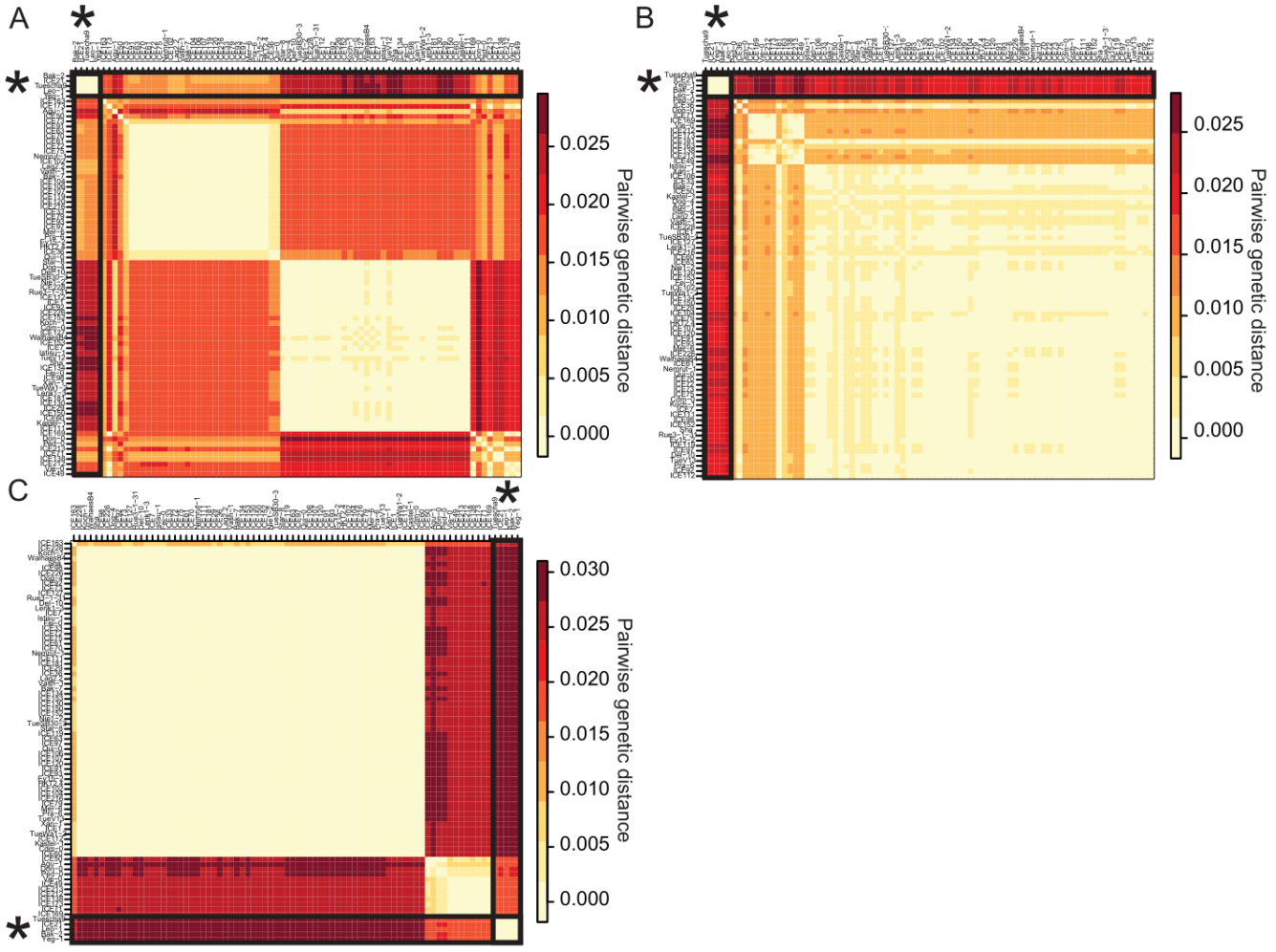


Fig S4. Pairwise genetic distances for *DM10* three candidate genes across 80 accessions. A-C. Heatmaps of pairwise genetic distances among 80 alleles of At5g58120 (A), *ROS3* (B) and *PHOT2* (C). Distances are the fraction of nucleotide sites at which two sequences are different. Asterisk highlights the group of five risk accessions causing hybrid necrosis when crossed to Cdm-0. These five accessions are genetically very similar in all three genes.

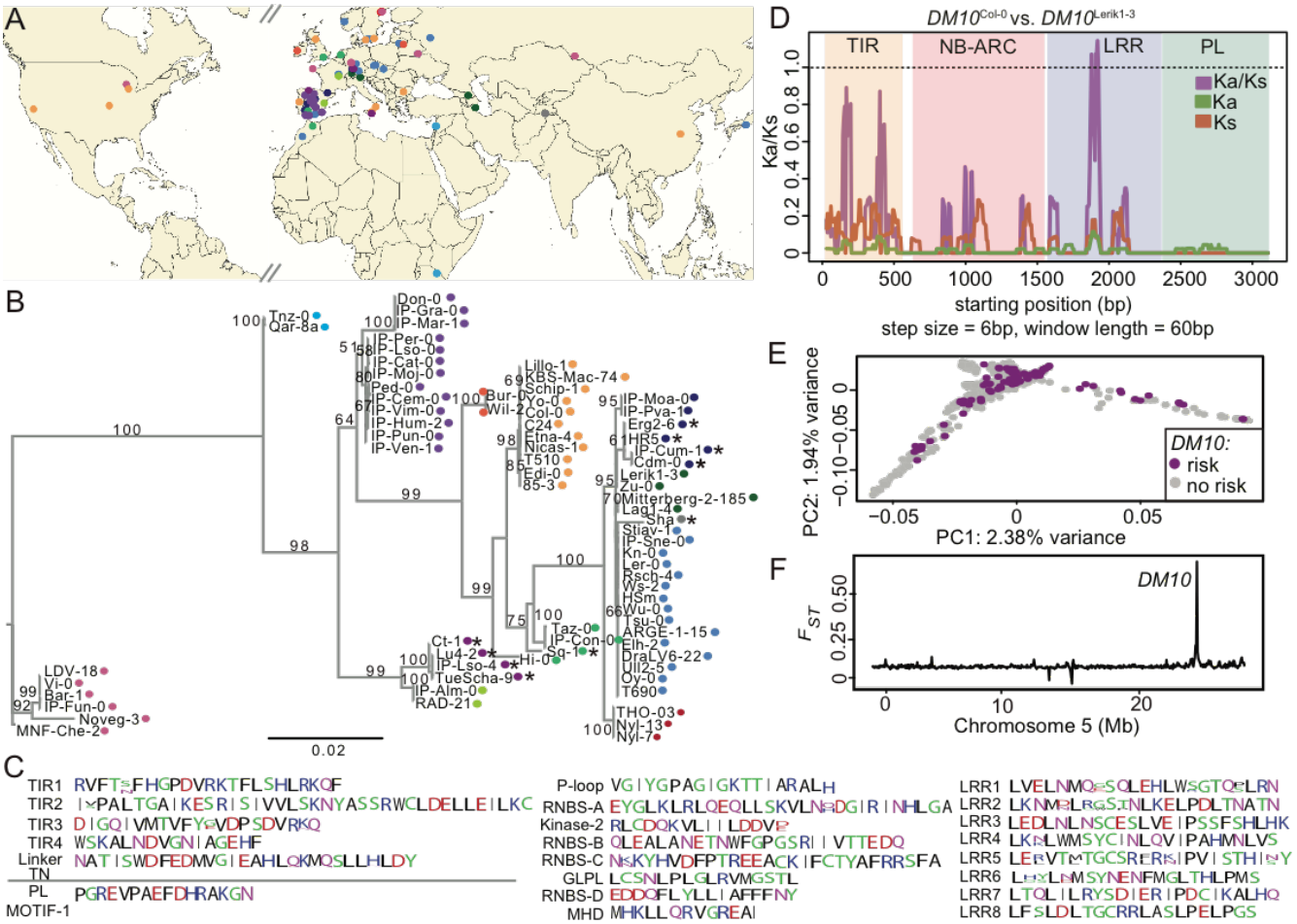


Fig S5. *DM10* natural variation. A. Geographic locations of 73 accessions carrying different *DM10* alleles. Each color indicates a similar *DM10* allele. **B.** ML tree of 73 CDS *DM10* sequences. 1,000 bootstrap replicates were performed, bootstrapping values are indicated on each branch, values above 50 are shown. Branch lengths in nucleotide substitutions are indicated. Asterisks indicate truncated DM10 proteins, colors as in A. **C.** DM10 motif consensus across 73 accessions. **D.** Ka/Ks ratio between $DM10^{Col\text{-}0}$ and $DM10^{Lerik1\text{-}3}$ **E.** Whole-genome PCA of 1001 Genomes accessions. **F.** F_{ST} between *DM10* risk and non-risk accessions across chromosome 5, only one peak is found in the region where *DM10* is located.

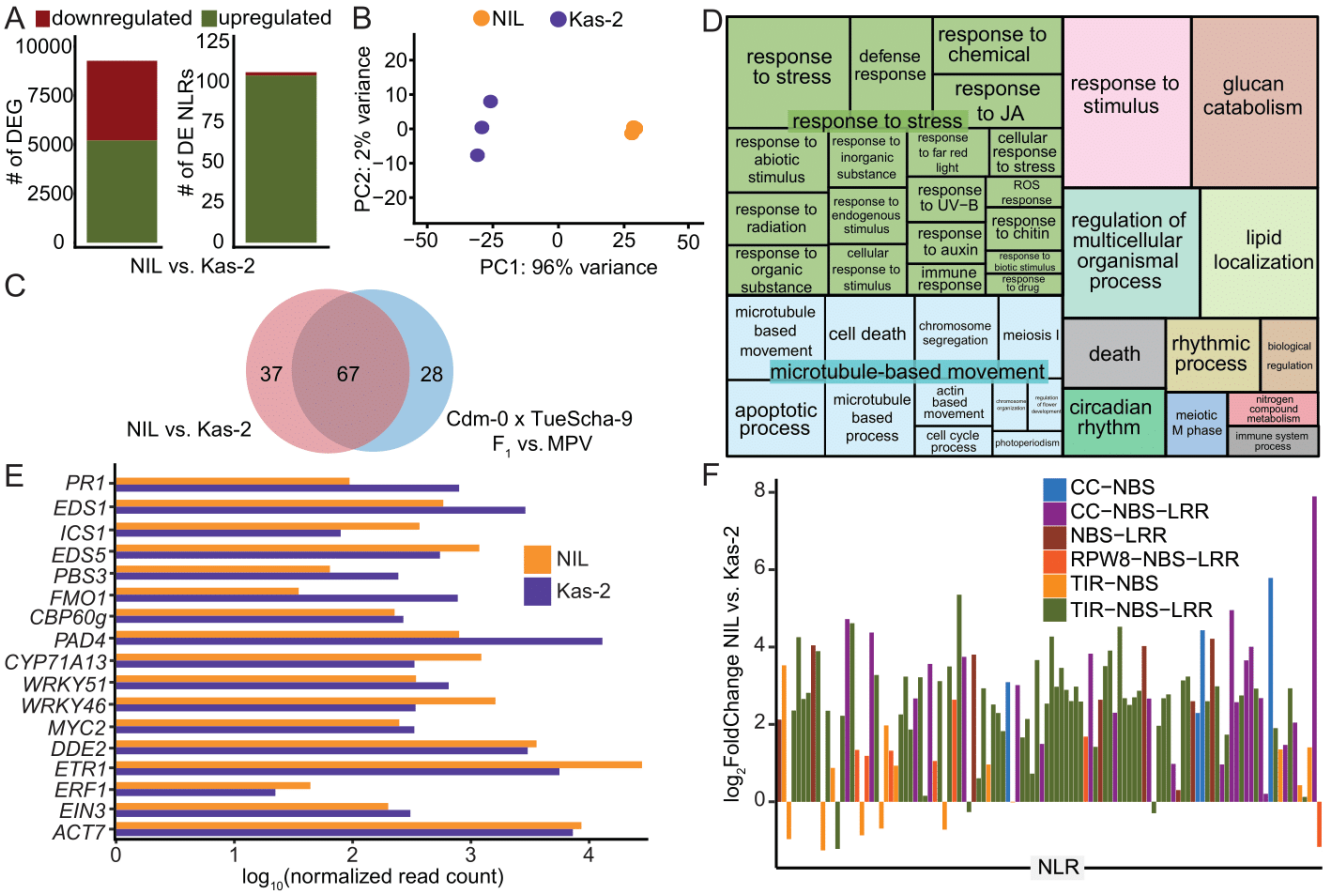


Fig S6. RNA-seq analysis of Ler/Kas-2 NIL vs. Kas-2 plants. A. Significantly ($|\log_2 FoldChange| > 1$, padj value < 0.01) differentially expressed genes (DEG) overall (left) and NLR genes (right) between Ler/Kas-2 near-isogenic line (NIL) and Kas-2 (Atanasov 2018). **B.** PCA of gene expression variance separating Ler/Kas-2 NIL and Kas-2. **C.** Intersection of differentially expressed NLRs between Ler/Kas-2 NIL and Kas-2 and between the Cdm-0 x TueScha-9 mid-parent value (MPV) and F_1 hybrid. **D.** REVIGO Gene Ontology treemap of the top 1000 DEG between Ler/Kas-2 NIL and Kas-2. Size of the square represents $-\log_{10}(p \text{ value})$ of each GO term (Table S2). **E.** $\log_{10}$(normalized read count) of defense-related marker genes between Ler/Kas-2 NIL and Kas-2 (Table S3). **F.** NLR expression changes between Ler/Kas-2 NIL and Kas-2 (Table S4). The NLR gene order follows Fig 1G.

Supplemental Methods

Pairwise genetic distances were calculated using the dist.DNA function in the ape R-package (v5.2) (Paradis and Schliep 2019). The identified NLR motif consensus across 73 DM10 proteins was visualized using WebLogo (Crooks et al. 2004).

Supplemental References

1001 Genomes Consortium. 2016. "1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana." *Cell* 166 (2): 481–91.

Atanasov, Kostadin E., Changxin Liu, Alexander Erban, Joachim Kopka, Jane E. Parker, and Rubén Alcázar. 2018. "NLR Mutations Suppressing Immune Hybrid Incompatibility and Their Effects on Disease Resistance." *Plant Physiology* 177 (3): 1152–69.

Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* 6:e4958.

Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res.* 14:1188–1190.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100.

Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–528.

A Case of Inbreeding Depression in a Natural *Arabidopsis arenosa* Population

A. Cristina Barragan[1], Maximilian Collenberg[1], Merjin Kerstens[1a], Rebecca Schwab[1], Felix Bemm[1,b], Ilja Bezrukov[1], Doubravka Požárová[2], Filip Kolář[2], Detlef Weigel[1]*

[1]Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany

[2]Department of Botany, Faculty of Science, Charles University, 128 01 Prague, Czech Republic

[a]Current address: Department of Molecular Biology, University of Wageningen, 6708, Wageningen, Netherlands

[b]Current address: KWS Saat, 37574 Einbeck, Germany

*Corresponding author: weigel@tue.mpg.de (D.W.)

Keywords: *A. arenosa*, natural populations, deleterious phenotypes, inbreeding depression

Abstract

Hybrid incompatibility in plants is usually the result of pairwise deleterious epistatic interactions between one or two loci, which often, but not always, encode for components of the immune system. In *A. thaliana*, the geographical co-occurence of incompatible alleles in natural settings has been shown. What remains elusive, is whether co-occurring incompatible alleles also exist in natural populations of outcrossing plant species, and if so, how common these are. To address this question, we screened over two thousand naturally occuring *A. arenosa* hybrid plants in search for potential incompatibilities. We show that while deleterious phenotypes are common and heritable in these plants, their molecular phenotype differs from that seen in incompatible *A. thaliana* hybrids. In addition, we identified a genomic region associated with one of these abnormal phenotypes through linkage mapping, and show that this region is highly homozygous in affected individuals, indicating that inbreeding depression rather than pairwise genetic incompatibilities may, at least in some cases, give rise to the deleterious phenotypes observed.

Introduction

Plant hybrid incompatibility cases that follow the Bateson-Dobzhansky-Muller (BDM) model have been mainly described in species or cultivated variants that are predominantly selfing (Krüger et al. 2002; Bomblies et al. 2007; Alcázar et al. 2009; Jeuken et al. 2009; Yamamoto et al. 2010; Chen et al. 2014; Chae et al. 2014; Todesco et al. 2014; Sicard et al. 2015; Deng et al. 2019; Sandstedt, Wu, and Sweigart 2020). In addition, alleles underlying incompatibility have been shown to naturally co-occur in *A. thaliana* individuals (Todesco et al. 2014). Whether these kinds of BDM incompatibilities exist between naturally co-occurring outcrossing plants, if so how common they are, remains unclear.

In selfing plants, slightly deleterious mutations tend to accumulate easier than in outcrossing plants, which are typically better at purging deleterious alleles (D. Charlesworth and Willis 2009). However, mutations that have strong negative effects on a plant would be under strong purifying selection in selfers, since these cannot be masked by heterozygosity (Arunkumar et al. 2015). Then again, many of the deleterious epistatic interactions between components of the immune system in incompatible *A. thaliana* hybrids are dominant or semi-dominant (although there are likely many more that have not yet been described which are recessive), and do not have to occur in a homozygous state to cause incompatibility in the first place (Chae et al. 2014; C. A. Barragan et al. 2019). Generally, natural selection is expected to eliminate genetic incompatibilities from populations where individuals are interbreeding, unless the advantage incompatible loci confer when present outweigh the disadvantages caused by their potential incompatibility. It is thus not clear how common deleterious epistatic interactions among naturally co-occurring outcrossing plants are expected to be.

A. arenosa is an obligate outcrossing relative of *A. thaliana* (Al-Shehbaz and O'Kane 2002; Koch and Matschinger 2007). These plants have either the ancestral diploid cytotype, or are tetraploid as a result of whole-genome duplication. The lack of populations of mixed ploidy indicates that a single ancient polyploidization event is likely (Arnold, Kim, and Bomblies 2015).

Here, we studied eight natural diploid *A. arenosa* populations from around the Carpathian Mountains, which is a center of genetic diversity of the species (Schmickl et al. 2012). We collected seeds and plant material from over 1,700 plants, sowed these seeds in the lab, and found that in these naturally occurring hybrid plants, here defined as the combination of any two parental genotypes, abnormal phenotypes are common and heritable. The transcriptional profile of two families segregating these abnormal phenotypes differ to that observed in incompatible *A. thaliana* hybrids. In addition, we identified a genomic region linked to the abnormal phenotype in one of these two sequenced families and observed that this region is highly homozygous, hinting at inbreeding depression, which is often

brought on by the expression of recessive deleterious alleles in a homozygous state (B. Charlesworth and Charlesworth 1999), as the underlying cause of this deleterious phenotype.

Results

Structure of Natural *Arabidopsis arenosa* Populations

Plant material and seeds were collected from eight different diploid *A. arenosa* populations in central and west Slovakia (**Fig 1A, S1A and Table S1**). For each plant we collected at least five, but usually hundreds of seeds. Seeds from a single mother plant are either siblings (if they share the same pollen donor) or half sibs (if they do not). We will refer to the immediate as well as all later-generation progeny from a single mother plant as families.

To study the population structure of the eight populations, we individually genotyped 345 of the 1,768 plants from which we had collected seeds by RAD-seq (Rowan et al. 2017). The resulting reads were mapped to a newly in-house generated *A. arenosa* reference genome (**Table S2**). A Principal Component Analysis (PCA) showed that, although variance between populations is explained by multiple small-effect components (**Fig S1B**), samples from the same population are genetically more similar to each other (**Fig 1B**). In addition, isolation by distance was observed, since the closer two populations are to each other, the more genetically similar they tend to be (**Fig 1C**, **S1C**). Genomic proportions that are identical by descent among individuals from the same population were relatively similar across all populations and ranged around from 96 to 97.5% (**Fig 1D**). The eight populations analyzed were estimated to have arisen from four or five ancestral populations (**Fig 1E, S1D**), with populations that are geographically closer tending to share a common ancestor e.g. the Horse and Bridge, and the Ranger (R) and Tesnina (T) populations (**Fig 1E**, **S1C**).

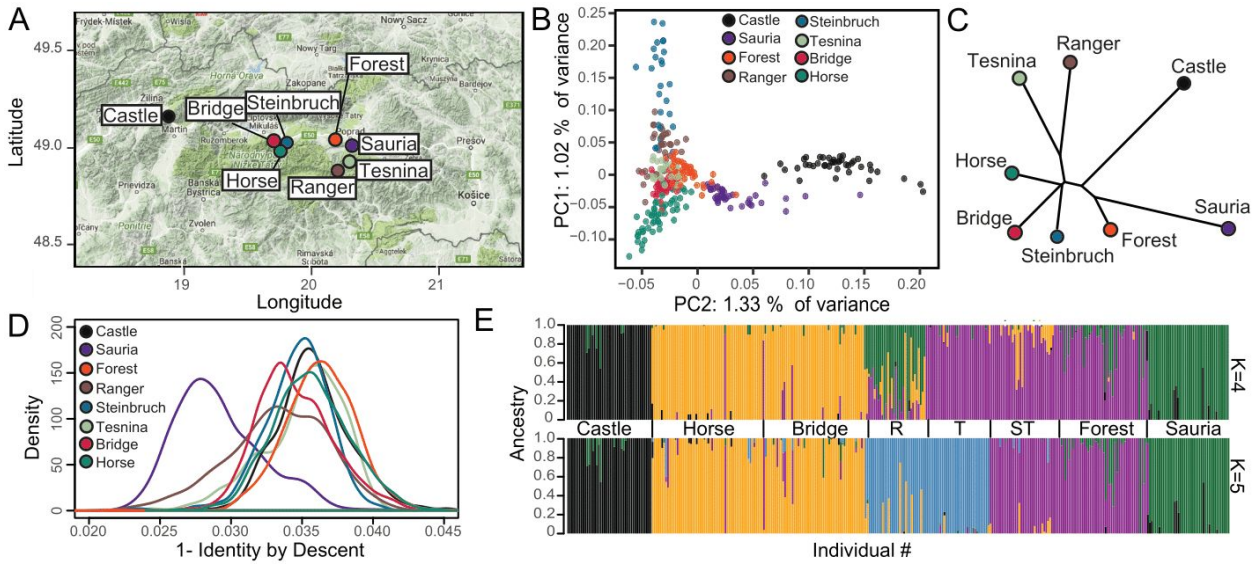


Fig 1. Structure of the sampled *A. arenosa* populations. A. Locations of the eight *A. arenosa* populations sampled. **B.** Genetic relationship between individual samples estimated by PCA. **C.** F_{ST}-based Neighbour-Joining Tree of the sampled populations. **D.** Identity by descent (1-IBS) among individuals from the same population. **E.** ADMIXTURE plot showing ancestry inferences (Alexander and Lange 2011). Four (K=4) or five (K=5) ancestral populations likely gave rise to the eight sampled populations. Ranger (R), Tesnina (T) and Steinbruch (ST).

Deleterious Phenotypes are Relatively Common and Heritable in *A. arenosa*

Because *A. arenosa* is an obligate outcrosser, all individuals are by definition hybrids. To study if hybrid incompatibilities naturally arise among co-occurring *A. arenosa* individuals, and if so, how prevalent these are, we sowed 4 to 6 seeds each from 461 families originating from all eight sampled populations. These plants, which we designate as F_1 individuals, were screened for abnormal phenotypes likely to reduce fitness in the field, such as small size, necrosis, chlorosis and developmental defects. If heritable, these could be caused by recessive or dominant deleterious alleles at a single locus, or by hybrid incompatibilities among two or more loci (**Fig 2A, Table S3**).

In 18% of the 461 families, at least one of the 4 to 6 plants screened showed some deleterious phenotype (**Fig 2B**), with the most common being chlorosis and reduced growth. The phenotypes ranged in severity from very mild and disappearing with age, to the plant not developing past the cotyledon stage and dying early on (**Fig 2C, D**). To test if these phenotypes were heritable, we created pseudo-F_2 populations for the 86 families with abnormal phenotypes. To this end, the 4 to 6 F_1 siblings were crossed to each other (sibcrosses) (**Fig 2A**). A total of 37 out of these 86 families tested produced abnormal offspring which often resembled the abnormal F_1 individuals from the same family, indicating that there is a genetic basis to many of the observed deleterious phenotypes. The fraction of

families with abnormal heritable phenotypes varied considerably between the eight geographic populations, from 1.7 to 16.6% (**Fig 2B, Table S3**).

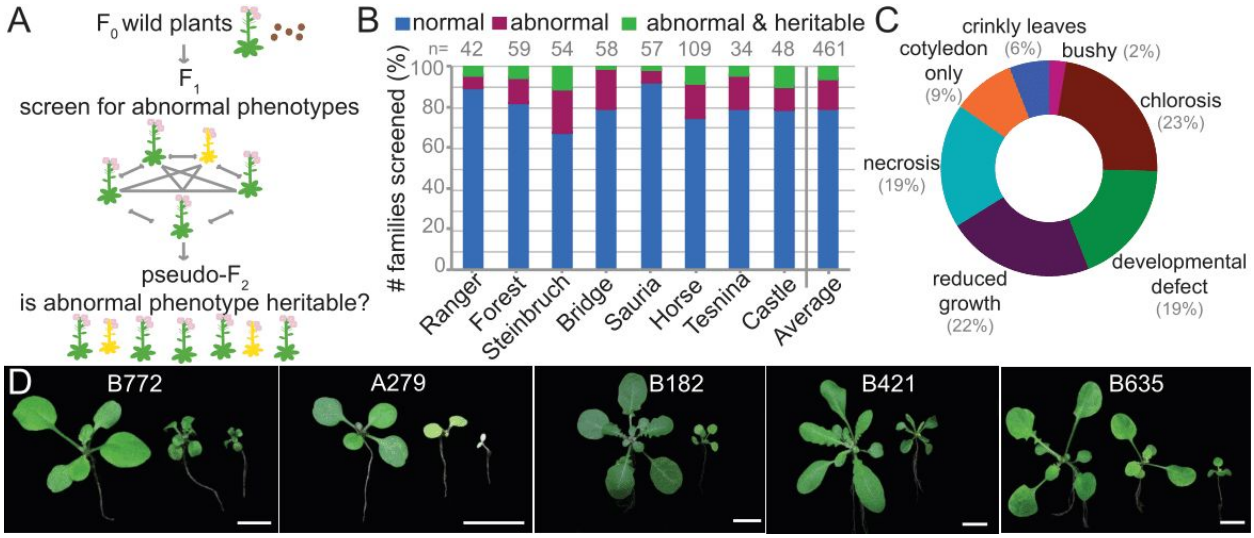


Fig 2. Screening for *A. arenosa* abnormal phenotypes. A. Creation of pseudo-F_2 population. F_1 seeds were collected from wild *A. arenosa* plants (F_0). The F_1 offspring were screened for abnormal phenotypes and, if present, these plants were crossed with their siblings to test whether this phenotype was recapitulated in the following generation (pseudo-F_2). **B.** Percentages of families showing abnormal phenotypes in the F_1 and pseudo-F_2 generation per geographic population (see Fig. 1A). The number of F_1 families screened per population is indicated at the top (n). **C.** Pie chart showing the most common deleterious phenotypes in F_1 and pseudo-F_2 plants. Some plants fell under more than one category. **D.** Examples of abnormal phenotypes from four independent families. Plants were between three and five weeks old. Scale bars represent 1 cm.

RNA-seq of Two *A. arenosa* Families Displaying Different Deleterious Phenotypes

To obtain first insight into molecular or physiological processes disturbed in the plants with deleterious abnormal phenotypes, we chose two families for RNA-seq analysis: A279 from the Sauria population and B772 from the Castle population (**Fig 2D**, **3A**). In both families, the pseudo-F_2 individuals consistently included plants with stunted growth and chlorosis in a milder or a more severe form (**Fig 2D, 3A**). Tissue was harvested from both normal and abnormal plants 17 days after germination, when the abnormal phenotype was clearly visible, but before plants carrying the more severe phenotype started dying (**Fig 3A**). A PCA showed that most of the variance in gene expression was driven by the difference between plants showing the abnormal phenotype and those that do not, this was especially seen in the A279 family, where the abnormal phenotype is more pronounced (**Fig 3B, C**). In abnormal plants of the A279 family, 8,962 genes out of 22,640 annotated genes were differentially expressed, whereas in B772 only 1,079 genes were misregulated (**Fig 3D**). There was substantial overlap between the misexpressed genes between the A279 and the B772 families, 532 genes (**Fig 3E**). A Gene Ontology

(GO) analysis of these overlapping differentially expressed genes (DEGs) showed that these were enriched for the terms "protein phosphorylation" and "response to stress" (**Fig 3F, Table S4**). We also assessed the top 100 and 500 DEGs for each family separately (**Fig S2A-D, Table S5**). Both sets were enriched for similar terms to those of their intersection, with the A279 DEGs being additionally enriched for "glucosinolate metabolic process" and "post-embryonic development" (**Fig S2A-D, Table S5**).

Previously, it has been shown that in several incompatible *A. thaliana* hybrids, defense response genes and genes involved in the salicylic acid (SA) pathway were greatly upregulated (Bomblies et al. 2007; A. C. Barragan et al. 2020). We therefore analyzed the expression of the *A. arenosa* orthologs of such genes, in addition to other marker genes (Papadopoulou et al. 2018) (**Table S6**). In the A279 family, very few of these genes were abnormally expressed, including *PBS3* orthologs, which were upregulated and are involved in disease resistance. However, most SA biosynthesis and signalling genes such as *PR1*, *EDS1*, *ICS1*, *EDS5* and *CBP60g*, many of which were induced in the *A. thaliana* hybrids, were not induced in the abnormal A279 plants. There was increased expression, however, of the stress response gene *PER57,* the SA-related transcription factor gene *WRKY28*, and of genes involved in jasmonic acid and ethylene biosynthesis or signalling such as *LOX2* and *DDE2* as well as *PDF1.2* and *EIN3*. It was notable that several genes coding for cytoskeleton components such as actin and tubulin were downregulated in the abnormal plants, possibly related to stunted growth.

In the B772 family, which had many fewer DEGs than A279, *EDS16*, *PR1* and *CBP60g* and *SARD1* were among the upregulated pathogenesis related genes, although to a much lesser extent than in incompatible *A. thaliana* hybrids (**Fig3G-H**).

Since the widespread induction of nucleotide binding site-leucine-rich repeat immune receptor genes (NLRs) was observed in incompatible *A. thaliana* hybrids (A. C. Barragan et al. 2020), the expression of *A. arenosa* NLR orthologs was compared between plants showing an abnormal and a normal phenotype. NLR induction was observed to a lesser extent than seen in *A. thaliana,* with genes from 7 out of 40 NLR orthogroups being upregulated in A279 and 3 of out 40 in BB72 family (**Fig S2E-F**, **Table S7**).

In short, both *A. arenosa* families showing deleterious abnormal phenotypes exhibit differential gene expression when compared to normal plants, especially the A279 family which shows a more pronounced abnormal phenotype. The overall transcriptional profile of these abnormal plants differs from that seen in incompatible *A. thaliana* hybrids, with limited evidence for autoimmunity.

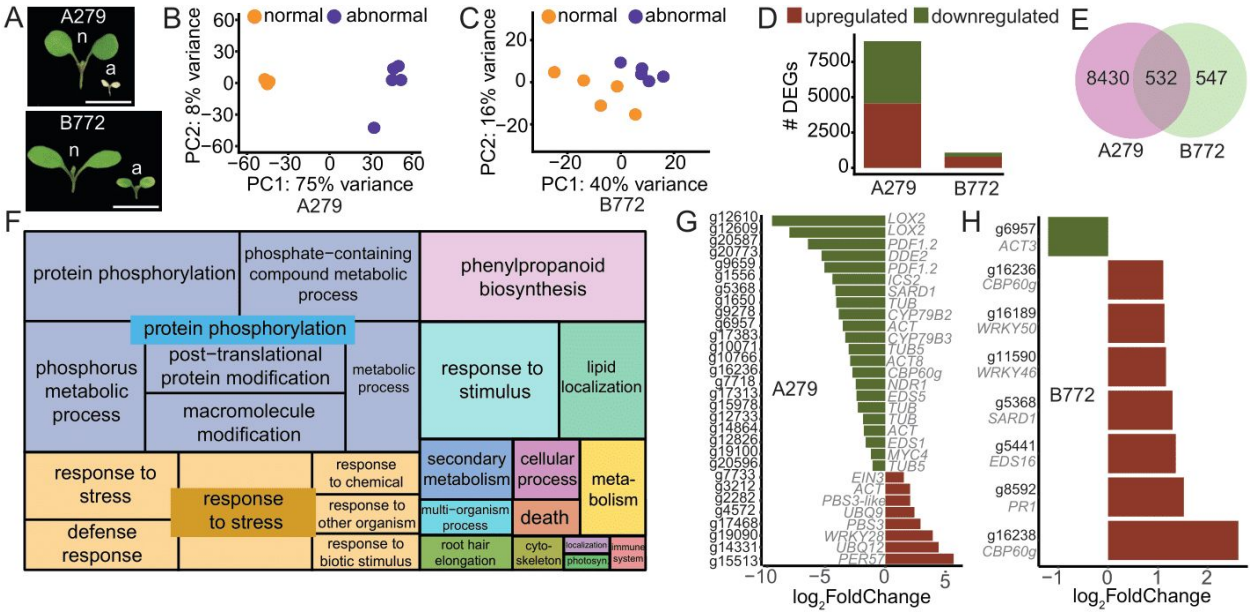


Fig 3. RNA-seq analysis of the A279 and B772 families. A. At 17 days after germination, normal (n) and abnormal (a) plants from the A279 and B772 families differ from each other, with abnormal plants showing reduced growth and chlorosis. Plants were grown at 16°C. Scale bar represents 0.5 cm. **B-C.** PCA of gene expression values. The main variance is between normal and abnormal plants. Each dot indicates one biological replicate, with five per family. **D.** Number differentially expressed genes (DEGs) which are either up- or downregulated. **E.** Intersection of DEGs between the A279 and B772 families. **F.** REVIGO Gene Ontology treemap of the DEGs in the intersection between the two families. Size of the square represents -$\log_{10}$(*p* value) of each GO term. **G-F.** -$\log_2$FoldChange of significantly (|$\log_2$FoldChange| >1, padj value < 0.05) changed marker genes between normal and abnormal plants. *A. thaliana* orthologs in grey.

QTL Mapping of Deleterious *A. arenosa* Phenotypes

To identify the genetic basis of the deleterious phenotypes observed, both normal and abnormal plants from four independent pseudo-F_2 families (B772, A279, B635 and B182) were individually genotyped by RAD-seq (Rowan et al. 2017) (**Fig 2A**). Quantitative trait locus (QTL) analysis did not reveal a clear genomic region associated with the abnormal phenotype in three out of the four families studied (**Fig S3A-D**).

In the B772 family, we identified a very clear QTL on chromosome six. This QTL interval, between 5.30-5.36 Mb, and with a maximum peak at 5.34 Mb and a LOD score of 32.9, contained 12 annotated genes (**Fig 4A, B**; **Table S8**). There was little recombination especially in the first half of the QTL interval in the B772 family (**Fig 4C**). These 12 genes were not obviously differentially expressed between normal and abnormal plants (**Table S8**). In a larger genomic region spanning 3.5-8.0 Mb on chromosome 6, four marginally significant QTL peaks were found (**Fig 4B**). This larger region contained 927 genes (**Table S9**), 49 of which were differentially expressed, but were not enriched for any

particular GO term (**Table S10**). There was one NLR gene, g15519, in this 4.5 Mb interval, with 80% amino acid identity of the encoded protein to the TIR-NLR CSA1 (Faigón-Soverna et al. 2006).

To examine the genomic region associated with the deleterious phenotype seen in the B772 family more closely, 52 pseudo-F_2 individuals (24 normal and 28 abnormal) were whole-genome sequenced. The fixation index (F_{ST}) between the plants showing an abnormal phenotype and those that did not was calculated. Concordant with the QTL analysis, a peak in the first half chromosome 6 was observed, with a maximal F_{ST} value of 0.41 between 5.7-5.8 Mb, followed by 5.3-5.4 Mb with a maximal F_{ST} value of 0.36. These two regions were part of the same linkage block (**Fig 4E)**. No other region with elevated F_{ST} values were observed across the entire genome (**Fig S3E)**. Taken together, this indicates that a region in the first half of chromosome six is both genetically differentiated in abnormal plants and linked to their phenotype.

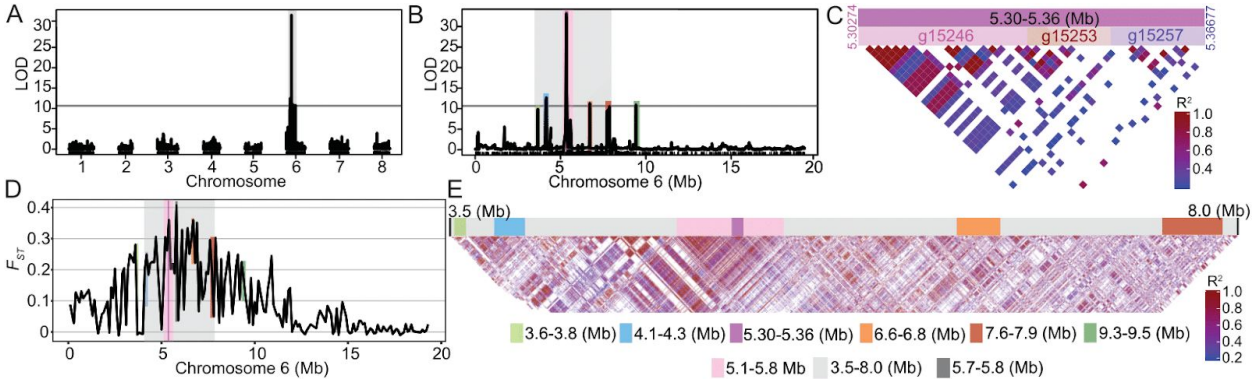


Fig 4. QTL analysis from the B772 family. A-B. A QTL peak is found on chromosome 6 (5.30-5.36 Mb). The horizontal lines indicate 0.05 significance threshold established with 1,000 permutations. **C.** Linkage disequilibrium (LD) across this same 60 kb region in chromosome 6. Strong linkage is observed in genes found between g15246 and g15253 and to a lesser extent to those found until g15257. **D.** Fixation index (F_{ST}) between normal and abnormal plants across chromosome six, the 3.5-8.0 Mb region is highlighted in grey. Colored lines indicate the position of the QTL peaks in B. **E.** LD plot from 3.5-8.0 Mb in chromosome 6. Colors same as in B and D. The region with the highest LOD and F_{ST} (5.1-5.8 Mb) are under LD (pink).

The Deleterious Phenotype in the B772 Family is Likely due to Inbreeding Depression

The fact that we found a single strong QTL peak for the abnormal phenotype in the B772 family did not directly speak to whether this due to a deleterious recessive mutation being exposed or to incompatibility between alleles at the same locus. The latter has been reported for two cases of hybrid incompatibility in *A. thaliana* (Chae et al. 2014; Todesco et al. 2014).

To distinguish between these possibilities, we extracted the genotype calls for the region surrounding the QTL interval on chromosome 6 (3.5-8.0 Mb), and compared the inbreeding coefficient (*F*) between normal and abnormal individuals. All abnormal individuals had a higher inbreeding coefficient than the

normal ones, with an average *F* of 0.93 (**Fig 5A, Table S11**). As a control, a similarly sized genomic region on chromosome 1 was analyzed; no differences were seen (**Fig S5A, Table S11**).

In addition to the inbreeding coefficient, we screened for runs of homozygosity (ROH) as an indication of inbreeding depression as the underlying cause for the abnormal B772 phenotype. ROH were identified in the 3.5-8.0 Mb region of chromosome 6 in all abnormal plants (**Fig 5B, Table S12**). This was not the case for other regions in the genome, such as 3.5-8.0 Mb of chromosome 1 (**Fig S4B, Table S12**).

Finally, to confirm that the abnormal individuals shared similar sequences in this extended region, a Neighbor-Joining Tree was calculated. Sequences from abnormal individuals clustered together for this region (**Fig 5C**), but not for the control region on chromosome 1.

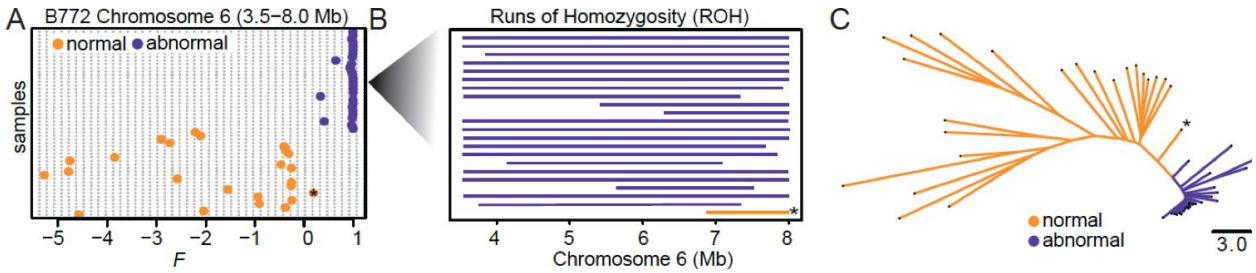


Fig 5. Homozygosity in normal and abnormal plants. A. Inbreeding coefficient (*F*) of plants showing an abnormal and a normal phenotype calculated for the 3.5-8 Mb region on chromosome 6. Abnormal plants have a much higher average *F*. **B.** Runs of homozygosity (ROH) from the same genomic region. Only plants with an abnormal phenotype show ROH with one exception of a shorter ROH in a normal plant, which is indicated by an asterisk in A-C. **C.** Neighbor-Joining Tree of the 3.5-8.0 Mb region on chromosome 6. Individuals cluster by phenotype. Branch lengths in nucleotide substitutions are indicated.

Discussion

Inbreeding depression is a result of higher levels of homozygosity in an individual, which can either make deleterious recessive mutations visible, or reduce the advantage certain alleles confer when present in a heterozygous state, known as overdominance (D. Charlesworth and Willis 2009). Many small-effect mutations across multiple loci that are maintained in natural populations at low frequencies are believed to be a common source giving rise to inbreeding depression (D. Charlesworth and Willis 2009). In our study, the fact that in three out of four tested families segregating an abnormal phenotype showed no clear region of the genome linked to this phenotype though QTL analysis, may indicate the involvement of multiple genomic regions in giving rise to these abnormal phenotypes.

The transcriptional profile in the two abnormal *A. arenosa* cases studied showed little indication for an ongoing autoimmune response. Autoimmunity due to hybrid necrosis in *A. thaliana* and other plants can

often be partially or completely suppressed by growing plants at elevated temperatures (Bomblies and Weigel 2007). The abnormal phenotypes in the A279 and the B772 family, were however not ameliorated by growing the plants at a higher temperature (**Fig S5A-B**).

Hybrid incompatibilities in *A. arenosa* which are caused by pairwise deleterious interactions following the BDM model, may be more likely to occur in plants originating from different populations, since different local selection pressures are guiding differential diversification and maintenance of immune genes, and where mismatches among these loci often give rise to incompatibilities in hybrids. In crosses between individuals from different populations, abnormal phenotypes seem to be relatively common as well (**Fig S5C-D, Table S13**). Future QTL analyses on these inter-populations crosses will help establish whether BDM incompatibilities may underlie these abnormal phenotypes among unrelated plants.

By occurring in a diploid and tetraploid state, *A. arenosa* is a great system to study the role ploidy may play in giving rise to both hybrid incompatibility and inbreeding depression. In the case of hybrid incompatibilities, a higher number of genes and gene copies may provide a larger basis for sequence diversification, which is positively correlated with genetic incompatibilities. On the other hand, the fitness cost many immune system components impart (D. Tian et al. 2003; Karasov et al. 2014), may limit their diversification, irrespective of its ploidy. In the case of inbreeding depression, polyploidy may increase the frequency of heterozygosity and mask recessive deleterious alleles, making inbreeding depression less frequent.

The possibility that pairwise genetic incompatibilities among interbreeding *A. arenosa* individuals occur is not excluded. In *A. thaliana*, co-occurring incompatible alleles seem to be more the exception rather than the rule. Further investigation of loci involved in potential *A. arenosa* incompatibilities through linkage mapping in addition to scanning the genome for the presence of long distance Linkage Disequilibrium to identify potential regions that do not co-occur as a sign of potential incompatibility, is warranted. Genomic areas under positive selection have been identified by scanning for loci that show an excess proportion of being identical by descent, among individuals in a population (Albrechtsen, Moltke, and Nielsen 2010; Han and Abney 2011). The same principle could be used to scan for regions under negative selection which could be candidates for loci giving rise to incompatibilities or acting as sources of inbreeding depression.

Methods

Plant material

All plants were stratified in the dark at 4°C for 5-8 days prior to planting on soil. Plants were grown in long days (16 h of light) at 16°C or 23°C at 65% relative humidity under 110 to 140 µmol m^{-2} s^{-1} light provided by Philips GreenPower TLED modules (Philips Lighting GmbH, Hamburg, Germany). The ploidy of each population sampled was estimated via flow-cytometry (one representative plant per population) and with nQuire (Weiß et al. 2018). Plants were collected under permit number (XXXX).

De novo genome assembly and annotation

An *A. arenosa* plant (701.a) from the Castle population (**Table S1**) was grown as described above. To reduce starch accumulation, 3-week-old plants were put into darkness for 30 h before harvesting. Sixteen grams of flash frozen leaf tissue were ground in liquid nitrogen and nuclei isolation was performed according to (Workman et al. 2018). High-molecular-weight DNA was recovered with the Nanobind Plant Nuclei Kit (Circulomics; SKU NB-900-801-01). A 35-kb template library was prepared with the SMRTbell® Express Template Preparation Kit 2.0, and size-selected with the BluePippin system according to the manufacturer's instructions (P/N 101-693-800-01, Pacific Biosciences, California, USA). In addition, a PCR-free library was prepared with the NxSeq® AmpFREE Low DNA Library Kit from Lucigen® according to the manufacturer's instructions. The final library was sequenced on a Pacific Biosciences Sequel instrument with Binding Kit 3.0. PacBio long-reads were assembled with Falcon (VXX) (Chin et al. 2016). The resulting contigs were first polished using the long-reads with the Arrow algorithm (v2.3.2; https://github.com/PacificBiosciences/GenomicConsensus), followed by a second polishing step with PCR-free short-reads using the Pilon algorithm (v1.22) (Walker et al. 2014). Lastly, the resulting contigs were scaffolded based on the *A. lyrata* assembly (v1) by REVEAL (v0.2.1) (Linthorst et al. 2015). The previously generated *A. arenosa* transcriptome sequencing data were mapped against the scaffolded genome assembly using HISAT (v2.0.5) (Kim, Langmead, and Salzberg 2015). Subsequently, the mapping results were used as extrinsic RNA sequencing evidence when annotating the genome using AUGUSTUS (v3.3.3) (Stanke et al. 2006). Transposable elements and repetitive regions were identified using RepeatModeler2 (v2.01) (Flynn et al. 2020). Repeat and transposable element masking was performed using RepeatMasker (v4.4.0) (A.F.A. Smit, R. Hubley & P. Green RepeatMasker at http://repeatmasker.org). Orthologous genes shared between *A. arenosa* and the current *A. thaliana* reference annotation from Araport11 were identified using Orthofinder (v2.4.0) (Emms and Kelly 2019). *A. arenosa* and *A. thaliana* protein fasta files were subsetted to only keep the primary transcript for orthologous assignment using the AGAT toolkit (v0.4.0) (https://github.com/NBISweden/AGAT).

Genotyping-by-sequencing and QTL mapping

Genomic DNA was from extracted from either F_0 individuals collected from the wild, or pseudo-F_2 plants with CTAB (cetyl trimethyl ammonium bromide) buffer (Doyle and Doyle 1987) and then purified through chloroform extraction and isopropanol precipitation (Ashktorab and Cohen 1992). Genotyping-by-Sequencing (GBS) using RAD-seq was used to genotype individuals with KpnI tags (Rowan et al. 2017). Briefly, libraries were single-end sequenced on a HiSeq 3000 instrument with 150 bp reads. Reads were processed with Stacks (v1.35)(Catchen et al. 2013) and mapped to our in-house *A. arenosa* reference (**Table S2**) with bwa-mem (v0.7.15)(H. Li 2013), variant calling was performed with GATK (v3.5)(McKenna et al. 2010). SNP filtering was performed with VCFtools (v0.1.14) (Danecek et al. 2011). Basic filtering criteria were to retain bi-allelic SNPs only, SNPs with at most 30% missing data, individuals with less than 40% missing data and SNPs with a minimum allele frequency of 0.01. Pseudo-F_2 plants were used as mapping populations for QTL analysis, which was performed using R/qtl (Broman et al. 2003). Analyses based on 345 individuals and 2,205 markers (F_0 individuals), and 227 individuals and 11,858 markers (B772), 162 individuals and 9,064 markers (B182), 183 individuals and 14,672 markers (A279) and 271 individuals and 6,110 markers (B635).

RNA sequencing

Five biological replicates of 21 day-old (17 days after germination) shoots of B772 and A279 plants were collected. RNA extraction was done as described in (Yaffe et al. 2012). Sequencing libraries were prepared using the TruSeq Total RNA Kit (illumina) and the Ribo-Zero Plant Kit (Illumina). Libraries were paired-end sequenced (150bp) in an Illumina HiSeq3000 (Illumina Inc., San Diego, USA) instrument. Reads were mapped against the in-house *A. arenosa* reference (**Table S2**) using bowtie2 (v2.2.6)(Langmead and Salzberg 2012). Default parameters were chosen unless mentioned otherwise. Transcript abundance was calculated with RSEM (v1.2.31) (B. Li and Dewey 2011). Differential gene expression analyses were performed using DESeq2 (v1.18.1) (Love, Huber, and Anders 2014). Genes with less than ten counts over all samples were removed from downstream analyses. Significant changes in gene expression between two genotypes were determined by filtering for genes with a $|\log_2\text{FoldChange}| > 1$ and padj value < 0.05. Plots were generated using the R package ggplot2 (v3.2.0) (Wickham 2009) and heatmaps were plotted using pheatmap (v1.0.8) (Kolde 2012). Gene Ontology (GO) analyses were performed using AgriGO (T. Tian et al. 2017) using the SEA method. The GO results were visualized with REVIGO treemap (Supek et al. 2011), for clearer visualization a maximum of

15 and GO categories with the lowest *p* values were plotted. The complete list of GO terms is found in Table S5 and S6.

Whole-genome sequencing

Libraries from 52 individuals (23 normal, 29 abnormal) from the B772 family were prepared using a modified protocol of (Picelli et al. 2014). Libraries were paired-end sequenced (150bp) in an Illumina HiSeq3000 (Illumina Inc., San Diego, USA) instrument. Reads were processed with Stacks (v1.35)(Catchen et al. 2013) and mapped to our in-house *A. arenosa* reference genome (**Table S2**) with bwa-mem (v0.7.15)(H. Li 2013), variant calling was performed with GATK (v3.5) (McKenna et al. 2010). Variant filtering criteria was the same as in RAD-seq. 43,885 SNPs were obtained after filtering, in the 3.5-8.0 Mb interval in chromosome 6 4,230 SNPs were found. In chromosome 1 this same interval contained 1,617 SNPs.

Population genetic analyses

Principal component analyses were calculated with smartPCA (Patterson, Price, and Reich 2006). F_{ST} was calculated with VCFtools (100 kb windows). Maps were created with the R-packages maps (v3.3) and ggmap (v3.0)(Kahle and Wickham 2013). Ancestral populations were estimated using ADMIXTURE (Alexander and Lange 2011). Identity by descent (1-IBD), linkage disequilibrium (r^2) (ld-window = 5kb), inbreeding coefficient (*F*) and runs of homozygosity (ROH) were calculated with PLINK (v1.90) (Purcell et al. 2007). For inbreeding analyses SNPs were pruned (window size=50kb, step size=5 variants, r^2 threshold=0.5) with PLINK. Sequences were visualized and aligned with Aliview (Larsson 2014) and Neighbor-Joining trees were estimated with Jalview (Clamp et al. 2004) and visualized in Figtree (v1.4.3) (Rambaut 2012).

Data Availability

Sequencing data can be found at the European Nucleotide Archive (ENA) under project numbers XXX (RNA-seq experiment) and XXX (*A. arenosa* assembly).

Funding

This work was supported by the Deutsche Forschungsgemeinschaft through the Collaborative Research Center (CRC1101), the Max Planck Society (to D.W.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

Acknowledgements

We thank Hung Vo, Frank Vogt, Josip Perkovic and Christa Lanz for technical support and Wei Yuan for guidance with the RNA extractions. We also thank Patrick Hüther and Claude Becker for helping with the sample transportation logistics and Gautam Shirsekar and Christian Kubica for discussion.

Author contributions

Conceptualization: ACB, FK, DW.

Formal analysis: ACB, MC, MK, FB.

Funding acquisition: DW.

Investigation: ACB, MC, RS, MK, IB, DP, FK.

Methodology: ACB.

Project administration: DW.

Supervision: DW.

Writing – original draft: ACB.

Writing – review & editing: ACB, DW.

Competing interests. The authors have declared that no competing interests exist.

References

Agorio, Astrid, Stéphanie Durand, Elisa Fiume, Cécile Brousse, Isabelle Gy, Matthieu Simon, Sarit Anava, et al. 2017. "An Arabidopsis Natural Epiallele Maintained by a Feed-Forward Silencing Loop between Histone and DNA." *PLoS Genetics* 13 (1): e1006551.

Albrechtsen, Anders, Ida Moltke, and Rasmus Nielsen. 2010. "Natural Selection and the Distribution of Identity-by-Descent in the Human Genome." *Genetics* 186 (1): 295–308.

Alcázar, Rubén, Ana V. García, Jane E. Parker, and Matthieu Reymond. 2009. "Incremental Steps toward Incompatibility Revealed by Arabidopsis Epistatic Interactions Modulating Salicylic Acid Pathway Activation." *Proceedings of the National Academy of Sciences* 106 (1): 334–39.

Alexander, David H., and Kenneth Lange. 2011. "Enhancements to the ADMIXTURE Algorithm for Individual Ancestry Estimation." *BMC Bioinformatics* 12 (June): 246.

Alhajturki, Dema, Subhashini Muralidharan, Markus Nurmi, Beth A. Rowan, John E. Lunn, Helena Boldt, Mohamed A. Salem, et al. 2018. "Dose-Dependent Interactions between Two Loci Trigger Altered Shoot Growth in BG-5 × Krotzenburg-0 (Kro-0) Hybrids of Arabidopsis Thaliana." *The New Phytologist* 217 (1): 392–406.

Al-Shehbaz, Ihsan A., and Steve L. O'Kane Jr. 2002. "Taxonomy and Phylogeny of Arabidopsis (brassicaceae)." *The Arabidopsis Book / American Society of Plant Biologists* 1 (September): e0001.

Arnold, Brian, Sang-Tae Kim, and Kirsten Bomblies. 2015. "Single Geographic Origin of a Widespread

Autotetraploid Arabidopsis Arenosa Lineage Followed by Interploidy Admixture." *Molecular Biology and Evolution* 32 (6): 1382–95.

Arunkumar, Ramesh, Rob W. Ness, Stephen I. Wright, and Spencer C. H. Barrett. 2015. "The Evolution of Selfing Is Accompanied by Reduced Efficacy of Selection and Purging of Deleterious Mutations." *Genetics* 199 (3): 817–29.

Ashktorab, H., and R. J. Cohen. 1992. "Facile Isolation of Genomic DNA from Filamentous Fungi." *BioTechniques* 13 (2): 198–200.

Barragan, A. Cristina, Maximilian Collenberg, Jinge Wang, Rachelle R. Q. Lee, Wei Yuan Cher, Fernando A. Rabanal, Haim Ashkenazy, Detlef Weigel, and Eunyoung Chae. 2020. "A Truncated Singleton NLR Causes Hybrid Necrosis in Arabidopsis Thaliana." *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msaa245.

Barragan, Cristina A., Rui Wu, Sang-Tae Kim, Wanyan Xi, Anette Habring, Jörg Hagmann, Anna-Lena Van de Weyer, et al. 2019. "RPW8/HR Repeats Control NLR Activation in Arabidopsis Thaliana." *PLoS Genetics* 15 (7): e1008313.

Bikard, David, Dhaval Patel, Claire Le Metté, Veronica Giorgi, Christine Camilleri, Malcolm J. Bennett, and Olivier Loudet. 2009. "Divergent Evolution of Duplicate Genes Leads to Genetic Incompatibilities within A. Thaliana." *Science* 323 (5914): 623–26.

Bomblies, Kirsten, Janne Lempe, Petra Epple, Norman Warthmann, Christa Lanz, Jeffery L. Dangl, and Detlef Weigel. 2007. "Autoimmune Response as a Mechanism for a Dobzhansky-Muller-Type Incompatibility Syndrome in Plants." *PLoS Biology* 5 (9): e236.

Bomblies, Kirsten, and Detlef Weigel. 2007. "Hybrid Necrosis: Autoimmunity as a Potential Gene-Flow Barrier in Plant Species." *Nature Reviews. Genetics* 8 (5): 382–93.

Broman, Karl W., Hao Wu, Saunak Sen, and Gary A. Churchill. 2003. "R/qtl: QTL Mapping in Experimental Crosses." *Bioinformatics* 19 (7): 889–90.

Catchen, Julian, Paul A. Hohenlohe, Susan Bassham, Angel Amores, and William A. Cresko. 2013. "Stacks: An Analysis Tool Set for Population Genomics." *Molecular Ecology* 22 (11): 3124–40.

Chae, Eunyoung, Kirsten Bomblies, Sang-Tae Kim, Darya Karelina, Maricris Zaidem, Stephan Ossowski, Carmen Martín-Pizarro, et al. 2014. "Species-Wide Genetic Incompatibility Analysis Identifies Immune Genes as Hot Spots of Deleterious Epistasis." *Cell* 159 (6): 1341–51.

Charlesworth, Brian, and Deborah Charlesworth. 1999. "The Genetic Basis of Inbreeding Depression." *Genetical Research*. https://doi.org/10.1017/s0016672399004152.

Charlesworth, Deborah, and John H. Willis. 2009. "The Genetics of Inbreeding Depression." *Nature Reviews. Genetics* 10 (11): 783–96.

Chen, Chen, Hao Chen, You-Shun Lin, Jin-Bo Shen, Jun-Xiang Shan, Peng Qi, Min Shi, et al. 2014. "A Two-Locus Interaction Causes Interspecific Hybrid Weakness in Rice." *Nature Communications* 5: 3357.

Chin, Chen-Shan, Paul Peluso, Fritz J. Sedlazeck, Maria Nattestad, Gregory T. Concepcion, Alicia Clum, Christopher Dunn, et al. 2016. "Phased Diploid Genome Assembly with Single-Molecule Real-Time Sequencing." *Nature Methods* 13 (12): 1050–54.

Clamp, Michele, James Cuff, Stephen M. Searle, and Geoffrey J. Barton. 2004. "The Jalview Java Alignment Editor." *Bioinformatics* 20 (3): 426–27.

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, et al. 2011. "The Variant Call Format and VCFtools." *Bioinformatics* 27 (15): 2156–58.

Deng, Jieqiong, Lei Fang, Xiefei Zhu, Baoliang Zhou, and Tianzhen Zhang. 2019. "A CC-NBS-LRR Gene Induces Hybrid Lethality in Cotton." *Journal of Experimental Botany* 70 (19): 5145–56.

Doyle, Jeff J., and Jan L. Doyle. 1987. "A Rapid DNA Isolation Procedure for Small Quantities of Fresh

Leaf Tissue." https://worldveg.tind.io/record/33886/.
Durand, S., N. Bouche, E. Perez Strand, O. Loudet, and C. Camilleri. 2012. "Rapid Establishment of Genetic Incompatibility through Natural Epigenetic Variation." *Current Biology: CB* 22 (4): 326–31.
Faigón-Soverna, Ana, Franklin G. Harmon, Leonardo Storani, Elizabeth Karayekov, Roberto J. Staneloni, Walter Gassmann, Paloma Más, Jorge J. Casal, Steve A. Kay, and Marcelo J. Yanovsky. 2006. "A Constitutive Shade-Avoidance Mutant Implicates TIR-NBS-LRR Proteins in Arabidopsis Photomorphogenic Development." *The Plant Cell* 18 (11): 2919–28.
Fishman, Lila, and Arpiar Saunders. 2008. "Centromere-Associated Female Meiotic Drive Entails Male Fitness Costs in Monkeyflowers." *Science* 322 (5907): 1559–62.
Han, Lide, and Mark Abney. 2011. "Identity by Descent Estimation with Dense Genome-Wide Genotype Data." *Genetic Epidemiology* 35 (6): 557–67.
Jeuken, Marieke J. W., Ningwen W. Zhang, Leah K. McHale, Koen Pelgrom, Erik den Boer, Pim Lindhout, Richard W. Michelmore, Richard G. F. Visser, and Rients E. Niks. 2009. "Rin4 Causes Hybrid Necrosis and Race-Specific Resistance in an Interspecific Lettuce Hybrid." *The Plant Cell* 21 (10): 3368–78.
Jiao, Wen-Biao, Vipul Patel, Jonas Klasen, Fang Liu, Petra Pecinkova, Marina Ferrand, Isabelle Gy, et al. 2020. "The Evolutionary Dynamics of Genetic Incompatibilities Introduced by Duplicated Genes in Arabidopsis Thaliana." https://doi.org/10.1101/2020.09.21.306035.
Kahle, David, and Hadley Wickham. 2013. "Ggmap: Spatial Visualization with ggplot2." *The R Journal* 5 (1): 144–61.
Karasov, Talia L., Joel M. Kniskern, Liping Gao, Brody J. DeYoung, Jing Ding, Ullrich Dubiella, Ruben O. Lastra, et al. 2014. "The Long-Term Maintenance of a Resistance Polymorphism through Diffuse Interactions." *Nature* 512 (7515): 436–40.
Koch, Marcus A., and Michaela Matschinger. 2007. "Evolution and Genetic Differentiation among Relatives of Arabidopsis Thaliana." *Proceedings of the National Academy of Sciences of the United States of America* 104 (15): 6272–77.
Kolde, Raivo. 2012. "Pheatmap: Pretty Heatmaps." *R Package Version* 61: 617.
Krüger, Julia, Colwyn M. Thomas, Catherine Golstein, Mark S. Dixon, Matthew Smoker, Saijun Tang, Lonneke Mulder, and Jonathan D. G. Jones. 2002. "A Tomato Cysteine Protease Required for Cf-2-Dependent Disease Resistance and Suppression of Autonecrosis." *Science* 296 (5568): 744–47.
Langmead, Ben, and Steven L. Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59.
Larsson, Anders. 2014. "AliView: A Fast and Lightweight Alignment Viewer and Editor for Large Datasets." *Bioinformatics* 30 (22): 3276–78.
Li, Bo, and Colin N. Dewey. 2011. "RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome." *BMC Bioinformatics* 12 (August): 323.
Li, Heng. 2013. "Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM." *arXiv [q-bio.GN]*. arXiv. http://arxiv.org/abs/1303.3997.
Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12): 550.
McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, et al. 2010. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data." *Genome Research* 20 (9): 1297–1303.
Moyle, Leonie C., Cathleen P. Jewell, and Jamie L. Kostyun. 2014. "Fertile Approaches to Dissecting Mechanisms of Premating and Postmating Prezygotic Reproductive Isolation." *Current Opinion in Plant Biology* 18 (April): 16–23.
Papadopoulou, Galini V., Anne Maedicke, Katharina Grosser, Nicole M. van Dam, and Ainhoa

Martínez-Medina. 2018. "Defence Signalling Marker Gene Responses to Hormonal Elicitation Differ between Roots and Shoots." *AoB Plants* 10 (3): ly031.

Patterson, Nick, Alkes L. Price, and David Reich. 2006. "Population Structure and Eigenanalysis." *PLoS Genetics* 2 (12): e190.

Picelli, Simone, Asa K. Björklund, Björn Reinius, Sven Sagasser, Gösta Winberg, and Rickard Sandberg. 2014. "Tn5 Transposase and Tagmentation Procedures for Massively Scaled Sequencing Projects." *Genome Research* 24 (12): 2033–40.

Plötner, Björn, Markus Nurmi, Axel Fischer, Mutsumi Watanabe, Korbinian Schneeberger, Svante Holm, Neha Vaid, et al. 2017. "Chlorosis Caused by Two Recessively Interacting Genes Reveals a Role of RNA Helicase in Hybrid Breakdown in Arabidopsis Thaliana." *The Plant Journal: For Cell and Molecular Biology* 91 (2): 251–62.

Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics* 81 (3): 559–75.

Rambaut, Andrew. 2012. "FigTree v1. 4."

Rieseberg, Loren H., and John H. Willis. 2007. "Plant Speciation." *Science* 317 (5840): 910–14.

Rowan, Beth A., Danelle K. Seymour, Eunyoung Chae, Derek S. Lundberg, and Detlef Weigel. 2017. "Methods for Genotyping-by-Sequencing." *Methods in Molecular Biology* 1492: 221–42.

Sandstedt, Gabrielle D., Carrie A. Wu, and Andrea L. Sweigart. 2020. "Evolution of Multiple Postzygotic Barriers between Species in the Mimulus Tilingii Species Complex." https://doi.org/10.1101/2020.08.07.241489.

Schmickl, Roswitha, Juraj Paule, Johannes Klein, Karol Marhold, and Marcus A. Koch. 2012. "The Evolutionary History of the Arabidopsis Arenosa Complex: Diverse Tetraploids Mask the Western Carpathian Center of Species and Genetic Diversity." *PloS One* 7 (8): e42691.

Sicard, Adrien, Christian Kappel, Emily B. Josephs, Young Wha Lee, Cindy Marona, John R. Stinchcombe, Stephen I. Wright, and Michael Lenhard. 2015. "Divergent Sorting of a Balanced Ancestral Polymorphism Underlies the Establishment of Gene-Flow Barriers in Capsella." *Nature Communications* 6 (August): 7960.

Smith, Lisa M., Kirsten Bomblies, and Detlef Weigel. 2011. "Complex Evolutionary Events at a Tandem Cluster of Arabidopsis Thaliana Genes Resulting in a Single-Locus Genetic Incompatibility." *PLoS Genetics* 7 (7): e1002164.

Supek, Fran, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. 2011. "REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms." *PloS One* 6 (7): e21800.

Tian, D., M. B. Traw, J. Q. Chen, M. Kreitman, and J. Bergelson. 2003. "Fitness Costs of R-Gene-Mediated Resistance in Arabidopsis Thaliana." *Nature* 423 (6935): 74–77.

Tian, Tian, Yue Liu, Hengyu Yan, Qi You, Xin Yi, Zhou Du, Wenying Xu, and Zhen Su. 2017. "agriGO v2.0: A GO Analysis Toolkit for the Agricultural Community, 2017 Update." *Nucleic Acids Research* 45 (W1): W122–29.

Todesco, Marco, Sang-Tae Kim, Eunyoung Chae, Kirsten Bomblies, Maricris Zaidem, Lisa M. Smith, Detlef Weigel, and Roosa A. E. Laitinen. 2014. "Activation of the Arabidopsis Thaliana Immune System by Combinations of Common ACD6 Alleles." *PLoS Genetics* 10 (7): e1004459.

Weiß, Clemens L., Marina Pais, Liliana M. Cano, Sophien Kamoun, and Hernán A. Burbano. 2018. "nQuire: A Statistical Framework for Ploidy Estimation Using next Generation Sequencing." *BMC Bioinformatics* 19 (1): 122.

Wickham, Hadley. 2009. *Ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. Springer Publishing Company, Incorporated.

Yaffe, Hila, Kobi Buxdorf, Illil Shapira, Shachaf Ein-Gedi, Michal Moyal-Ben Zvi, Eyal Fridman, Menachem

Moshelion, and Maggie Levy. 2012. "LogSpin: A Simple, Economical and Fast Method for RNA Isolation from Infected or Healthy Plants and Other Eukaryotic Tissues." *BMC Research Notes* 5 (January): 45.

Yamamoto, Eiji, Tomonori Takashi, Yoichi Morinaka, Shaoyang Lin, Jianzhong Wu, Takashi Matsumoto, Hidemi Kitano, Makoto Matsuoka, and Motoyuki Ashikari. 2010. "Gain of Deleterious Function Causes an Autoimmune Response and Bateson-Dobzhansky-Muller Incompatibility in Rice." *Molecular Genetics and Genomics: MGG* 283 (4): 305–15.

Inbreeding Depression in a Natural *Arabidopsis arenosa* Population

A. Cristina Barragan[1], Maximilian Collenberg[1], Merjin Kerstens[1a], Rebecca Schwab[1], Felix Bemm[1,b], Ilja Bezrukov[1], Doubravka Požárová[2], Filip Kolář[2], Detlef Weigel[1]*

[1]Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany

[2]Department of Botany, Faculty of Science, Charles University, 128 01 Prague, Czech Republic

[a]Current address: Department of Molecular Biology, University of Wageningen, 6708, Wageningen, Netherlands

[b]Current address: KWS Saat, 37574 Einbeck, Germany

*Corresponding author: weigel@tue.mpg.de (D.W.)

Supplemental Figures

Fig S1. Information on sampled *A. arenosa* populations. Related to Fig 1.

Fig S2. RNA-seq analysis of the A279 and B772 families. Related to Fig 3.

Fig S3. QTL analysis from the B772 family A-D. Related to Fig 4.

Fig S4. Homozygosity in normal and abnormal plants. Related to Fig 5.

Fig S5. Pseudo-F_2 and F_1 *A. arenosa* phenotypes.

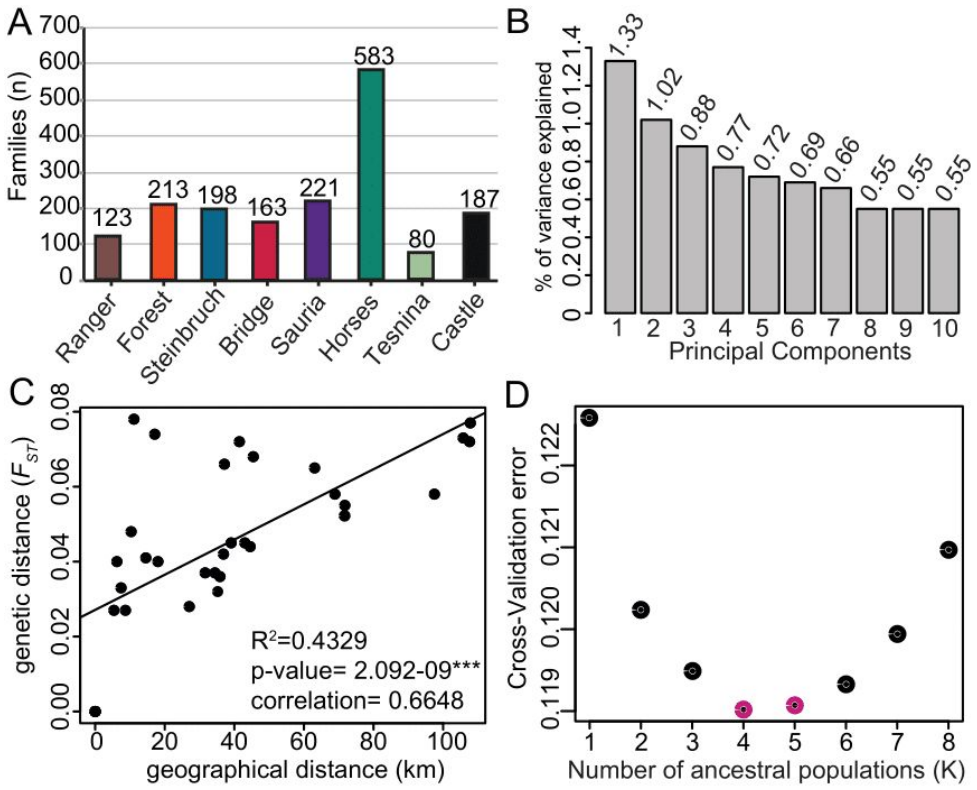


Fig S1. Information on sampled *A. arenosa* populations. A. Number of families sampled in each of the eight *A. arenosa* populations studied. **B.** Principal components and percentage of variance explained by each one. **C.** Correlation between genetic and geographical distance. **D.** ADMIXTURE cross-validation error. K=4 and K=5 have the lowest errors (pink). Related to Fig 1.

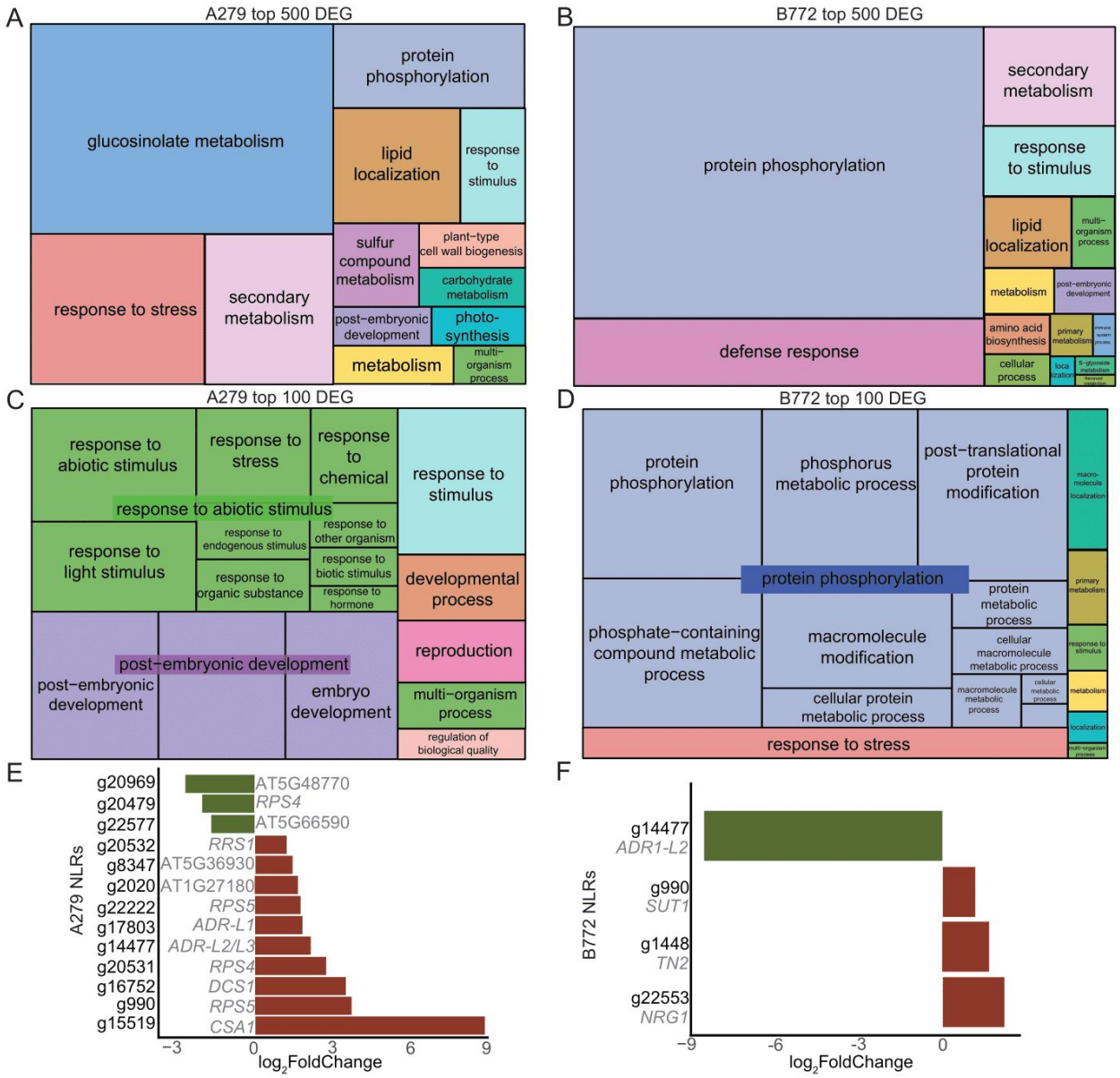


Fig S2. RNA-seq analysis of the A279 and B772 families. A-D. REVIGO Gene Ontology treemap of the top 500 (**A-B**) and the top 100 (**C-D**) DEG between normal and abnormal plants in the A279 and B772 families. Size of the square represents -$\log_{10}$(*p* value) of each GO term. **E-F.** Significant (|$\log_2$FoldChange| >1, padj value < 0.05) NLR orthogroup expression changes between normal and abnormal plants in A279 (**E**) and B772 (**F**). *A. thaliana* ortholog genes are in grey.

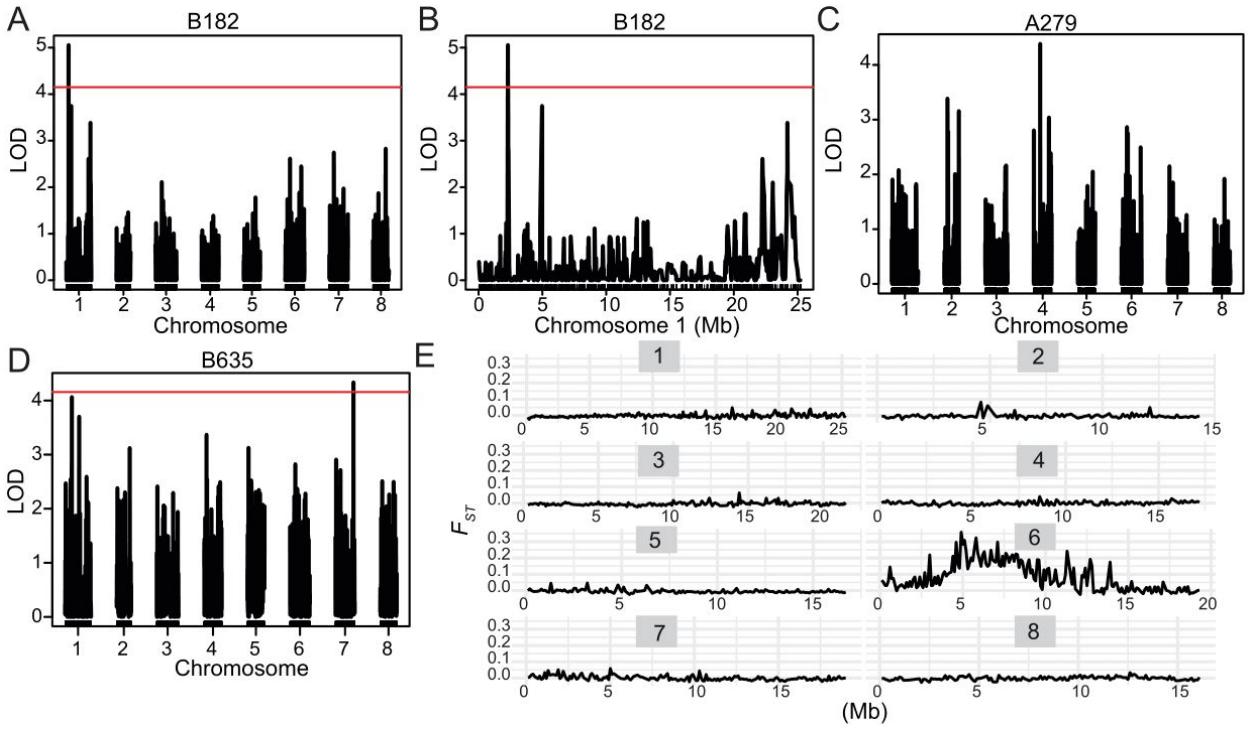


Fig S3. QTL analysis from the B772 family A-D. QTL analyses from the B182 (**A**, **B**), A279 (**C**) and B635 (**D**) families. The horizontal lines indicate 0.05 significance threshold established with 1,000 permutations. The absence of this family indicated the significance threshold is above the plotted values. **E.** Genome-wide fixation index (F_{ST}) in the B772 family. Higher values are seen exclusively in the first half of chromosome six.

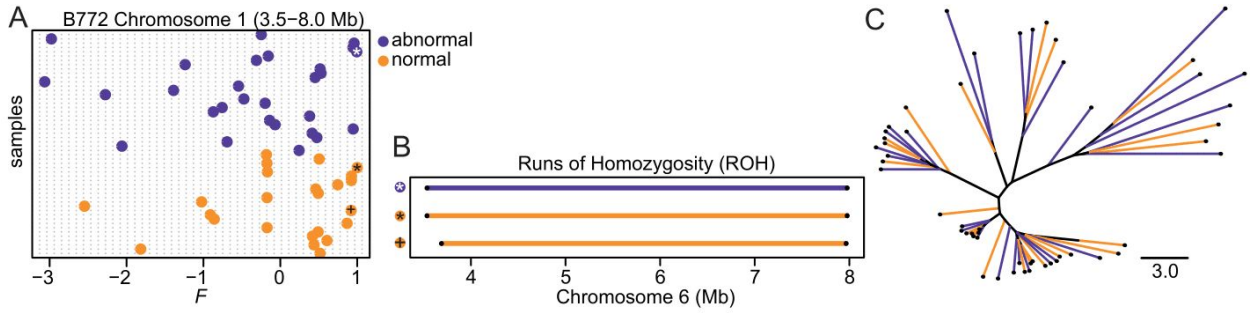


Fig S4. Homozygosity in normal and abnormal plants. A. Inbreeding coefficient (*F*) of plants showing an abnormal and a normal phenotype. Calculated from the 3.5-8 Mb region of chromosome 1. Abnormal plants do not tend to have a higher *F*. **B.** Runs of homozygosity (ROH) from the same genomic region. Only three plants show ROH. **C.** Neighbor-Joining tree of the 3.5-8.0 Mb region of chromosome 1. Individuals do not cluster by phenotype. Branch lengths in nucleotide substitutions are indicated.

Fig S5. Pseudo-F_2 and F_1 Phenotypes. A. Pseudo-F_2 plants from the B772 and **B.** the B635 families. Plants were grown at 16°C or 23°C. Temperature has no major effect on the abnormal phenotypes. Plants were five weeks old. **C.** F_1 B772 x A601 cross, no abnormal phenotypes are seen in contrast to pseudo-F_2 B772 plants. Plants were five weeks old. Scale bar represents 1cm.