ORIGINAL ARTICLE

Molecular Plant Pathology WILEY

The NLRomes of *Zea mays* NAM founder lines and *Zea luxurians* display presence–absence variation, integrated domain diversity, and mobility

Shawn Thatcher | **Mark Jung** | **Gayathri Panangipalli** | **Kevin Fengler** | **Abhijit Sanyal** | **Bailin Li** | **Victor Llaca** | **Jeffrey Habben**

Corteva Agriscience, Johnston, Iowa, USA

Correspondence
Jeffrey Habben, Corteva Agriscience, Johnston, Iowa, USA.
Email: jeffrey.habben@corteva.com

Abstract

Plant pathogens cause significant crop loss worldwide, and new resistance genes deployed to combat diseases can be overcome quickly. Understanding the existing resistance gene diversity within the germplasm of major crops, such as maize, is crucial for the development of new disease-resistant varieties. We analysed the nucleotide-binding leucine-rich repeat receptors (NLRs) of 26 recently sequenced diverse founder lines from the maize nested association mapping (NAM) population and compared them to the R gene complement present in a wild relative of maize, *Zea luxurians*. We found that NLRs in both species contain a large diversity of atypical integrated domains, including many domains that have not previously been found in the NLRs of other species. Additionally, the single *Z. luxurians* genome was found to have greater integrated atypical domain diversity than all 26 NAM founder lines combined, indicating that this species may represent a rich source of novel resistance genes. NLRs were also found to have very high sequence diversity and presence–absence variation among the NAM founder lines, with a large NLR cluster on Chr10 representing a diversity hotspot. Additionally, NLRs were shown to be mobile within maize genomes, with several putative interchromosomal translocations identified.

KEYWORDS
disease resistance, NLR, *Zea luxurians*, *Zea mays*

1 | INTRODUCTION

Plant disease resistance gene complements are the result of millions of years of co-evolution with pathogens. Resistance genes encode proteins that form a multilayer defence mechanism which can detect pathogen-associated molecular patterns or damage-associated molecular patterns through extracellular pattern recognition receptors (PRRs) and small, secreted pathogen effectors through intracellular nucleotide-binding leucine-rich repeat receptors (NLRs) (Jones & Dangl, 2006; Monteiro & Nishimura, 2018; Zipfel, 2014). PRRs are primarily transmembrane domain-containing proteins in which extracellular domains interact with pathogen-associated molecular patterns or damage-associated molecular patterns. This interaction can cause a conformational change that initiates a signalling cascade through the action of an intracellular kinase domain (Tang et al., 2017). Effectors are excreted by plant pathogens for a variety of purposes, including suppression of the plant defence responses that are triggered by PRRs (Irieda et al., 2019). These proteins are

often small, and typically undergo rapid evolution, meaning that the complement of pathogen effectors which plants interact with is in a constant state of flux (Newman et al., 2013; Sanchez-Vallet et al., 2018). As a result, plant species typically harbour hundreds of different NLRs, which have high sequence diversity and presence–absence variation (PAV) (Shang et al., 2022; Van de Weyer et al., 2019). NLRs have frequently been found to underlie dominant resistance phenotypes in many crop species, including rice, soybean, wheat, and maize (Deng et al., 2022; Liu, Kang, et al., 2020; Saintenac et al., 2013; Thatcher et al., 2022; Wang et al., 2021).

NLRs are typically composed of a nucleotide-binding (NB) domain, a series of leucine-rich repeats (LRR), and an N-terminal region which may include a coiled-coil, Toll/interleukin-1 receptor (TIR), or resistance to powdery mildew 8 (RPW8) domain (Shao et al., 2016). The relative abundance of these different N-terminal domains varies significantly across different plant species, with *Zea mays* possessing almost exclusively N-terminal coiled-coil domains. In addition to these canonical domains, NLRs occasionally harbour atypical integrated domains (Sarris et al., 2016). Integrated domains are thought to arise via rare recombination events that result in domains with high similarity to effector targets being integrated into NLR genes, which then detect the presence of effectors though direct interaction (Grund et al., 2019). Depending on their domain complement, NLRs can detect the presence of pathogen effectors through (i) direct interaction of an effector with canonical NLR domains, (ii) direct interaction of an effector with an integrated domain that mimics the effector's host target, or (iii) interaction with a host gene targeted by an effector (guardee) to detect alteration of its normal state by the pathogen (Cesari, 2018; Cesari et al., 2014; van der Hoorn & Kamoun, 2008). The direct interaction mechanism of integrated domain NLRs makes them particularly amenable to engineering, which has been shown to be able to expand the resistance spectrum of *RGA5* in rice (Liu et al., 2021). Additionally, "helper" NLRs typically contain all canonical NLR domains but do not function in direct detection of pathogen effectors, but instead act to transmit the signal of a "sensor" NLR (Wu et al., 2017). Helper NLRs have been found in a variety of plant species, and can be specific to a single sensor NLR or interact with a wide variety of sensors (Saile et al., 2020).

The NLR complements of dozens of plant species have been identified thus far, and recent work by Van de Weyer et al. revealed a nearly complete picture of the *Arabidopsis* NLRome, representing a significant step forward in the understanding of intraspecific NLR diversity (Van de Weyer et al., 2019). Since this initial work in *Arabidopsis*, the pangenomes of numerous plant species have been sequenced, leading to a much more complete picture of the total NLR complement of many species, including rice, barley, wheat, rapeseed, soybean, pigeon pea, lupin, and a number of other important crops and wild relatives (Dolatabadian et al., 2020; Garg et al., 2022; Jayakodi et al., 2020; Li et al., 2021; Liu, Du, et al., 2020; Shang et al., 2022; Walkowiak et al., 2020; Zhao et al., 2020). Such work has revealed extremely variable numbers of NLRs and levels of NLR diversity, with species NLR counts ranging from a few dozen to thousands per line. Maize is known to possess fewer NLRs than most other crop species, averaging less than 150, compared to rice (>450), soybean (>300), barley (>460), and wheat (>2500). However, a recent analysis of the NLR complement contained within the pangenome of lupin revealed that it possessed even lower counts (<60) and a surprisingly low level of diversity even within wild accessions (Garg et al., 2022). How different plant species cope with disease pressure using such extremely variable amounts of NLRs remains an open question.

Earlier studies in *Arabidopsis* and rice found that NLR PAV largely takes place through the expansion and contraction of large physically compact clusters of NLRs in a few locations within the genome (Jacob et al., 2013; Meyers et al., 2003). These regions are thought to represent evolutionary hotspots, where different NLRs may rapidly recombine to generate new sequence diversity (van Wersch & Li, 2019). Such loci are often associated with functional disease resistance genes, and analysis in several *Brassica* species has indicated that there is a correlation between complex loci containing densely packed NLRs and the presence of functional resistance genes (Cai et al., 2021; Dolatabadian et al., 2020; Golicz et al., 2016; Zhang et al., 2021). These hotspots have been implicated in conferring resistance to a wide array of different pathogens, including turnip mosaic virus and clubroot in *Brassica rapa* and powdery mildew in *Brassica napus* (Jin et al., 2014; Kato et al., 2013; Li et al., 2016). Cross-species comparisons within *Brassica* have shown that such complex NLR loci are subject to frequent birth and death events, with all members of some large NLR clusters showing high sequence homology to a singleton NLR of a closely related species (Fu et al., 2019; Zhang et al., 2021). More recently conducted pangenomic analyses have further shed light on the high level of intraspecies NLR diversity. One of the most extreme examples of this phenomenon is wheat, which was found to average over 2500 NLRs per line (Walkowiak et al., 2020). Comparison of 16 wheat lines revealed that individual lines contained up to 170 unique NLRs and several hundred NLRs that were only shared with two or fewer other lines. Surprisingly, only one third of the NLRs present in each wheat line were shared across all other lines surveyed. Such findings highlight the importance of sequencing diverse germplasm to obtain a more complete picture of the NLR repertoire of plant species.

At the protein level, early work in lettuce, as well as the more recent work in *Arabidopsis*, has shown that the different domains of NLRs are subject to differential evolutionary pressures (Meyers et al., 1998). NB-ARC domains, which are functional ATPases that control the activation states of NLRs, typically show high conservation, while LRRs and coiled-coil domains have much higher amino acid diversity (Qi et al., 2012). Additionally, there is evidence that sensor NLRs, which must keep pace with rapidly evolving pathogen effectors, may have much higher diversity than helper NLRs, which must maintain the ability to interact with multiple different immune system components (Wu et al., 2017). Knowledge about the array of NLRs throughout the maize genome is therefore key to identifying putative sensor NLRs that may be responsible for resistance to a given pathogen.

The maize nested association mapping (NAM) population was created from 26 maize founder lines in order to identify the genetic basis of complex traits (Yu et al., 2008). The recent effort to sequence and assemble the founder lines, which represent a diverse set of germplasm, presents an opportunity to obtain a much more complete picture of the resistance gene complement of this important crop species (Hufford et al., 2021). Maize pathogens cause significant crop loss annually, making it crucial to identify new sources of resistance genes (Mueller, 2016). Maize is thought to have been domesticated during a single event roughly 9000 years ago, implying that a significant portion of the resistance gene diversity in maize's wild ancestors may have been lost in modern-day varieties through the initial domestication event and subsequent breeding (Matsuoka et al., 2002; Yang et al., 2019). A more complete picture of the NLR complement of wild relatives may therefore enable the identification of new resistance genes that are not currently present in maize germplasm, a phenomenon that has been reported in several other species (Liu, Du, et al., 2020; Munoz-Amatriain et al., 2013; Radwan et al., 2008).

In this work, we sought to identify the NLR complement of maize NAM genomes to better understand the level of diversity, PAV, and integrated domains. Additionally, a wild relative of maize that was found to be resistant to several maize diseases, *Zea luxurians* (accession PI422162), was sequenced in order to compare its NLR complement to those of domesticated maize varieties. We found that the NAM founder lines contain a high level of diversity in their NLR complement and possess a number of integrated domains not previously associated with NLRs. The single *Z. luxurians* accession contained more different types of integrated domains than all NAM lines combined, indicating that this wild ancestor may represent a source of novel NLRs. We also identified putative interchromosomal translocations of NLRs, in which translocated NLRs were found to have maintained high similarity to the original sequence and expression levels of their nontranslocated counterparts. We posit that the natural movement of NLRs within the maize genome has direct regulatory implications for the movement of resistance genes via genome editing technologies, such as CRISPR (Svitashev et al., 2016).

2 | RESULTS

2.1 | Identification of NLRs in maize NAM founder lines and *Z. luxurians*

To identify NLRs in the 26 NAM founder lines, we employed HMMer to search for canonical NLR domains, including NB-ARC, coiled-coil, RPW8, TIR, and LRR domains (Eddy, 1995). In total, 2658 genes containing an NB-ARC domain were found across the NAM founder lines, with an average number of 102 per line (Tables 1 and S4). To search for regions with the potential to encode NLRs but no annotated NLR gene models, we translated each NAM genome into all six possible reading frames and employed HMMer to search for NB-ARC hits, stitching together partial hits within 10,000 bp to account

TABLE 1 Number of genomic NB-ARC domains, genes with NB-ARC domains, and genes with NB-ARC domains and LRR domains.

Genome	Genomic NB-ARCs	NB-ARC genes	NLR genes	NB-ARC genes in clusters (%)
B73	152	98	79	49.0
B97	154	111	92	59.5
CML103	163	95	75	52.6
CML228	148	103	81	51.5
CML247	135	100	79	54.0
CML277	145	106	81	55.7
CML322	143	103	81	59.2
CML333	160	105	86	58.1
CML52	164	99	81	57.6
CML69	160	109	86	57.8
HP301	137	98	84	43.9
Il14H	144	105	81	55.2
Ki11	154	103	84	53.4
Ki3	152	106	82	53.8
Ky21	156	103	81	50.5
M162W	142	102	81	58.8
M37W	167	110	89	55.5
Mo18W	149	100	77	52.0
Ms71	140	99	78	49.5
NC350	164	110	87	58.2
NC358	157	112	84	56.3
Oh43	156	106	86	56.6
Oh7B	140	91	73	46.2
P39	146	95	73	53.7
Tx303	136	92	73	45.7
Tzi8	150	113	86	57.5

for potential splicing. This analysis yielded an additional 47 genomic regions that could potentially encode NLRs, but did not have RNA-seq-based gene model assemblies resulting from any of the tissue types surveyed. These results are largely in agreement with the recent findings of Hufford et al. and highlight the relatively low NLR content of maize compared to other crop species, such as rice (Shang et al., 2022). Most of these genes also included LRR (77.0%) and N-terminal coiled-coil domains (77.9%). Unlike in dicots, TIR and RPW8 domains were extremely rare, only appearing in 1.3% and 2.6% of NB-ARC domain-containing genes, respectively. TIR domains were never found in combination with LRRs, while RPW8 NLRs almost always contained LRRs. A total of 14 potential domain architectures were found to occur in at least five different genes across the NAM founder lines (Figure 1a).

The majority of NLRs (57%) had canonical structures, with a coiled-coil region, followed by an NB-ARC domain, terminating in a series of LRRs. Some alternative structures were abundant, including proteins containing only a coiled-coil and an NB-ARC domain
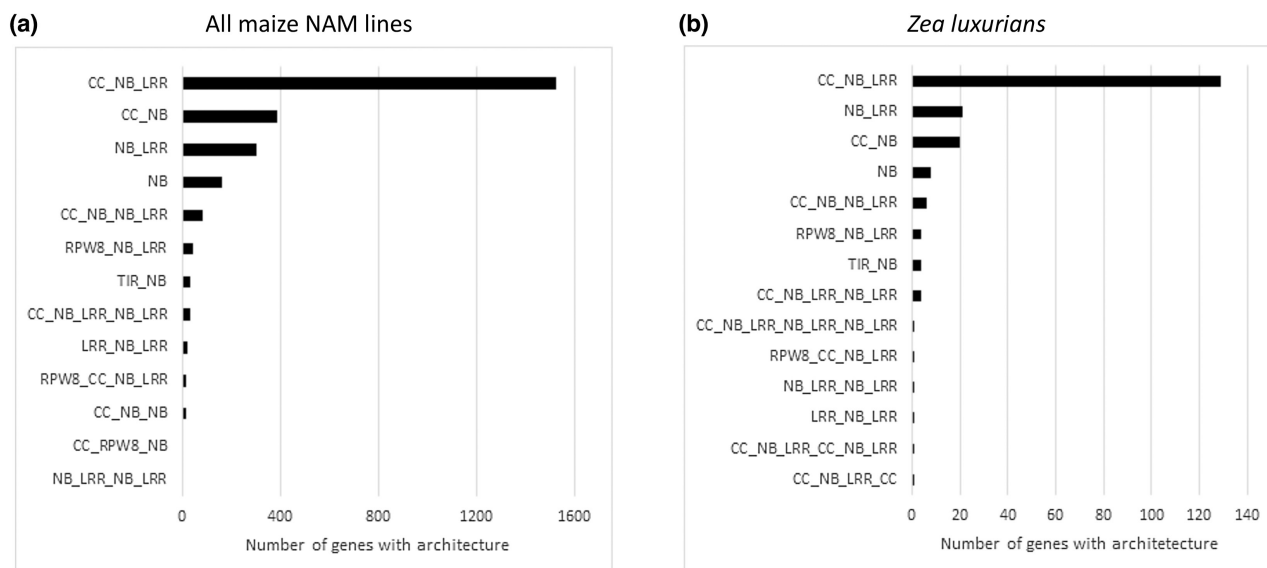
FIGURE 1 Distribution of domain architectures in maize and *Zea luxurians*. The abundance of different domain architectures involving canonical NLR domains is shown for (a) the sum of all 26 maize NAM lines and (b) a single *Z. luxurians* line. CC: coiled-coil, NB: NB-ARC, LRR: leucine-rich repeat, TIR: Toll/interleukin 1 receptor, RPW8: resistance to powdery mildew 8 domain.

(14.6%), proteins containing only an NB-ARC domain and LRRs (11.4%), and proteins with an NB-ARC and no other canonical NLR domains (6.1%). Interestingly, several genes were identified that may be the result of a two-NLR fusion. For example, 25 of the NAM founder lines contained an NLR on Chr6 that had a coiled-coil–NB-ARC–LRR–NB-ARC–LRR structure, with a C-terminal integrated no apical meristem-associated (NAM-associated) domain (Cheng et al., 2012).

Wild relatives of crop species, including maize, have previously been shown to represent valuable sources of novel disease resistance traits (Lennon et al., 2016; Mammadov et al., 2018). Our greenhouse assays indicated that the wild maize relative *Z. luxurians* PI422162 was resistant to southern corn rust, grey leaf spot, and northern leaf blight (data not shown). To assess the NLR diversity of this wild relative, we sequenced its genome via PacBio. The resulting assembly was 4.9 Gb and contained 179 hybrid scaffolds with an N50 value of 95.9 Mb. The nonscaffolded sequences consisted of 2990 contigs with a total length of 696.6 Mb. This large genome size indicated that, unlike the highly inbred NAM lines, this wild relative possessed a very high level of heterozygosity. RNA-seq and de novo prediction were then employed to annotate the coding regions of the genome. We then applied HMMer to the predicted *Z. luxurians* proteome to determine its NLR complement. A larger number of NLRs was identified in *Z. luxurians* compared to the NAM lines, with a total of 202 genes containing NB-ARCs found. Although the heterozygosity present in the *Z. luxurians* genome may have partially caused this increase, it probably does not account for all of it, given the higher overall diversity of integrated domains in *Z. luxurians* (see Section 2.2) as well as novel sequenced-based clusters that formed when *Z. luxurians* NLRs were clustered together with NAM line NLRs (see Section 2.3). The overall distribution of different domain architectures was nearly identical to that in the NAM population, with 63.8% of the NLRs possessing the canonical coiled-coil, NB-ARC, LRR structure. A total of 14 different architectures were identified, all of which were also present in the NAM founder lines (Figure 1b).

2.2 | Integrated domains in NLRs of the maize NAM founder lines and *Z. luxurians*

We next employed HMMer to search for any atypical integrated domains within the NAM NLRome. After identifying all domains via HMMer, custom Python scripts were employed to filter out all hits that overlapped canonical NLR domains. The resulting set of potential atypical domains was then filtered loosely (E-value ≤ 0.1) and strictly (E-value ≤ 0.01 and at least 40% of the domain covered). The loosely filtered set contained a number of "domains of unknown function" and canonical domains with very poor coverage. Although the majority of these hits are probably false positives, some may represent true integrated domains that have undergone significant divergence after their neofunctionalization and are reported in Table S1. After filtering for high-confidence domain calls and collapsing redundant domains, a total of 19 strictly filtered unique integrated domains were found across all NAM NLRs (Figure 2a).

The most frequent integrated domain was a kinase, which appeared in two or three NLRs in each NAM founder line. The next most common was a paired amphipathic helix domain, followed by a NAM-associated domain, which occurred in the unique two-NLR fusion gene described earlier. Although the unfiltered set included many low-quality domain hits found in only a single gene, the more strictly filtered set only included two domains that appeared uniquely in a single gene in only one NAM founder line (zf-RVT and UvsW). A single NLR, occurring in 14 of the NAM founder lines, possessed two integrated domains, which were found in the middle
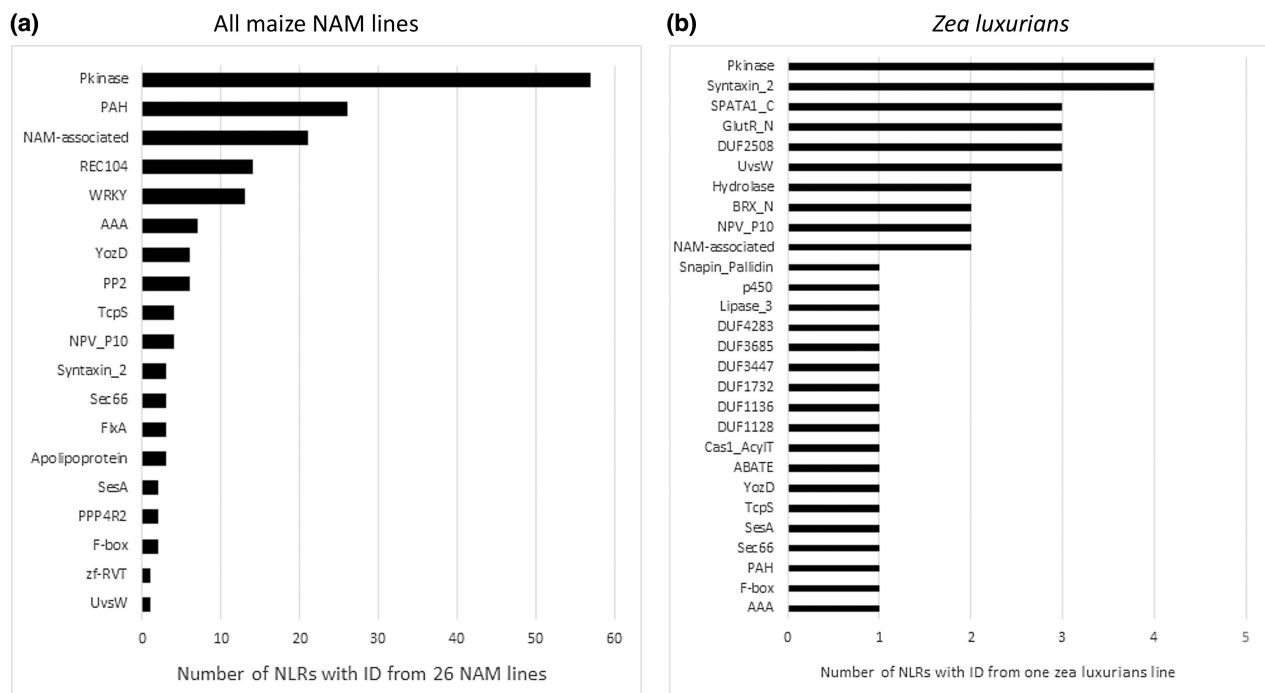
FIGURE 2 Distribution of integrated domains in maize and *Zea luxurians*. Atypical integrated domain counts identified via HMMer searches of NLR proteins are shown for (a) the sum of all 26 NAM genomes and (b) a single *Z. luxurians* line.

of the protein between the NB-ARC domain and LRRs (Meiotic recombination factor, REC104) and at the end of the protein (tyrosine kinase). Even when employing a loose filter, very little overlap was found with the *Arabidopsis* pan-NLRome, with only kinase and AAA-ATPase domains appearing in both sets (Van de Weyer et al., 2019). Surprisingly, a similarly low level of integrated domain overlap was found with the recently published rice NLRome, with only integrated NAM domains found in both sets (Shang et al., 2022).

We next employed HMMer against the NLRome of *Z. luxurians*. Although only one *Z. luxurians* accession was included, higher integrated domain diversity was found within it than in all NAM founder lines combined, with a total of 28 integrated domains identified in the strictly filtered set (Figure 2b). Additionally, 12 out of the 19 integrated domains found in the NAM lines (63%) were also found in the *Z. luxurians* NLR complement (Figure 2b). *Z. luxurians* NLRs contained a higher proportion of domains with no known function. Despite the very low level of integrated domain overlap when comparing both maize and *Z. luxurians* to rice, one uncharacterized integrated domain found in *Z. luxurians* (DUF4283) was also found in the rice NLRome. *Z. luxurians* contained two genes with multiple integrated domains, including one with an N-terminal kinase and a C-terminal hydrolase.

2.3 | Distribution of NLRs in the NAM genomes reveals high PAV

We next examined the distribution of all identified maize NLRs throughout the NAM genomes. Because NLRs are known to have high PAV and diversity, we hypothesized that there may be cases

where a genomic region contains a pseudogene in one line, but a fully functional gene in other lines. To address this, we employed HMMer to scan all maize genomes for regions with the potential to encode NB-ARC and LRR domains and removed those that overlapped with annotated genes. This analysis revealed potential pseudogenes present in all 26 lines, ranging from 35 in CML247 to 68 in CML103 (Table 1). The number of pseudogenes present in a given line was uncorrelated ($r^2 = 0.035$) with the number of annotated genes, making it unlikely that variable annotation quality was the reason that some lines had higher pseudogene counts. We further employed NLR-Annotator to confirm the results of our HMMer-based genomic scans, which yielded similar results (Steuernagel et al., 2020).

To assess whether pseudogenes in one line were closely related to functional genes in other lines, we extracted all nucleotides encoding the NB-ARC domains of all genes and pseudogenes. These sequences were then aligned in a pairwise manner, and alignments were used to generate phylogenetic trees via the maximum-likelihood method. All pairwise comparisons examined contained several pseudogenes from one line that clustered closely with genes from another line. Such clusters could be found across many different chromosomal locations. The B73:Mo18W pairwise comparison revealed a number of these instances, including a Chr3 gene in B73 (*Zm00001e018195*) which is located at 129 Mb. The Mo18W genome contains nucleotides with the potential to encode an NB-ARC domain that clusters very closely with the B73 gene NB-ARC domain (98.3% sequence identity), but no actual gene was found to be produced at this locus. Similar genomic/genic NB-ARC clusters were found throughout the genome, including Chr1 (Mo18W *Zm000034a005849*), Chr2 (Mo18W *Zm00034a016521*), Chr4

(Mo18W *Zm000034a031848*), Chr6 (B73 *Zm00001e031193*), Chr7 (Mo18W *Zm00034a051957*), and Chr10 (B73 *Zm00001e039226*). Although misannotation could account for some cases of potential pseudogenization, it probably does not explain most of them because all genomes were annotated in an identical manner, using RNA-seq data from the same set of tissues as part of the NAM sequencing project (Hufford et al., 2021).

To obtain a picture of the physical distances between and the overall distribution of NLRs in the maize genome, we plotted the physical locations of all NLR genes and pseudogenes in the 26 NAM founder lines, with B73 shown as a representative line (Figure 3). The overall distribution of NLRs was mostly consistent across the different NAM founder lines. NLRs were found to be distributed as singletons and small groups throughout the genome, but many existed in a few large clusters of variable size in which many NLRs were concentrated in a small genomic space. For the purpose of this analysis, physically clustered genes where those considered to reside within 1 Mb of another NLR. Expanding this distance to 2 Mb did not dramatically change the number of clustered NLRs, with an average of 54% of NLRs clustered using a 1 Mb cut-off and an average of 59% clustered using a 2 Mb cut-off. This rate was very similar to the one found in the recently published rice NLRome (60.0%), despite the significantly different total NLR counts of the two species.

The largest physical cluster exists on Chr10, which contains from 6 (CML52) to 20 (CML333 and CML69) NLR genes and represents an average of 14% of the total NLR content in maize genomes. This cluster also contains a large number of genomic NB-ARCs without definitive gene models, with the most extreme example being M37W, which had 17 NLR genes and 18 genomic regions with potential to encode NB-ARCs, but gene models derived from RNA-seq data (Figure 3). Unsurprisingly this cluster also has a high degree of PAV and allelic diversity, and sequence-based clustering revealed that it actually comprises two groups that are distinct at the sequence level but in very close proximity physically (Table S3).

Next, we analysed the PAV of NLRs that were not physically clustered across the 26 NAM lines and plotted their distribution (Figure 3). Such genes are denoted by asterisks and displayed as partially filled boxes where more light blue colour indicates presence in more NAM lines and more light red colour indicates absence in more NAM lines. The vast majority of unclustered NLRs were present in all or nearly all NAM lines, although several rarer genes were identified. For example, Chr5 contains two regions capable of producing NLRs at approximately 20 and 32 Mb. Each region contains an annotated NLR gene in roughly half of the NAM lines, although only CML52 produces transcripts at both locations. A small number of locations have the potential to encode NLRs, but no transcription was detected, and no gene models were assembled across any of the NAM lines. Such regions, including a Chr8 location (B73: 73 Mb), two Chr9 locations (B73: 121 and 124 Mb), and a Chr10 location (B73: 121 Mb), may represent genes that were completely lost due to
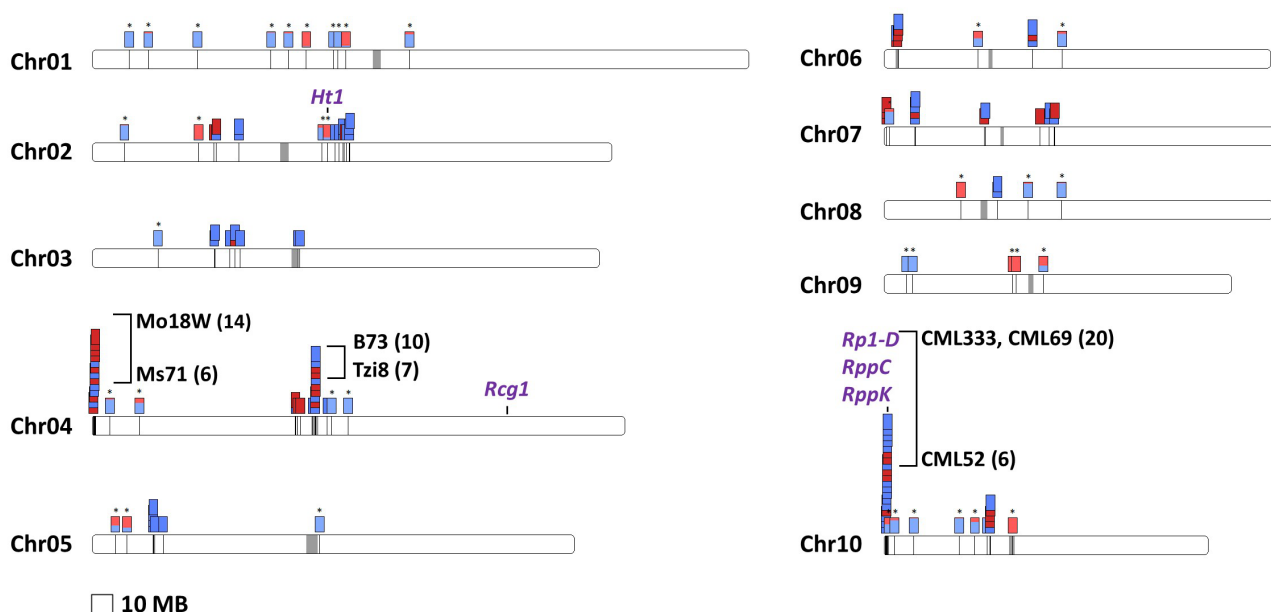

FIGURE 3 Distribution of NLR genes and pseudogenes across B73 chromosomes. NLR genes are represented with blue boxes, while potential pseudogenes are shown as red boxes. Genes that occur within 1 Mb of another NLR are stacked vertically in order to represent physical clusters. Genes not found within physical clusters were assessed for occurrence across the 26 NAM lines and are displayed as two-coloured boxes (denoted with asterisks), where the amount of light blue denotes the fraction of NAM lines with a gene present at the location and light red denotes the fraction of NAM lines without a gene at that location. The NAM lines with the maximum and minimum number of NLRs found in the three largest physical clusters are denoted in parentheses. Loci containing cloned NLRs, including those absent from all NAM lines, are denoted in purple.

pseudogenization or could produce functional NLRs in other maize lines not included in this dataset.

The percentage of genes containing NB-ARCs that existed within clusters was positively correlated ($r^2 = 0.50$) with the total number of NB-ARC-containing genes in a given genome, indicating that much of the differences in NLR counts between NAM genomes arose from expansion or contraction of clustered genes. Pairwise comparisons of NAM lines showed that on average, there was a difference of 8.2 clustered NB-ARC-containing genes but only 3.3 singleton NB-ARC-containing genes. The most extreme case can be seen when comparing Oh7B (91 NB-ARC genes) and Tzi8 (113 NB-ARC genes), where variation in clusters accounts for 100% of the 22-gene difference.

Overlaying the locations of functional maize NLRs onto the B73 genome revealed a bias towards highly variable regions with high copy number variation or PAV. Three of the cloned NLRs fell within the very large Chr10 NLR cluster, including the well-known common rust resistance gene *Rp1-D* and the recently cloned southern rust resistance genes *RppC* and *RppK* (Chen et al., 2022; Collins et al., 1999; Deng et al., 2022). The other two maize resistance loci, including the anthracnose stalk rot resistance gene pair *Rcg1a* and *Rcg1b* as well as the northern corn leaf blight locus *Ht1*, did not reside in physical clusters (Broglie et al., 2009; Thatcher et al., 2022). However, they had much higher rates of PAV compared to most singletons, with *Rcg1* being absent from all NAM lines (approximate location denoted based on marker data) and *Ht1* only being present in a handful of lines, with very low expression in lines in which it did exist.

Although paired NLRs in a head-to-head configuration have been reported to exist frequently in other species (Grund et al., 2019; Lee et al., 2021; Stein et al., 2018), we found only one potential NLR pair in the NAM founder lines. This pair was located at approximately 235 Mb on Chr2 in 10 of the NAM founder lines (B73: *Zm00001e011300* and *Zm00001e011302*) and was separated by around 8 kb in all lines. This distance is greater than that typically associated with paired NLRs (van Wersch & Li, 2019), and a small ribosomal protein was also annotated between the NLRs, although it had no detectible expression in any of the lines. Despite the distance and potential intervening gene, these NLRs appeared to be highly co-regulated, averaging an r^2 of 0.97 across different tissue types (Table S2). A survey of the teosinte genome also revealed only a single head-to-head NLR pair, which was separated by only 595 bp. Both genes in the pair lacked any LRR domains and did not share any significant sequence homology with the maize pair.

We next employed OrthoAug to cluster the protein sequences of all NAM NLRs in order to determine their relationships (Table S3) (Ekseth et al., 2014). We identified 158 clusters, 20 of which we classified as "core" NLR clusters, with all NAM founder lines containing at least one member. A total of 15 clusters were present in all but one NAM founder line and 11 were missing from only two NAM founder lines. On average, clusters contained at least one member in 16 out of the 26 NAM founder lines, indicating that PAV was the norm for most NLRs across the lines.

2.4 | Chromosomal translocation of NLRs

NLRs are known to be a very diverse group of genes, with high PAV, high Ka:Ks ratios, and frequent intergenic crossovers. We therefore hypothesized that NLRs may also be mobile within the genomes of maize lines. In order to look for NLR mobility, we examined the previously generated OrthoAug clusters for outliers with different chromosomes or significantly different positions relative to other members. Although the vast majority of NLRs (98.7%) resided in groups that contained similar positions on the same chromosome, several outliers were also present (Table S3).

The most dramatic outliers were found in Oh7B, which contained 11 NLRs on Chr9 that clustered with Chr10 NLRs from all other NAM founder lines. This apparent NLR translocation spanned genes that normally range from approximately 1.5 Mb to 28.4 Mb on Chr10. Earlier research using chromosomal probes identified this nonreciprocal translocation several years ago (Albert et al., 2010), but the fact that it resulted in the movement of the largest NLR cluster in maize was not previously noted. Interestingly, the translocated NLRs had an average similarity of 99.23% protein identity to their closest Chr10 orthologues in the other NAM founder lines.

Several putative smaller translocations were also identified from our clustering approach, which we examined in more detail. Initial clustering was carried out at the protein level, but this approach requires genes to be expressed or predicted correctly, making it possible for a gene that was only annotated correctly in one line to be incorrectly clustered because it was not assembled or predicted correctly in other lines. To address this possible false clustering, we employed BLAST to search the genomic regions encoding NLRs that clustered with genes on other chromosomes against the genomes of all NAM founder lines.

Several rare translocations that were predicted from protein clustering were also reproduced at the nucleotide level, including a putative Chr2-to-Chr10 translocation. One mixed cluster contains 8 Chr10 genes and 10 Chr2 genes, while the other contains 3 Chr10 genes and 10 Chr2 genes. These proteins were reclustered using MUSCLE, followed by both nearest-neighbour joining and maximum-likelihood models, which both gave the same result: two distinct clusters, which are formed from a mixture of Chr2 and Chr10 genes (Figure 4a) (Edgar, 2004; Saitou & Nei, 1987). Within the two major clusters, subclusters did break out by chromosomal location, but these distances were relatively minor.

One alternative explanation besides transposition is that the rapidly evolving nature of NLRs caused two separate clusters to undergo convergent evolution. To assess the likelihood of this, we performed a separate analysis clustering only the NB-ARC domains, which are typically under purifying selection. This analysis revealed the same clustering pattern as the full protein sequences, indicating that convergent evolution of the rapidly evolving portions of the proteins does not explain the similarity of these genes (Figure 4b). Subsequent expression analysis revealed relatively low Manhattan distances for pairwise comparisons within these
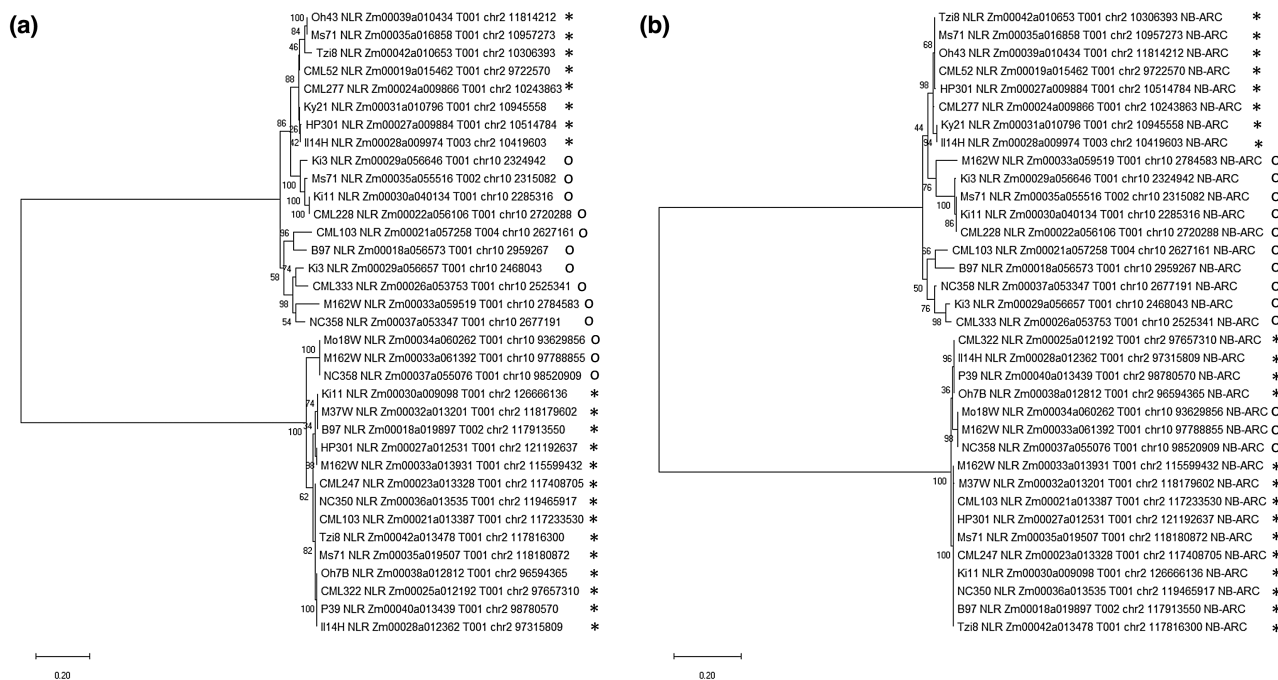
(a)

(b)

FIGURE 4 Chr2 and Chr10 NLRs form two distinct mixed clusters. Whole protein (a) and NB-ARC domains (b) were aligned and clustered via maximum likelihood. Bootstrap values are representative of 50 replicates. Names are displayed in the following format: genome, gene class, transcript name, chromosome, chromosome position. Chr2 genes are marked with "*," while Chr10 genes are marked with "o."

clusters, providing further evidence for their relatedness (see Section 2.6). Although transposition therefore remains the most likely explanation, the most closely related interchromosome homologues in the mixed clusters (Mo18W Chr10 *Zm00034a060252* and Tzi8 Chr2 *Zm00042a013478*) were only 89.6% identical and 92.3% similar.

2.5 | Relationship of *Z. luxurians* to maize NAM NLRs

We reclustered all NAM NLR proteins with *Z. luxurians* NLRs to determine the relatedness of their NLR complements (Table S3). Out of the 202 NLRs found in *Z. luxurians*, 167 (82.7%) clustered with maize NLRs. The addition of *Z. luxurians* NLRs did not substantially alter the maize clustering, although one previously unclustered gene (NC350 Chr10 3.8 Mb *Zm00036a038916*) clustered with a *Z. luxurians* gene (*dpzl0021g0539850*.725). This raises the possibility that other NLRs which are present at low frequency in maize may also be found in its wild relatives. Furthermore, the 35 *Z. luxurians* NLRs that did not cluster with any NAM founder lines may represent a source of new resistance gene diversity.

2.6 | NLR gene expression

We next examined the expression of NLR genes across a variety of tissue types present as part of the recent maize NAM sequencing effort (Hufford et al., 2021). This analysis contained a total of

11 tissue types, including leaf, leaf base, leaf tip, shoot, root, plant embryo, endosperm, anther, tassel inflorescence, and ear inflorescence. These RNA-seq data were originally intended for transcriptome annotation and most tissues only contained two biological replicates, reducing the statistical power of differential expression testing. We therefore only used the data to assess broad expression differences across tissues and have noted all cases where the two biological replicates are substantially divergent (≥2-fold difference), and a third biological replicate would be required to get a more accurate expression estimate (Table S4). We also supplemented these public data with additional RNA-seq libraries, which contained four biological replicates from each NAM founder line constructed from R1 leaves, a developmental stage at which plants often encounter pathogen challenge in the field.

NLRs were found to be expressed at a significant level across all tissues surveyed (average fragments per kilobase of transcript per million mapped reads [FPKM] of 6.75), with the highest average expression found in vegetative tissue (Table S4). Endosperm had the lowest median NLR expression, followed by embryo, anther, ear inflorescence, and tassel. All vegetative tissues had similar levels of average NLR expression, with shoot having the lowest average NLR expression (4.52 FPKM) and leaf base having the highest (7.85 FPKM). Surprisingly, although anther had a low median NLR expression (2.20 FPKM), one NLR present on Chr5 (B73: *Zm00001e013342*) was found in all lines and had the highest average expression of any NLR in any tissue, averaging 632.71 FPKM across the 26 lines. This gene had very low diversity across all lines, contained all canonical NLR domains, and was also expressed well above the average NLR expression in all vegetative tissues (average 33.56 FPKM). This

gene's extremely high expression and low diversity indicate that it may not play a canonical NLR role, although its exact function is currently unknown. We also sought to determine if the various domain architectures of NLRs noted earlier possessed different average expression levels across tissues. In general, NLRs that lacked LRR domains were expressed at a slightly lower level than those containing the canonical coiled-coil, NB-ARC, and LRR domains (average FPKM of 4.26 compared to 6.94). Interestingly, although only two distinct RPW8 NLRs were found in the NAM founder lines, they both possessed above average expression levels (average 18.98 FPKM).

Next, we sought to determine whether the clusters formed based on sequence homology also had similar expression patterns. Average Manhattan distances, which represent average differences in expression between each gene in each tissue for each pairwise comparison, were calculated for log-transformed expression values within each group, and the resulting average cluster distances were then compared to each other to identify clusters with very low and very high expression diversity. Analysis of the resulting distances revealed that most sequence-based clusters shared similar expression patterns across the 11 different tissue types, although some clear outliers could be found (Table S4). A clear cluster-wide tissue preference could be seen for some groups of NLRs, including a Chr2 root-preferential cluster (B97: *Zm00018a019897*) and a Chr4 endosperm-specific cluster (CML228: *Zm00022a035566*), but a vast majority of the sequence-based clusters (154/159) showed no strong tissue-specific expression. The rare tissue-specific expression patterns may have bearing on resistance gene selection for diseases that are known to invade specific tissues. Interestingly, the Chr10-to-Chr2 translocation that resulted in a sequence-based cluster containing a mixture of genes from different chromosomes also possessed a Manhattan distance that was similar to that of clusters containing nonmixed genes (23.2, compared to an average of 25.4).

Although most of the sequence-based clusters contained genes with similar expression patterns, a small number of clear outliers could be seen. An NLR found at approximately 24 Mb on Chr10 of every maize line (B73: *Zm00001e039226*) was predominantly expressed in leaf base tissue (average 6.80 FPKM), with very little expression in other tissues of most maize lines (average 0.49 FPKM). However, four NAM lines (CML103, M37W, NC358, and Tx303) had high average expression of this gene in several other tissue types, including R1 leaf (34.72 FPKM), leap tip (47.66 FPKM), shoot (19.82 FPKM), anther (6.74 FPKM), and tassel (3.34 FPKM). Clustering using the protein sequences of this gene from each NAM line indicated high sequence conservation, with the expression outliers not clustering apart from other members at the sequence level. Expression of some other NLRs was lost in specific NAM lines, including an NLR located at approximately 31 Mb on Chr2 (B73: *Zm00001e009019*), which had near ubiquitous expression across all tissues in all lines (average 7.74 FPKM), but had very little expression in CML52 (average FPKM 0.37). This gene was expressed at the leaf R1 stage (3.32 FPKM), indicating that it was not a pseudogene.

2.7 | Diversity within clusters at the whole gene and domain level

To investigate the degree of diversity within the previously generated sequence clusters, we calculated Shannon entropy at both the whole protein and the protein region level. Average Shannon entropy across whole proteins ranged from 0 to 1.27, with an average of 0.28 (Table S5). Although the cluster sizes examined ranged from 2 members up to 36, there was no significant correlation ($r^2 = 0.04$) between average Shannon entropy and group size. In general, groups with integrated domains had slightly higher protein-wide entropy than those without (0.39 versus 0.25), which fits with their proposed role in direct interactions with rapidly changing pathogen effectors.

Next, we sought to determine whether entropy varied across the different regions of the NLR proteins within each cluster. After Shannon entropy was calculated at each position within each cluster, these values were binned into the following protein regions: coiled-coil region, NB-ARC domain, spacer (region between NB-ARC and start of LRRs), LRRs, LRR spacers (regions in between LRRs), C-terminal region, and integrated domains. Coiled-coil regions, which have been proposed to play a role in inter- and intraprotein interaction, tended to have higher entropy than the whole protein (0.31). It has previously been reported that NB-ARC domains tend to have higher conservation than average within NLRs, and we found that this was broadly true across the clusters from the NAM founder lines (average Shannon entropy of 0.10). Spacer sequences between the NB-ARC domain and the LRR region also had low entropy on average (0.18). LRRs have been noted to have higher than average diversity, and we also found that they had high average Shannon entropy within clusters (0.38). Interestingly, the spacer regions between different LRRs on average had a similar level of entropy (0.38), but on a per-cluster basis, diversity of LRRs was often uncorrelated with diversity of LRR spacer regions. For example, many clusters had high LRR spacer entropy and low LRR entropy, while others showed the opposite pattern. The C-terminal regions of NLRs, which harbour no annotated domains, tended to have the highest level of entropy (average 0.53).

Integrated domains on average had a higher level of entropy than whole proteins (0.45). Despite typically being highly conserved in other genes, integrated NLR kinase domains in NLRs had very high entropy (0.68), possibly due to their transition away from catalytic activity and towards effector binding (Lai et al., 2016). The highest level of entropy was found in the integrated NAM domains (0.78), including those found in the novel Chr6 fused NLR, which also possessed well above average whole protein entropy (0.68). Still, some integrated domains were highly conserved, and these conserved integrated domains could be found in clusters with both high and low diversity at the whole protein level. For example, the Sec66 domain, which has been proposed to be involved in protein translocation, had extremely low entropy (0.04) within its 24-member cluster, despite this cluster having very high entropy at the whole protein level (0.69). Overall, the majority of clusters with integrated domains tended to have high entropy either within the integrated domain or
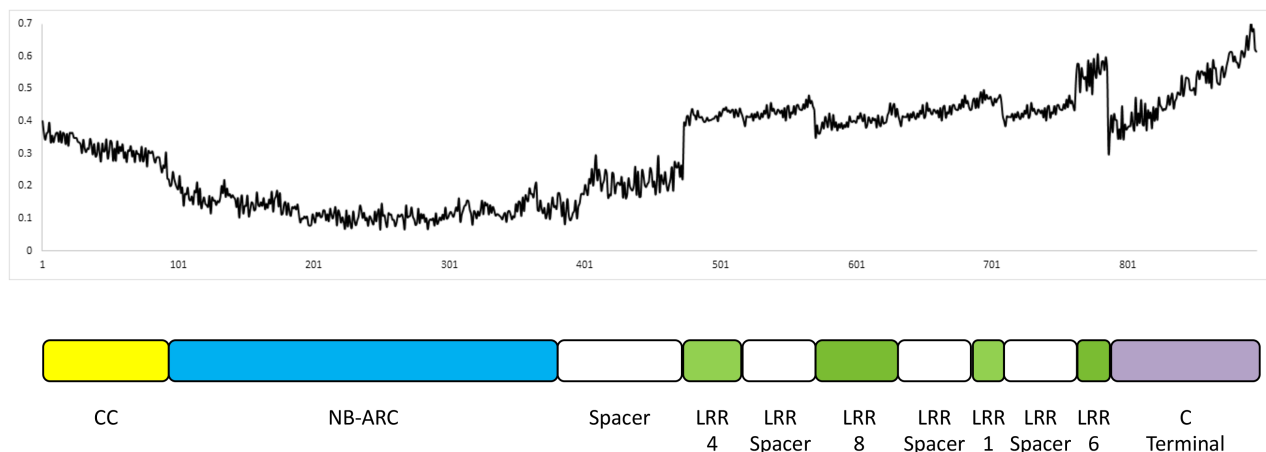
FIGURE 5 Average Shannon entropy across different NLR features. Shannon entropy of a composite constructed by averaging the entropy of all 158 NLR clusters.

at the whole protein level, which may be reflective of their proposed role in direct effector binding.

Finally, to get an overall picture of the average structure and entropy within maize NLRs, we constructed a "composite" NLR by averaging the entropy patterns of the most common domains. We first calculated the average Shannon entropy of each position within each domain/protein region for all clusters (Figure 5). For regions of variable size (spacers and C-terminal regions), the positions of entropy values were placed into 100 bins, with each bin representing 1% of the domain's total size in a given cluster, before averaging. Only the four common LRR HMM models were included in the resulting composite NLR (LRR_1, LRR_4, LRR_6, and LRR_8). These LRR domains showed the second highest level of entropy, with only the C-terminal domains having higher average values. The resulting composite NLR shows clear variability of NLR entropy throughout the different canonical domains (Figure 5).

3 | DISCUSSION

In this work, we identified the total NLR complement of the 26 maize NAM founder lines as well as a wild relative of maize, *Z. luxurians*. NLRs were found to have very high levels of PAV and allelic diversity and were distributed unevenly across maize genomes, with a single cluster on Chr10 representing a significant portion of the total complement of almost all lines. The physical clustering seen across the maize genome correlates well with sequence-based clustering, enabling physical placement of NLRs based on sequence alone. The ability to infer physical location from sequence is beneficial for techniques such as resistance gene enrichment sequencing (Jupe et al., 2013).

The total number of NLRs varied considerably across different NAM lines, with expansion and contraction of physically compact clusters accounting for the majority of the differences. The most extreme example was found in a comparison of the Chr10 cluster,

which varied from 6 (CML52) to 20 (CML333 and CML69) NLRs (Figure 3). Such highly compact NLR regions have been shown to contain functional NLRs in maize and other species, including the first cloned maize gene for common rust, *Rp1-D*, as well as the recently cloned southern corn rust resistance genes *RppC* and *RppK* (Chen et al., 2022; Collins et al., 1999; Deng et al., 2022). This pattern can also be seen in one of the most well-studied NLR clusters, the barley *Mla* locus (Wei et al., 1999). This locus is highly variable and comprises several related NLRs that confer resistance to several different pathogens, highlighting the benefits of NLR clusters (Bettgenhaeuser et al., 2021). The presence of such homologous sequences in close proximity probably enables a high rate of recombination that results in rapid gene gain, loss, and change (Leister, 2004; Michelmore & Meyers, 1998). These data indicate that individual NLRs within clusters are probably much more dispensable, but the clusters themselves serve as an essential reservoir for novel pathogen resistance genes. The much higher conservation of unclustered NLRs could be indicative of a more conserved role, although exceptions to this do exist. While the majority of unclustered NLRs had little PAV, the *Ht1* locus, which confers resistance to northern corn leaf blight, only had assembled NLR genes in eight of the NAM lines (Figure 3) (Thatcher et al., 2022). None of the NAM line alleles appear to contain the functional *Ht1* allele, but the comparatively high level of PAV at this unclustered locus once again points to high variability being indicative of functional NLR loci. These findings are in keeping with work in several other species, which has repeatedly pointed to the role of NLR clusters with high PAV in conferring disease resistance (Dolatabadian et al., 2020; Walkowiak et al., 2020; Zhang et al., 2021).

Despite a high level of variability in total NLR count, maize contained significantly fewer total NLRs than most other plants, such as *Arabidopsis* (>200), rice (>450), barley (>460), and wheat (>2,500) (Li et al., 2021; Shang et al., 2022; Van de Weyer et al., 2019; Walkowiak et al., 2020). Even accounting for possible missing annotation via an HMMer-based genomic scanning analysis, the total number of maize NLRs was far lower than the total

number in most other plant species. Although we found significantly more NLRs in the *Z. luxurians* accession, the majority clustered closely with NAM line NLRs and the increased count was partially driven by high heterozygosity (Table S3). The recently released pangenome of lupin revealed an even lower number of NLRs (67) and dramatically lower levels of diversity, with nearly all NLR genes fixed in both breeding lines and wild accessions (Garg et al., 2022). How maize and other species with relatively low NLR counts cope with disease pressure despite dramatically lower NLR counts and diversity requires additional investigation. Although NLRs have been found to confer resistance to several maize diseases, other resistance gene classes, such as PRRs, have also been implicated (Deng et al., 2022; Thatcher et al., 2022; Yang et al., 2021). A systematic study of the prevalence and diversity of these other maize R gene classes may therefore shed light on how maize is adapted to its relatively low NLR content.

Analysis of NLR expression across a wide array of tissue types indicated that genes in most sequence-based clusters shared tissue-specific expression patterns. The majority of NLRs were expressed ubiquitously, although some clear root-preferential clusters existed. A small number of outliers within sequence-based clusters exhibited different expression patterns compared to the rest of the cluster, including a Chr10 NLR that had leaf base-specific expression in most lines, but much broader expression in other lines. Such outliers may be indicative of neofunctionalization, although additional studies are needed to assess this possibility.

We also found evidence for mobility of NLRs within maize genomes, including several Mb of Chr10 that were found to have translocated into Chr9 in Oh7B, taking with it the largest NLR cluster in maize. NLRs in this translocated cluster were found to have high similarity to NLRs from other lines lacking the translocation. Another translocation was also identified between Chr10 and Chr2, which was evidenced by similarity at the whole protein and NB-ARC domain levels, genomic sequence similarity, and expression pattern similarity. These results are in keeping with findings in other species, where the uneven distribution of NLRs throughout the genome is thought to be the result of translocations (Borrelli et al., 2018). We found that translocated maize NLRs maintained high sequence similarity and similar expression patterns when compared to their nontranslocated counterparts.

Genome editing represents a powerful new technology for crop improvement, enabling both precise editing of gene sequences and the insertion of whole genes via template-based approaches (Svitashev et al., 2016). Here, we report that resistance gene translocations have previously occurred in maize, indicating that template-based genome editing mimics a process by which resistance genes have naturally moved among maize genomes. Furthermore, some of the translocated genes identified in this work were shown to have maintained similar expression profiles to their nontranslocated counterparts, indicating that genes moved via technologies such as CRISPR are likely to retain their native expression patterns and functions. Taken together, these results

strengthen the case for editing cloned resistance genes into susceptible maize lines in order to improve their disease tolerance (Yin & Qiu, 2019).

Our analysis of the entropy of NLRs at the protein level revealed several interesting findings. Different clusters were found to have significantly different levels of entropy, which may affect their likelihood of containing sensor NLRs that are responsible for resistance phenotypes. Previous work has shown that NLRs that directly interact with pathogen effectors have higher levels of diversity, probably to keep pace with the rapid nature of pathogen effector evolution (Prigozhin & Krasileva, 2021; Sanchez-Vallet et al., 2018). Additionally, entropy was found to vary substantially across different protein regions, with NB-ARC domains showing very low entropy, while LRRs, coiled-coil domains, and C-terminal domains possessed very high entropy. These findings are largely in keeping with the proposed roles of the canonical NLR domains. Interestingly, genes containing integrated domains were found to possess above average entropy levels both within and outside of those domains, which lends further support to the direct effector interaction model that has been proposed for this class of NLRs (Kroj et al., 2016).

Our survey of maize integrated domains identified several domains not previously known to be associated with NLRs. The integrated domains present in the NAM founder lines varied substantially, with only integrated kinases being present in all lines. The prevalence of this particular domain in NLRs is probably a response to targeting of kinase pathways by pathogen effectors, which has been reported in several other species (Shan et al., 2008; Zhang et al., 2010). Although paired amphipathic helix domains have not been reported as effector targets, they are known to have integrated into the NLRs of other species and may be targeted by pathogens due to their role in protein–protein interaction of transcription factors (Bowen et al., 2010; Kroj et al., 2016). We also found a novel integrated domain structure, where one gene contained both an N-terminal REC104 domain and a mid-protein kinase domain. Despite the lack of overlap of actual domains with the recently published *Arabidopsis* pan-NLRome, roughly similar rates of integrated domains were found in the two species, with 5.0% of the *Arabidopsis* pan-NLRome reported to contain integrated domains and 6.7% of the NAM founder lines possessing at least one. Rice, which had an integrated domain frequency of 8.9%, only shared one integrated domain in common with maize (NAM), while rice and *Arabidopsis* shared no common integrated domains. This lack of conservation in integrated domain repertoire indicates that NLR integrated domains may undergo rapid rates of gene birth and death. Perhaps most interestingly, we found that the integrated domain complement of the single *Z. luxurians* genome had higher diversity than all NAM founder lines combined, highlighting the potential utility of mining wild maize relatives for novel disease resistance genes. Taken together, our findings substantially expand on the existing knowledge of maize NLR diversity and mobility and provide new insight into the NLR complement of a wild maize relative.

4 | EXPERIMENTAL PROCEDURES

4.1 | NLR annotation and integrated domain analysis

Maize NAM genomes, gene annotations, and predicted proteins were obtained from MaizeGDB (https://www.maizegdb.org/). HMMer was used to identify conserved domains in the predicted protein sequences with default settings. For canonical NLR domain annotations, the following Markov models were obtained from Pfam (https://www.maizegdb.org/): NB-ARC, RP8W, TIR_2, TIR_1, TIR-like, Rx_N (coiled-coil), LRR_1, LRR_2, LRR_3, LRR_4, LRR_5, LRR_6, LRR_8, and LRR_9. For integrated domain analysis, all Pfam hidden Markov models were used. Hits that overlapped canonical domains were removed, and similar domains were collapsed through custom Python scripts and manual curation. The resulting set was filtered both loosely (E-value ≤0.1) and strictly (E-value <0.01, at least 40% coverage of the domain). For identification of NLR pseudogenes, each genome was translated into all six possible reading frames. HMMer was then used to search for NB-ARC domains and resulting partial hits within 10,000 bp were stitched together via custom Python scripts to identify genomic regions with the potential to code for NLRs. Genomic hits which overlapped annotated genes were then removed by combining genomic hit information with GFF file information from each NAM founder line. For genomic hits, at least 60% of the NB-ARC domain was required to be covered after stitching hits together to consider a region as having the potential to encode an NLR.

4.2 | NLR clustering

All predicted proteins from the NAM founder lines were clustered to determine their relationships. Diamond (v. 0.9.31.132) was first used to perform an all-by-all BLAST with default settings, followed by clustering with OrthAgogue (v. 1.0.2) with strict orthologues (Buchfink et al., 2015). Clusters were named using the convention "chromosome_position-in-MB_cluster-size." Clusters were then manually examined for outliers where genes from different chromosomes were clustered together. NB-ARC domain clustering was carried out by pairwise alignment using MUSCLE, followed by construction of a phylogenetic tree with a maximum-likelihood tree using 50 bootstraps, through MEGA software (v. 10.0.5) (Kumar et al., 2018).

4.3 | *Z. luxurians* genome assembly and annotation

For *Z. luxurians* genome sequencing and assembly, the Pacific BioSciences (Menlo Park, CA, USA) Sequel platform was used to generate long-read data. Ten-hour movies from 27 SMRT cells with v5 chemistry were filtered to a minimum subread length of 1 kb, yielding 53.2× genome coverage with a read N50 of 34.3 kb. Raw subreads were then trimmed and corrected via MECAT2 (github.com/xiaochuanle/MECAT2) with default parameters. The resulting dataset had a read N50 of 37.0 kb, with 31× genome coverage. Canu v. 1.8 (https://github.com/marbl/canu) was used to assemble the corrected reads into contigs. The following changes were made to the default parameters: "correctedErrorRate = 0.065, corMhapSensitivity = normal, ovlMerDistinct = 0.99." A minimum contig length of 30 kb was required. Additional sequence polishing was carried out by aligning raw PacBio subreads to the contig assembly with pbmm2 v. 0.12.0, followed by applying the pbbioconda (https://github.com/PacificBiosciences/pbbioconda) Genomic Consensus package (v. 2.3.2) Arrow algorithm to identify and correct remaining consensus errors in the contigs.

Next, the Bionano Genomics (San Diego, California) Saphyr platform was used to generate genome maps using the Direct Label and Stain system. DLE-1-labelled molecule data were filtered to generate a dataset with a molecule N50 of 295 kb and 364× coverage. This dataset was assembled via the Bionano Genomics Access software platform (Solve3.2.2_08222018) with the configuration file optArguments_nonhaplotype_noES_noCut_DLE1_saphyr.xml. The genome map assembly consisted of a total of 348 maps with a map N50 of 68.1 Mb and a total map length of 5255.4 Mb.

Polished contigs were used to generate hybrid scaffolds and Bionano maps with Bionano Genomics Access software (Solve3.3_10252018) using the DLE-1 configuration file hybridScaffold_DLE1_config.xml. The final result consisted of 179 hybrid scaffolds with a scaffold N50 of 95.9 Mb and a total scaffold length of 4906.1 Mb. The non-scaffolded sequences consisted of 2990 contigs with a total length of 696.6 Mb.

The *Z. luxurians* genome was annotated by first masking repeats via RepeatMasker (http://www.repeatmasker.org/). The repeat-masked genome was then used as the input for gene predictors. The de novo gene prediction programs Fgenesh (licensed v. 7.2.2), Augustus (v. 2.7 http://bioinf.uni-greifswald.de/augustus/), and SNAP (v. 2006-07-28 https://github.com/KorfLab/SNAP) were run with default settings, with training sets "monocots," "maize," and "rice," respectively. Expressed sequence tag (EST), cDNA, and long-read evidence-based gene structure modellers GMAP (v. 03-25-2018 http://research-pub.gene.com/gmap/) and PASA (v. 2.2.2 https://github.com/PASApipeline/PASApipeline/wiki) were run with the –max-intron-size parameter adjusted to 60,000. The protein evidence-based gene structure modeller SPLAN (v. 2.1.3 http://www.genome.ist.i.kyoto-u.ac.jp/~aln_user/spaln/) was run with default settings. Maize ESTs and cDNA from the public v4 annotations as well as internal gene models from other maize lines (data not published) were used as the evidence set for PASA. Poales ESTs, cDNA sequences from NCBI, and monocot transcripts from Phytozome were used as additional closely related species evidence for gene prediction with GMAP (parameters –min-identity 80 –min-coverage 95). Uniref100 plant protein sequences were employed as the evidence dataset for gene structure prediction with SPLAN. All gene annotation files were run through EvidenceModeler and the output was utilized to refine the gene boundaries in PASA. The final PASA

annotation file was merged with a tRNA predictions file from tRNA-ScanSE to obtain the final structural annotation file, along with fasta sequences of genes, cDNAs, coding sequences, and proteins.

4.4 | Plant growth conditions

During the seedling stage, all plants were grown in a greenhouse under 16 h/8 h light/dark conditions with a total of 25 µmol m^{-2} day^{-1} of total photosynthetic light, a vapour pressure deficit of 12 mbar, and day/night temperatures of 25/22°C (week 1), 22/19°C (week 2), 19/17°C (week 3), and 17/14°C (week 4). Because of the photoperiod sensitivity of some NAM founder lines, from week 5 until maturity, plants were grown under 12 h/12 h day/night conditions at 26/20°C (day/night) to reduce delayed flowering and maturity.

4.5 | RNA-seq library construction

Leaves from plants grown under the conditions listed above were sampled at the flowering stage (50–70 days, depending on the line), and their total RNA was isolated from ground frozen tissue with RNeasy (Qiagen Inc.), according to the manufacturer's protocol. Total RNA was then analysed for quality and quantity with the Bioanalyzer RNA Nano kit (Agilent Technologies) and normalized to 1 µg input per sample. Sequencing libraries were prepared according to Illumina Inc. (San Diego, CA) TruSeq mRNA-Seq protocols. mRNAs were isolated via attachment to oligo(dT) beads, fragmented, and reverse transcribed into cDNA by random hexamer primers with Superscript II reverse transcriptase (Life Technologies). The resulting cDNAs were end repaired, 3′-A-tailed, and ligated with Illumina indexed TruSeq adapters. Ligated cDNA fragments were PCR-amplified with Illumina TruSeq primers, purified with AmpureXP Beads (Beckman Coulter Genomics), and checked for quality and quantity with the Agilent TapeStation 4200 system with D1000 ScreenTape. Libraries were combined into one sequencing pool and normalized to 2 nM. The pool was denatured according to Illumina sequencing protocols, hybridized, and clustered on two flow cell lanes of a NovaSP flow cell using a NovaSeq 6000 platform. Single-end 50-base sequences and 8-base dual-index sequences were generated on a NovaSeq 6000 platform according to Illumina protocols. Data were trimmed for quality with a minimum threshold of Q13 and the resulting sequences were split by index identifier.

4.6 | Expression analysis

Publicly available RNA-seq expression data were obtained from a short-read repository (Sequence Read Archive [SRA], https://www.ncbi.nlm.nih.gov/sra, accessions ERX3793507–ERX3793986). RNA-seq reads obtained from SRA, as well as those generated through our own library construction and sequencing, were then quantified by running Salmon (v. 1.1.0) against the transcriptome of each NAM founder line, with GC bias correction (Patro et al., 2017). Transcript expression per library was then converted to gene expression per tissue using the DESeq2 package in R (v. 1.31.6) (Love et al., 2014). Because much of the SRA data only contained two replicates, quantifications which were significant (>0.3 FPKM) but differed by ≥2-fold between replicates or were calculated from only one library were noted as potentially unreliable in the supplemental data. Intracluster expression variability was assessed by calculating the average pairwise Manhattan distance of log-transformed expression values for each cluster using the spatial distance module of SciPy (v. 1.5.4).

4.7 | Shannon entropy assessment

Shannon entropy was calculated by first employing MUSCLE (v. 3.8.31) to align all proteins within each cluster, followed by construction of a consensus sequence based on these alignments. The entropy of each position was then assessed using the following formula:

$$X = - \sum P(x_i) \log P(x_i),$$

where x equals each position within the consensus protein and $\sum$ represents the sum of the variability at each position. For the diversity calculation, gaps in the alignment were treated as the 21st amino acid. Consensus sequences were then assessed via HMMer to identify conserved domains, followed by calculation of the average Shannon entropy within each domain for each gene cluster.

DATA AVAILABILITY STATEMENT
Genome sequence of *Z. luxurians* and RNA sequencing data from the 24 NAM founder lines are available at the SRA at www.ncbi.nlm.nih.gov/sra/ under accessions GSE206952 and JAOPSE000000000.

ORCID
Shawn Thatcher 🔟 https://orcid.org/0000-0002-4598-8863
Bailin Li 🔟 https://orcid.org/0000-0002-0678-6580

REFERENCES
Albert, P.S., Gao, Z., Danilova, T.V. & Birchler, J.A. (2010) Diversity of chromosomal karyotypes in maize and its relatives. *Cytogenetic and Genome Research*, 129, 6–16.

Bettgenhaeuser, J., Hernandez-Pinzon, I., Dawson, A.M., Gardiner, M., Green, P., Taylor, J. et al. (2021) The barley immune receptor Mla recognizes multiple pathogens and contributes to host range dynamics. *Nature Communications*, 12, 6915.

Borrelli, G.M., Mazzucotelli, E., Marone, D., Crosatti, C., Michelotti, V., Vale, G. et al. (2018) Regulation and evolution of NLR genes: a close interconnection for plant immunity. *International Journal of Molecular Sciences*, 19, 1662.

Bowen, A.J., Gonzalez, D., Mullins, J.G., Bhatt, A.M., Martinez, A. & Conlan, R.S. (2010) PAH-domain-specific interactions of the

Arabidopsis transcription coregulator SIN3-LIKE1 (SNL1) with telomere-binding protein 1 and ALWAYS EARLY2 Myb-DNA binding factors. *Journal of Molecular Biology*, 395, 937–949.

Broglie, K.E., Butler, K.H., Da Silva Conceição, A., Frey, T.J., Hawk, J.A., Multani, D.S. et al. (2009) *Polynucleotides and methods for making plants resistant to fungal pathogens*. Google Patents. Patent WO2006107931A3.

Buchfink, B., Xie, C. & Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12, 59–60.

Cai, X., Chang, L., Zhang, T., Chen, H., Zhang, L., Lin, R. et al. (2021) Impacts of allopolyploidization and structural variation on intraspecific diversification in *Brassica rapa*. *Genome Biology*, 22, 166.

Cesari, S. (2018) Multiple strategies for pathogen perception by plant immune receptors. *New Phytologist*, 219, 17–24.

Cesari, S., Bernoux, M., Moncuquet, P., Kroj, T. & Dodds, P.N. (2014) A novel conserved mechanism for plant NLR protein pairs: the "integrated decoy" hypothesis. *Frontiers in Plant Science*, 5, 606.

Chen, G., Zhang, B., Ding, J., Wang, H., Deng, C., Wang, J. et al. (2022) Cloning southern corn rust resistant gene *RppK* and its cognate gene *AvrRppK* from *Puccinia polysora*. *Nature Communications*, 13, 4392.

Cheng, X., Peng, J., Ma, J., Tang, Y., Chen, R., Mysore, K.S. et al. (2012) NO APICAL MERISTEM (MtNAM) regulates floral organ identity and lateral organ separation in *Medicago truncatula*. *New Phytologist*, 195, 71–84.

Collins, N., Drake, J., Ayliffe, M., Sun, Q., Ellis, J., Hulbert, S. et al. (1999) Molecular characterization of the maize *Rp1-D* rust resistance haplotype and its mutants. *The Plant Cell*, 11, 1365–1376.

Deng, C., Leonard, A., Cahill, J., Lv, M., Li, Y., Thatcher, S. et al. (2022) The RppC-AvrRppC NLR-effector interaction mediates the resistance to southern corn rust in maize. *Molecular Plant*, 15, 904–912.

Dolatabadian, A., Bayer, P.E., Tirnaz, S., Hurgobin, B., Edwards, D. & Batley, J. (2020) Characterization of disease resistance genes in the *Brassica napus* pangenome reveals significant structural variation. *Plant Biotechnology Journal*, 18, 969–982.

Eddy, S.R. (1995) Multiple alignment using hidden Markov models. *Proceedings 16th International Conference on Intelligent Systems for Molecular Biology*, 3, 114–120.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792–1797.

Ekseth, O.K., Kuiper, M. & Mironov, V. (2014) orthAgogue: an agile tool for the rapid prediction of orthology relations. *Bioinformatics*, 30, 734–736.

Fu, Y., Zhang, Y., Mason, A.S., Lin, B., Zhang, D., Yu, H. et al. (2019) NBS-encoding genes in *Brassica napus* evolved rapidly after allopolyploidization and co-localize with known disease resistance loci. *Frontiers in Plant Science*, 10, 26.

Garg, G., Kamphuis, L.G., Bayer, P.E., Kaur, P., Dudchenko, O., Taylor, C.M. et al. (2022) A pan-genome and chromosome-length reference genome of narrow-leafed lupin (*Lupinus angustifolius*) reveals genomic diversity and insights into key industry and biological traits. *The Plant Journal*, 111, 1252–1266.

Golicz, A.A., Bayer, P.E., Barker, G.C., Edger, P.P., Kim, H., Martinez, P.A. et al. (2016) The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nature Communications*, 7, 13390.

Grund, E., Tremousaygue, D. & Deslandes, L. (2019) Plant NLRs with integrated domains: unity makes strength. *Plant Physiology*, 179, 1227–1235.

Hufford, M.B., Seetharam, A.S., Woodhouse, M.R., Chougule, K.M., Ou, S., Liu, J. et al. (2021) De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science*, 373, 655–662.

Irieda, H., Inoue, Y., Mori, M., Yamada, K., Oshikawa, Y., Saitoh, H. et al. (2019) Conserved fungal effector suppresses PAMP-triggered immunity by targeting plant immune kinases. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 496–505.

Jacob, F., Vernaldi, S. & Maekawa, T. (2013) Evolution and conservation of plant NLR functions. *Frontiers in Immunology*, 4, 297.

Jayakodi, M., Padmarasu, S., Haberer, G., Bonthala, V.S., Gundlach, H., Monat, C. et al. (2020) The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature*, 588, 284–289.

Jin, M., Lee, S.S., Ke, L., Kim, J.S., Seo, M.S., Sohn, S.H. et al. (2014) Identification and mapping of a novel dominant resistance gene, *TuRB07* to *Turnip mosaic virus* in *Brassica rapa*. *Theoretical and Applied Genetics*, 127, 509–519.

Jones, J.D. & Dangl, J.L. (2006) The plant immune system. *Nature*, 444, 323–329.

Jupe, F., Witek, K., Verweij, W., Sliwka, J., Pritchard, L., Etherington, G.J. et al. (2013) Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *The Plant Journal*, 76, 530–544.

Kato, T., Hatakeyama, K., Fukino, N. & Matsumoto, S. (2013) Fine mapping of the clubroot resistance gene *CRb* and development of a useful selectable marker in *Brassica rapa*. *Breeding Science*, 63, 116–124.

Kroj, T., Chanclud, E., Michel-Romiti, C., Grand, X. & Morel, J.B. (2016) Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytologist*, 210, 618–626.

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35, 1547–1549.

Lai, S., Safaei, J. & Pelech, S. (2016) Evolutionary ancestry of eukaryotic protein kinases and choline kinases. *Journal of Biological Chemistry*, 291, 5199–5205.

Lee, H.Y., Mang, H., Choi, E., Seo, Y.E., Kim, M.S., Oh, S. et al. (2021) Genome-wide functional analysis of hot pepper immune receptors reveals an autonomous NLR clade in seed plants. *New Phytologist*, 229, 532–547.

Leister, D. (2004) Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene. *Trends in Genetics*, 20, 116–122.

Lennon, J.R., Krakowsky, M., Goodman, M., Flint-Garcia, S. & Balint-Kurti, P.J. (2016) Identification of alleles conferring resistance to gray leaf spot in maize derived from its wild progenitor species teosinte. *Crop Science*, 56, 209–218.

Li, Q., Jiang, X.M. & Shao, Z.Q. (2021) Genome-wide analysis of NLR disease resistance genes in an updated reference genome of barley. *Frontiers in Genetics*, 12, 694682.

Li, Q., Li, J., Sun, J.L., Ma, X.F., Wang, T.T., Berkey, R. et al. (2016) Multiple evolutionary events involved in maintaining homologs of resistance to *powdery mildew 8* in *Brassica napus*. *Frontiers in Plant Science*, 7, 1065.

Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S. et al. (2020) Pan-genome of wild and cultivated soybeans. *Cell*, 182, 162–176 e13.

Liu, M.H., Kang, H., Xu, Y., Peng, Y., Wang, D., Gao, L. et al. (2020) Genome-wide association study identifies an NLR gene that confers partial resistance to *Magnaporthe oryzae* in rice. *Plant Biotechnology Journal*, 18, 1376–1383.

Liu, Y., Zhang, X., Yuan, G., Wang, D., Zheng, Y., Ma, M. et al. (2021) A designer rice NLR immune receptor confers resistance to the rice blast fungus carrying noncorresponding avirulence effectors. *Proceedings of the National Academy of Sciences of the United States of America*, 118, e2110751118.

Love, M.I., Huber, W. & Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 550.

Mammadov, J., Buyyarapu, R., Guttikonda, S.K., Parliament, K., Abdurakhmonov, I.Y. & Kumpatla, S.P. (2018) Wild relatives of maize, rice, cotton, and soybean: treasure troves for tolerance to biotic and abiotic stresses. *Frontiers in Plant Science*, 9, 886.

Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez, G.J., Buckler, E. & Doebley, J. (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 6080–6084.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H. & Michelmore, R.W. (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *The Plant Cell*, 15, 809–834.

Meyers, B.C., Shen, K.A., Rohani, P., Gaut, B.S. & Michelmore, R.W. (1998) Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. *The Plant Cell*, 10, 1833–1846.

Michelmore, R.W. & Meyers, B.C. (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research*, 8, 1113–1130.

Monteiro, F. & Nishimura, M.T. (2018) Structural, functional, and genomic diversity of plant NLR proteins: an evolved resource for rational engineering of plant immunity. *Annual Review of Phytopathology*, 56, 243–267.

Mueller, D.S. (2016) Corn yield loss estimates due to diseases in the United States and Ontario, Canada from 2012 to 2015. *Plant Health Progress*, 17, 12.

Munoz-Amatriain, M., Eichten, S.R., Wicker, T., Richmond, T.A., Mascher, M., Steuernagel, B. et al. (2013) Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome. *Genome Biology*, 14, R58.

Newman, M.A., Sundelin, T., Nielsen, J.T. & Erbs, G. (2013) MAMP (microbe-associated molecular pattern) triggered immunity in plants. *Frontiers in Plant Science*, 4, 139.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. & Kingsford, C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14, 417–419.

Prigozhin, D.M. & Krasileva, K.V. (2021) Analysis of intraspecies diversity reveals a subset of highly variable plant immune receptors and predicts their binding sites. *The Plant Cell*, 33, 998–1015.

Qi, D., Deyoung, B.J. & Innes, R.W. (2012) Structure-function analysis of the coiled-coil and leucine-rich repeat domains of the RPS5 disease resistance protein. *Plant Physiology*, 158, 1819–1832.

Radwan, O., Gandhi, S., Heesacker, A., Whitaker, B., Taylor, C., Plocik, A. et al. (2008) Genetic diversity and genomic distribution of homologs encoding NBS-LRR disease resistance proteins in sunflower. *Molecular Genetics and Genomics*, 280, 111–125.

Saile, S.C., Jacob, P., Castel, B., Jubic, L.M., Salas-Gonzales, I., Backer, M. et al. (2020) Two unequally redundant "helper" immune receptor families mediate *Arabidopsis thaliana* intracellular "sensor" immune receptor functions. *PLoS Biology*, 18, e3000783.

Saintenac, C., Zhang, W., Salcedo, A., Rouse, M.N., Trick, H.N., Akhunov, E. et al. (2013) Identification of wheat gene *Sr35* that confers resistance to Ug99 stem rust race group. *Science*, 341, 783–786.

Saitou, N. & Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4, 406–425.

Sanchez-Vallet, A., Fouche, S., Fudal, I., Hartmann, F.E., Soyer, J.L., Tellier, A. et al. (2018) The genome biology of effector gene evolution in filamentous plant pathogens. *Annual Review of Phytopathology*, 56, 21–40.

Sarris, P.F., Cevik, V., Dagdas, G., Jones, J.D. & Krasileva, K.V. (2016) Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biology*, 14, 8.

Shan, L., He, P., Li, J., Heese, A., Peck, S.C., Nurnberger, T. et al. (2008) Bacterial effectors target the common signaling partner BAK1 to disrupt multiple MAMP receptor-signaling complexes and impede plant immunity. *Cell Host & Microbe*, 4, 17–27.

Shang, L., Li, X., He, H., Yuan, Q., Song, Y., Wei, Z. et al. (2022) A super pan-genomic landscape of rice. *Cell Research*, 32, 878–896.

Shao, Z.Q., Xue, J.Y., Wu, P., Zhang, Y.M., Wu, Y., Hang, Y.Y. et al. (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, 2095–2109.

Stein, J.C., Yu, Y., Copetti, D., Zwickl, D.J., Zhang, L., Zhang, C. et al. (2018) Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nature Genetics*, 50, 285–296.

Steuernagel, B., Witek, K., Krattinger, S.G., Ramirez-Gonzalez, R.H., Schoonbeek, H.J., Yu, G. et al. (2020) The NLR-annotator tool enables annotation of the intracellular immune receptor repertoire. *Plant Physiology*, 183, 468–482.

Svitashev, S., Schwartz, C., Lenderts, B., Young, J.K. & Mark Cigan, A. (2016) Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. *Nature Communications*, 7, 13274.

Tang, D., Wang, G. & Zhou, J.M. (2017) Receptor kinases in plant–pathogen interactions: more than pattern recognition. *The Plant Cell*, 29, 618–637.

Thatcher, S., Leonard, A., Lauer, M., Panangipalli, G., Norman, B., Hou, Z. et al. (2022) The northern corn leaf blight resistance gene *Ht1* encodes an nucleotide-binding, leucine-rich repeat immune receptor. *Molecular Plant Pathology*, 24, 758–767.

Van de Weyer, A.L., Monteiro, F., Furzer, O.J., Nishimura, M.T., Cevik, V., Witek, K. et al. (2019) A species-wide inventory of NLR genes and alleles in *Arabidopsis thaliana*. *Cell*, 178, 1260–1272 e14.

van der Hoorn, R.A. & Kamoun, S. (2008) From guard to decoy: a new model for perception of plant pathogen effectors. *The Plant Cell*, 20, 2009–2017.

van Wersch, S. & Li, X. (2019) Stronger when together: clustering of plant NLR disease resistance genes. *Trends in Plant Science*, 24, 688–699.

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M.T., Brinton, J. et al. (2020) Multiple wheat genomes reveal global variation in modern breeding. *Nature*, 588, 277–283.

Wang, W., Chen, L., Fengler, K., Bolar, J., Llaca, V., Wang, X. et al. (2021) A giant NLR gene confers broad-spectrum resistance to *Phytophthora sojae* in soybean. *Nature Communications*, 12, 6263.

Wei, F., Gobelman-Werner, K., Morroll, S.M., Kurth, J., Mao, L., Wing, R. et al. (1999) The *Mla* (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. *Genetics*, 153, 1929–1948.

Wu, C.H., Abd-El-Haliem, A., Bozkurt, T.O., Belhaj, K., Terauchi, R., Vossen, J.H. et al. (2017) NLR network mediates immunity to diverse plant pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 8113–8118.

Yang, C.J., Samayoa, L.F., Bradbury, P.J., Olukolu, B.A., Xue, W., York, A.M. et al. (2019) The genetic architecture of teosinte catalyzed and constrained maize domestication. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 5643–5652.

Yang, P., Scheuermann, D., Kessel, B., Koller, T., Greenwood, J.R., Hurni, S. et al. (2021) Alleles of a wall-associated kinase gene account for three of the major northern corn leaf blight resistance loci in maize. *The Plant Journal*, 106, 526–535.

Yin, K. & Qiu, J.L. (2019) Genome editing for plant disease resistance: applications and perspectives. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 374, 20180322.

Yu, J., Holland, J.B., Mcmullen, M.D. & Buckler, E.S. (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics*, 178, 539–551.

Zhang, Y., Edwards, D. & Batley, J. (2021) Comparison and evolutionary analysis of brassica nucleotide binding site leucine rich repeat (NLR) genes and importance for disease resistance breeding. *Plant Genome*, 14, e20060.

Zhang, J., Li, W., Xiang, T., Liu, Z., Laluk, K., Ding, X. et al. (2010) Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a *Pseudomonas syringae* effector. *Cell Host & Microbe*, 7, 290–301.

Zhao, J., Bayer, P.E., Ruperao, P., Saxena, R.K., Khan, A.W., Golicz, A.A. et al. (2020) Trait associations in the pangenome of pigeon pea (*Cajanus cajan*). *Plant Biotechnology Journal*, 18, 1946–1954.

Zipfel, C. (2014) Plant pattern-recognition receptors. *Trends in Immunology*, 35, 345–351.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.