Conifers Concentrate Large Numbers of NLR Immune Receptor Genes on One Chromosome

Yannick Woudstra [1], Hayley Tumas [1], Cyril van Ghelder[2], Tin Hang Hung [1], Joana J. Ilska[3], Sebastien Girardi[4,5], Stuart A'Hara[6], Paul McLean[6], Joan Cottrell[6], Joerg Bohlmann[7], Jean Bousquet[4], Inanc Birol [8], John A. Woolliams[3], and John J. MacKay[1,*]

[1]Department of Biology, University of Oxford, Oxford OX1 3RB, UK

[2]INRAE, Université Côte d'Azur, CNRS, ISA, Sophia Antipolis 06903, France

[3]The Roslin Institute, Royal (Dick) School of Veterinary Science, University of Edinburgh, Easter Bush, Midlothian EH25 9RG, UK

[4]Canada Research Chair in Forest Genomics, Forest Research Centre, Université Laval, Québec, QC, Canada G1V 0A6

[5]Institute for Systems and Integrative Biology, Université Laval, Québec, QC, Canada GIV 0A6

[6]Forest Research, Northern Research Station, Roslin, Midlothian EH25 9SY, UK

[7]Michael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada V6T 1Z4

[8]Canada's Michael Smith Genome Sciences Centre, Vancouver, BC, Canada V5Z 4S6

*Corresponding author: E-mail: john.mackay@biology.ox.ac.uk.

Accepted: May 21, 2024

Abstract

Nucleotide-binding domain and leucine-rich repeat (NLR) immune receptor genes form a major line of defense in plants, acting in both pathogen recognition and resistance machinery activation. NLRs are reported to form large gene clusters in limber pine (*Pinus flexilis*), but it is unknown how widespread this genomic architecture may be among the extant species of conifers (Pinophyta). We used comparative genomic analyses to assess patterns in the abundance, diversity, and genomic distribution of NLR genes. Chromosome-level whole genome assemblies and high-density linkage maps in the Pinaceae, Cupressaceae, Taxaceae, and other gymnosperms were scanned for NLR genes using existing and customized pipelines. The discovered genes were mapped across chromosomes and linkage groups and analyzed phylogenetically for evolutionary history. Conifer genomes are characterized by dense clusters of NLR genes, highly localized on one chromosome. These clusters are rich in TNL-encoding genes, which seem to have formed through multiple tandem duplication events. In contrast to angiosperms and nonconiferous gymnosperms, genomic clustering of NLR genes is ubiquitous in conifers. NLR-dense genomic regions are likely to influence a large part of the plant's resistance, informing our understanding of adaptation to biotic stress and the development of genetic resources through breeding.

Key words: resistance genes, NBS-LRR, genomic architecture, comparative genomics, gene family evolution, gene clusters.

Significance

NLR immune receptor genes are important in pest, disease, and drought resistance of plants. In the giga-genomes of conifers, they concentrate on very small chromosomal regions. These regions act as important reservoirs for NLR diversity and can be used in breeding to improve the resilience of conifer trees.

Introduction

Disease resistance is one of the key areas of study in plant genetics and evolution with implications for conservation and ecosystem health, as well as breeding. Decades of research have improved our understanding of the identity and interplay of major gene families involved in disease resistance with functions in detection and defense

(Ngou et al. 2022). One of the first events in plant defense mechanisms is pathogen recognition, in which the nucleotide-binding domain and leucine-rich repeat (NBS-LRR or NLR) immune receptor gene family plays a central role (Duxbury et al. 2021). The products of NLR genes occur intracellularly and can bind directly to specific pathogenic effectors (pathogen-encoded proteins) or detect modifications of plant proteins induced by such effectors, thus activating a cascade of defense mechanisms upon perception (Ngou et al. 2022). NLRs provide a typical example of an evolutionary arms race in which pathogenic effectors evolve to evade detection by host NLRs, which, in turn, evolve to recognize the new variants (Chen et al. 2023; Chia and Carella 2023). NLRs are therefore diverse and abundant in many plant species with several hundred different NLR genes found in a range of land plant lineages (Barragan and Weigel 2021). While NLRs have been studied extensively in angiosperms (i.e., Cucurbitaceae [Lin et al. 2013], Rosaceae [Jia et al. 2015], and Solanaceae [Seo et al. 2016]; Angiosperm NLR Atlas [Liu et al. 2021c]), studies on conifers are rare (Liu et al. 2019; Van Ghelder et al. 2019; Ence et al. 2022).

NLRs exhibit a conserved tripartite structure consisting of the following: (i) a nonconserved *N*-terminal domain; (ii) a conserved central nucleotide-binding domain (NB-ARC, defined as "a nucleotide-binding adaptor shared by APAF-1, certain *R* gene products, and CED-4" (van der Biezen and Jones 1998)); and (iii) a C-terminal leucine-rich repeat (LRR) domain that can vary in length. These resistance genes seem to have originated before the rise of land plants (Shao et al. 2019) and have since diversified into three main classes based on the character of the *N*-terminal domain: CNL, RNL, and TNL. CNLs (*N*-terminal "coiled-coil" domain) and RNLs (*N*-terminal "resistance to powdery mildew 8 (RPW8) domain") are closely related classes, whereas TNLs (*N*-terminal "Toll interleukin-1 receptor" domain) form a distinct class. All three classes are unusually abundant and diverse in conifers, with an RNL diversity that is distinctly higher than in any other group of land plants (Van Ghelder et al. 2019). Furthermore, a genomic distribution analysis in limber pine (*Pinus flexilis* E.James) revealed an unbalanced intragenomic and intrachromosomal distribution of NLRs (Liu et al. 2019). In this species, one chromosome contained a dense cluster of NLRs comprising mainly TNLs, indicating a high rate of tandem duplications. Besides disease resistance, NLRs have been shown to be responsive to drought stress when investigated in the conifer white spruce (*Picea glauca* (Moench) Voss) (Van Ghelder et al. 2019). NLR-rich genomic regions are therefore particularly interesting candidates for genomic breeding purposes and genetic resource management in conifers, especially under climate change and the spread of tree pests and diseases.

Conifers are found in diverse ecosystems and several species are used in productive forestry across the globe, some of which involves breeding programs (Mullin and Lee 2013). Breeding in conifers traditionally relies on pedigree analysis and phenotypic evaluations such as growth rates, wood yield, and properties, and susceptibility to biotic and abiotic threats (White et al. 2007). Emerging and intensifying threats to the health of conifers in natural populations and managed forests involve a range of biotic stressors including oomycetes, fungi, herbivorous insects, and nematodes (Mota et al. 1999; Mitton and Ferrenberg 2012; Brar et al. 2018; Jakoby et al. 2019). In response to these challenges, molecular tools and genomic resources are being developed to support both fundamental research and diverse applications such as an acceleration of breeding outputs (Neale and Kremer 2011; Stocks et al. 2019; Bousquet et al. 2021). For example, investigations have linked genetic resistance to fusiform rusts in *Pinus taeda* L. to TNL-encoding sequences and to genomic clustering of resistance genes (Wilcox et al. 1996; Quesada et al. 2014; Ence et al. 2022). Genomic selection methods may have the potential to be developed to enhance resistance, such as rust resistance in *P. taeda* L. (Ence et al. 2022) and insect resistance in Norway spruce (*Picea abies* (L.) H.Karst.) (Lenz et al. 2020). However, the success of this relies on a more complete understanding of the genomic architecture of conifer NLRs (Ence et al. 2022).

The large size (often ≥10 Gb) and relative complexity of conifer nuclear genomes has challenged whole genome sequence assembly (e.g. Birol et al. 2013; Nystedt et al. 2013; Zimin et al. 2014), but new methods have greatly improved the contiguity of genome assemblies as seen in *Sequoiadendron giganteum* (Lindl.) J.Buchholz (Scott et al. 2020), *Taxus chinensis* (Pilg.) Rehder (Xiong et al. 2021), and *Pinus tabuliformis* Carrière (Niu et al. 2022). In some species that lack such assemblies, high-density genetic maps are available to probe genome architecture (Bernhardsson et al. 2019; Liu et al. 2019; Gagalova et al. 2022). Together, these assembled genome sequences and genetic maps, along with diverse transcriptomes (e.g. Van Ghelder et al. 2019), open the doors to more comprehensive analyses of resistance genes and their genomic architecture in conifer trees. Considering the relatively high level of genome conservation across conifers, we may predict that genomic clustering as observed in *P. flexilis* (Liu et al. 2019) will occur across conifer taxa, that is, to the extent they result from shared ancestral evolutionary events.

Considering the potential benefits of genomic breeding with NLR dense genome segments for conservation and industry, we investigated NLR gene clustering patterns across conifers. We leveraged recently published diploid high-density linkage maps and chromosome-level whole genome assemblies for genomic mapping of NLR genes. The results for conifers were contrasted with nonconifer gymnosperms from the Ginkgoales and Cycadales. To elucidate

the evolutionary trajectories toward the observed clustering patterns, NLR genes were analyzed in a phylogenetic framework. Our results indicate consistently uneven genomic distribution patterns of NLR genes across all conifers, with large and heavily concentrated reservoirs of NLR genes located on specific chromosomes. This knowledge on resistance genes will be informative for both breeding and conservation of these economically and ecologically important trees.

Results

NLR immune receptor gene diversity and abundance were annotated in the genomes of six conifer species (members of the Pinaceae, Cupressaceae, and Taxaceae) and two other gymnosperms (*Ginkgo biloba* (Ginkgoales) and *Cycas panzhihuaensis* (Cycadales)), with varying levels of contiguity. We deployed an automated NLR annotation pipeline and additional manual BLAST procedures on recently published high-density linkage maps and chromosome-level whole genome assemblies to physically map the genomic distribution of NLR genes in gymnosperms. We discovered consistent patterns of genomic clustering of NLR genes among conifers, but not in other gymnosperms. In each of the analyzed conifers, a particularly dense cluster of NLR genes occurred on a single chromosome that contained between 18% and 34% of the total number of NLR genes within only a short segment of a few Mb (or cM in the case of linkage maps).

We restricted our analysis to diploid genomes to avoid potential issues of false discovery for duplicated genes, due to incomplete phasing of the haplotypes in polyploid genome assemblies. We therefore omitted the genome assembly of the hexaploid *Sequoia sempervirens* Endl. (Neale et al. 2022) from our analysis.

Our scan of genome assemblies and high-density linkage maps for conifers and other gymnosperms did include the recently published mega-genome assembly (25.4 Gb) of the diploid *P. tabuliformis* (Niu et al. 2022), where some chromosomes surpass 2 Gb in length. Unfortunately, the NLR Annotator pipeline (Steuernagel et al. 2020) is currently insufficiently programmed for such large contigs, limiting the output of our NLR analysis on this genome assembly. Furthermore, our preliminary results indicated extremely high NLR numbers (>4,000) in this assembly, which we hypothesize to be an overestimation, considering the large disjunction with other conifers (1,002 in *S. giganteum*) and members of the *Pinus* genus (639 in *P. flexilis*). We therefore omitted *P. tabuliformis* from further analysis.

Gymnosperm NLR Abundance and Diversity

The number of NLR genes discovered in conifer genomic datasets varied nearly 2-fold (Fig. 1), between 533 (*T. chinensis*) and 1,002 (*S. giganteum*). Of the nonconifers, *G. biloba* contained 585 NLR genes, which is similar to the average found in conifers (Table 1). By contrast, there were only 136 NLR genes found in the Cycad genome. Of the conifers, the *S. giganteum* genome contained the highest number of NLR genes (1,002) followed by (*P. flexilis*), which had considerably less genes (639). In all conifers, the TNL class was the most abundant, contributing between 33% (81/245, *Picea sitchensis*) to 59% (376/639, *P. flexilis*) to the total number of NLR genes (Table 1). By contrast, the CNL class was the most abundant in *G. biloba*, in which it contributed 43% to the total number of NLR genes. RNLs were more abundant in conifers than in other gymnosperms. This class was proportionally most abundant in *Picea* species (16% to 20%), with half of the RNLs in *P. glauca* and *P. sitchensis* consisting only of RPW8 domains. The number of unclassified NLR genes (missing or ambiguous C-terminal domain) grew proportionally with the total number of NLR genes, representing 16% to 25% of the genes in the complete genomic datasets.

Chromosomal and Intrachromosomal NLR Distributions

In each of the conifer species we analyzed, one chromosome displayed a disproportionately high NLR content (Table 1 and Fig. 2A), a phenomenon that was absent in nonconiferous gymnosperms. In conifers, the chromosome that displayed the highest clustering of NLR genes contained between 29% (*T. chinensis*) and 42% (*P. flexilis*) of the total number of NLR genes found in the respective genomes (Table 1). Even when there was incomplete genome coverage, the results for *Picea* species were within this range, the only outlier being *P. sitchensis* at 23%. Large numbers of NLR genes form dense clusters on these NLR-rich chromosomes or linkage groups (Fig. 2B to E). These clusters contain high proportions of the total amount of NLR genes found in the genomic dataset, with up to 34% (*P. flexilis*, Fig. 2C) of NLR genes concentrated in the space of 21% (44 cM) of a chromosome (linkage group).

Although the distribution of NLR genes was nonrandom in all gymnosperms analyzed, it was considerably less clustered in *C. panzhihuaensis* and *G. biloba* compared with the full genomic datasets of the conifers we analyzed (Table 1). In the non-conifer species, only one small cluster of 13 genes (10% of the total) was found on Chromosome #10 in *C. panzhihuaensis* (Fig. 3) and a small cluster of 45 genes (7.9% of the total) occurred on Chromosome #3 in *G. biloba*, both proportionally smaller than the clusters in conifer species (on average 25% of the total). Nonuniform distribution of NLR genes in conifers not only occurs between but also within chromosomes (Fig. 3) with large regions of NLR-rich chromosomes being devoid of NLR genes, particularly in *T. chinensis*.

Intragenomic Diversification and Evolution of NLR Genes

Phylogenetic relationships between intragenomic NLR genes followed expected class distinctions in all conifers

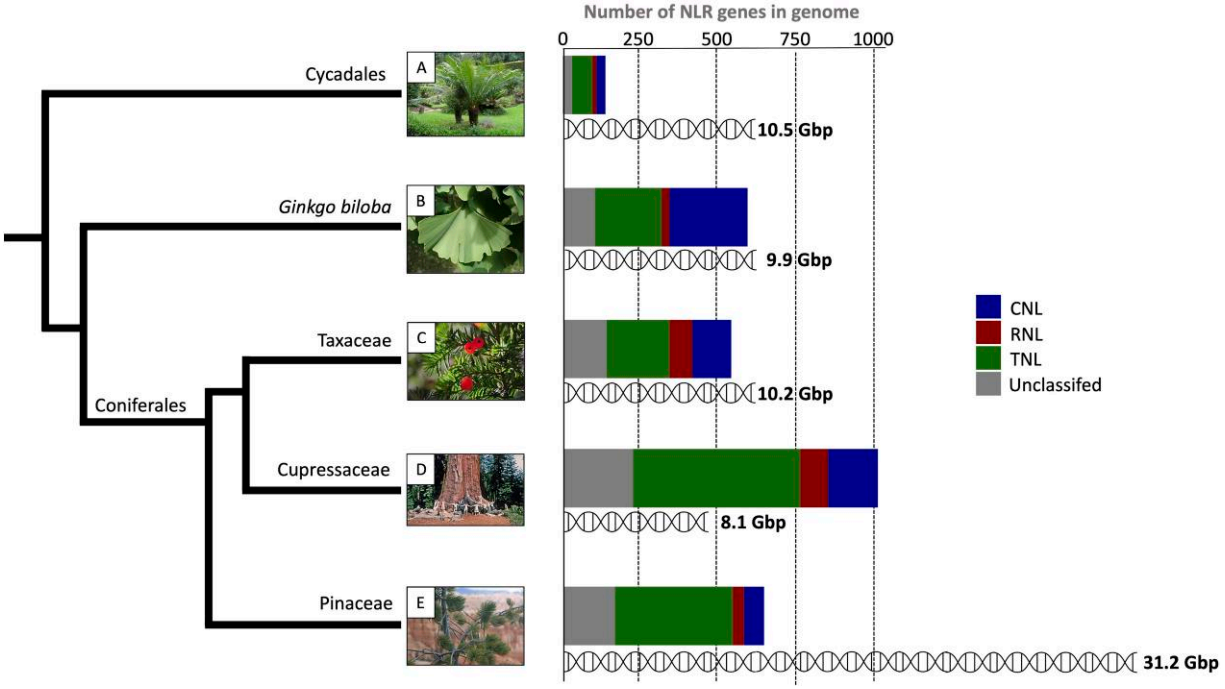

Fig. 1.—Overview of main gymnosperm clades, with total number of NLR genes and their corresponding categories indicated as bar charts, as found in this study (Table 1). Cladogram is based on the phylogeny presented in Leslie et al. (2018). Genome sizes are indicated beneath each bar chart and are based on the chromosome-level assemblies used in this study (see "Genomic Distribution of NLR Genes in Pinaceae and Other Conifer Families" section) or, in the case of Pinaceae (*P. flexilis*), obtained from the Kew Plant DNA C-values database (release 7.1, Pellicer and Leitch 2020). Gymnosperms invariably have large genomes (~10 Gb), but display large variations in NLR gene numbers. Despite the 3-fold increase in genome size observed in Pinaceae, the number of discovered NLR genes remained within the average range of conifers. Pictures were obtained from the Wikimedia Commons repository (https://commons.wikimedia.org) and correspond to the broader taxonomic clades: A—*Cycas rumphii* Miq., Andy King; B—*Ginkgo biloba* L., Susanna Giaccai; C—*Taxus baccata* L., Mykola Swarnyk; D—*Sequoiadendron giganteum*, W. Bulach; E—*Pinus flexilis*, Greg Woodhouse.

Table 1

Number of NLRs discovered in each taxon of this study, separated by class

Species	Total #NLR (on chr.)	#CNL	#RNL (RPW8 only)	#TNL	#Uncl.	Densest chr. (% of total)	#Observed/ #predicted[a]	*P*-value of distribution
Cycas panzhihuaensis°	136 (127)	29	12 (8)	65	30	24 (19%)	2.02	0.02958
Ginkgo biloba	585 (570)	236	23 (5)	221	105	83 (14%)	2.04	6.701e−08
Taxaceae								
Taxus chinensis	533 (496)	129	68 (29)	201	135	146 (29%)	4.30	<2.2e−16
Cupressaceae								
Sequoiadendron giganteum°	1002	161	83 (17)	534	224	312 (31%)	2.47	<2.2e−16
Pinaceae								
Pinus flexilis[b]	639	63	21 (–)	375	180	266 (42%)	3.55	<2.2e−16
Picea abies[c]	173	38	29 (9)	85	21	68 (39%)	3.37	4.779e−05
Picea glauca[c]	273	76	41 (21)	105	51	89 (33%)	3.26	1.072e−07
Picea sitchensis[c]	245	73	50 (26)	81	41	57 (23%)	2.61	0.003825

The number of NLRs mapped onto chromosomes is recorded separately for *Ginkgo biloba* and *T. chinensis* as some NLR genes were located on unassigned scaffolds. Some RNL genes only contained the RPW8 domain, and these are recorded separately for each taxon. All species have a chromosome count of $n = 12$ per haploid genome, except those indicated with ° ($n = 11$).

[a]Listed for densest chromosome only.

[b]*P. flexilis* results are based on the results from Liu et al (2019).

[c]The results for *Picea* taxa are based on NLR annotation of high-density linkage maps or incomplete genome assemblies that were built using high-density linkage maps, details in the Materials and Methods section.

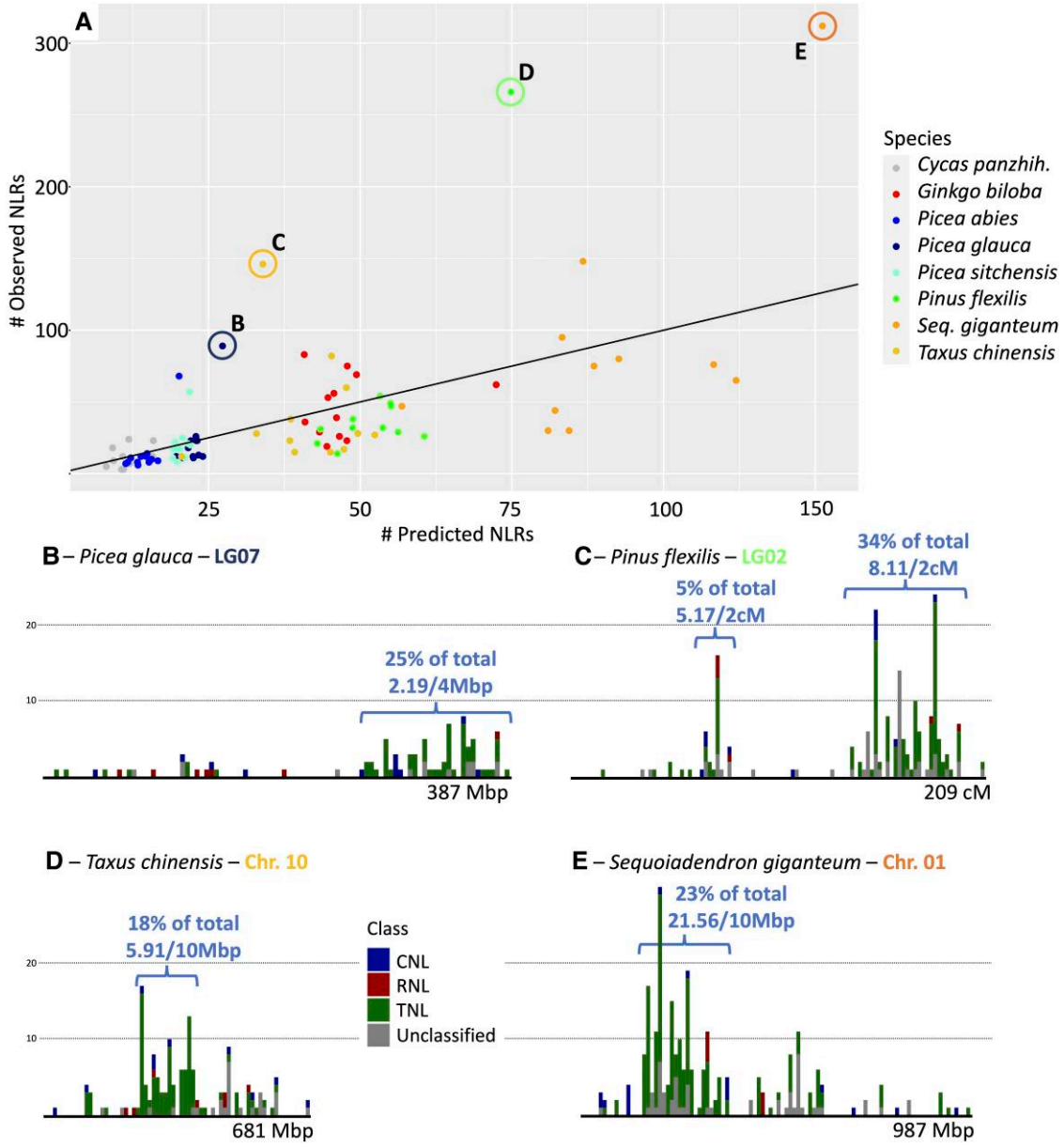

Fig. 2.—Chromosomal distribution of NLR genes across gymnosperms. A) Scatter plot indicating the observed number of NLR genes versus the expected number of NLR genes (calculated based on the length of the chromosome and the total number of NLR genes discovered in the genome; see "Genomic Distribution of NLR Genes in Pinaceae and Other Conifer Families" section) for each chromosome in each taxon analyzed in this study. The black line follows the function $y = x$, indicating a perfectly homogeneous distribution of NLR genes over the chromosomes. Deviations from this line therefore indicate a non-homogeneous distribution. Highly deviant chromosomes of four taxa are highlighted and have their intrachromosomal NLR distribution displayed in histograms (B–E). Bin width equals ±1% of the length of the largest chromosome in the genome of the respective taxon (see "Genomic Distribution of NLR Genes in Pinaceae and Other Conifer Families" section). The colors indicate NLR class as determined with the NLR Annotator (Steuernagel et al. 2020) and manual BLASTs (see "NLR Classification" section). Ultra-dense NLR clusters are indicated for each taxon.

and *C. panzhihuaensis* but not in *G. biloba* (Fig. 4). Although conifer TNLs show a strong overall monophyletic correlation, *T. chinensis* (Fig. 4D) is the only gymnosperm where all TNLs share one most recent common ancestor (MRCA). *S. giganteum* (Fig. 4C) even has a small monophyletic clade of TNLs nested within the larger CNL/RNL clade. In all noncycad gymnosperms, there is at least one monophyletic RNL clade with a different MRCA than the main CNL clade. All conifer CNL clades contained a few derived RNL sequences restricted to one phylogenetic subclade of the CNL clade. Both the *T. chinensis* and *S. giganteum* genomes contain a smaller monophyletic clade with a unique MRCA that contains almost exclusively unclassified NLR genes.

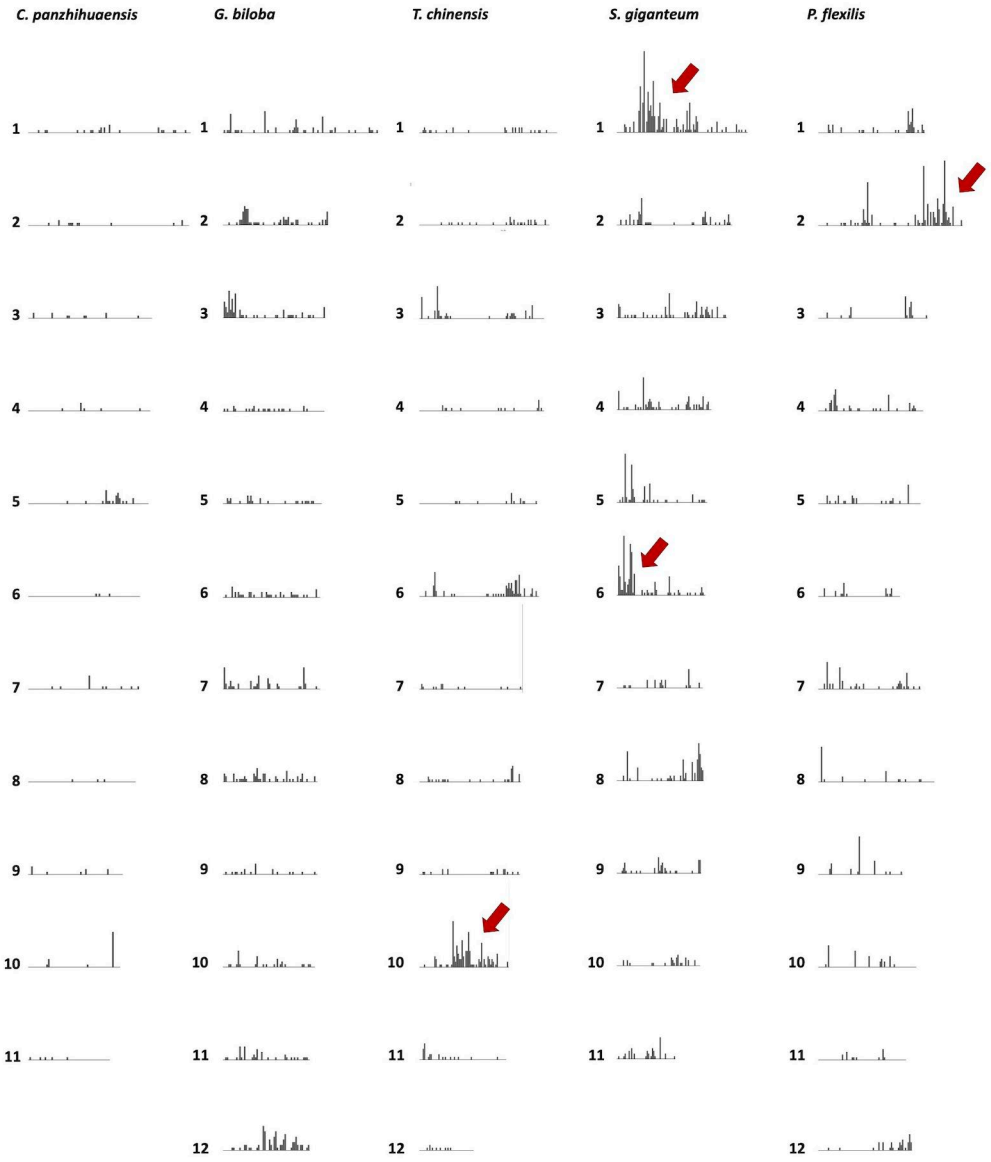

Fig. 3.—Histogram plots of NLR genes on each chromosome within the genome of five gymnosperm lineages. Chromosomes are ordered based on the ordering in the respective assembly and do not reflect synteny. Red arrows indicate particularly dense clusters.

The densest clusters in conifers are mainly composed of TNLs (Fig. 2B to E), which correlates with their intrachromosomal diversification (Fig. 4G). Most TNLs that occur on the same chromosome are also phylogenetically correlated. In the Pinaceae, this phylogenetically correlated diversification even occurred on the same syntenic linkage group (supplementary material figs. S1 to S4, Supplementary Material online).

Discussion

The study of NLR immune receptor evolution and genomic architecture in conifers is motivated in part by the increased threats from diverse biotic aggressors. The 615 species of extant conifers span all continents except Antarctica and are classified in eight families with the largest being Pinaceae, Cupressaceae, and Podocarpaceae (Farjon and Filer 2013). Forest trees, including conifers, have several co-evolved biotic aggressors ranging from rust diseases to herbivorous insects that may damage or even kill trees. However, more severe attacks, range expansions, or species introductions, and infection of previously unknown hosts have become increasingly prevalent and linked to climate change (Teshome et al. 2020). Major herbivorous insects are expanding their range and intensifying damage levels in conjunction with climate change, such as mountain

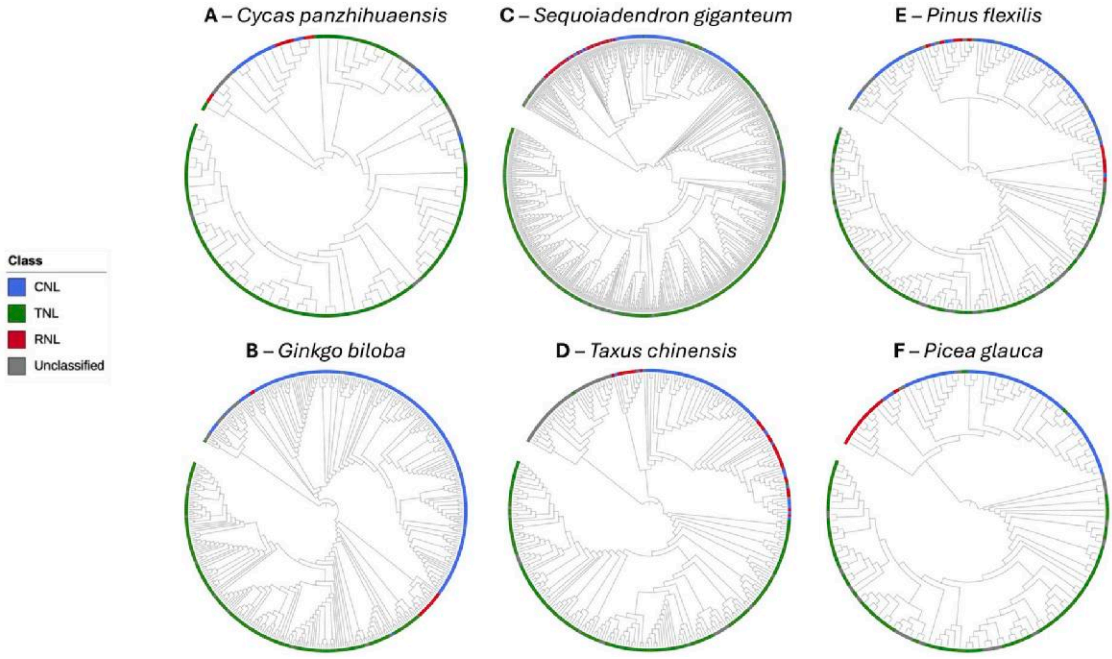

Fig. 4.—Intragenomic phylogenetic relationships of NLR genes based on the conserved central NB-ARC domain, calculated with maximum likelihood algorithms using IQTree v1.6.12 (Nguyen et al. 2015; (see "NLR Phylogenies" section) and annotated using the iTOL web server (Letunic and Bork 2021). The color strips around the circular trees indicate NLR class as determined with the NLR Annotator (Steuernagel et al. 2020) and manual BLASTs (see "NLR Classification" section). Main gymnosperm lineages represented by Cycadales (A), Ginkgoales (B), Cuppressaceae (C), Taxaceae (D), and Pinaceae: Pinus (E), and Picea (F).

pine beetles (e.g. Mitton and Ferrenberg 2012). New diseases or outbreaks in several conifers are being linked to Oomycetes, namely, species of *Phytophthora* de Barys (e.g. Brasier and Webber 2010; Brar et al. 2018). There is growing evidence of the involvement of NLRs in resistance, particularly of TNLs to rust diseases affecting pines (Quesada et al. 2014; Amerson et al. 2015; Ence et al. 2022), but much less is known when it comes to insects and newly emerging biotic threats. Understanding the genomic architecture of NLR genes will therefore be critical for informing breeding programs and other conservation practices seeking to mitigate the effects of climate change.

Ubiquitous Genomic Clustering of NLR Genes in Conifers

Through comparative genome-wide analysis of NLR immune receptor genes in conifers, we elucidated a conserved nonrandom chromosomal distribution of NLR genes. In all conifer genomes analyzed here we found one chromosome containing a disproportionately large number of NLR genes (Fig. 2). Without exception, these chromosomes contained dense clusters of NLR genes (Fig. 3), comprising on average a quarter of the total number of NLR genes in the genome (Fig. 2B to E). These clusters predominantly contain TNL genes, which share an ancestral origin (Figs. 4 and 5). High-density gene regions are a hallmark of conifer genomes (Pavy et al. 2017), hinting at tandem duplications, as detected frequently for NLR genes (Pavy et al. 2017) and for other conifer gene families (e.g. Guillet-Claude et al. 2004). We therefore conclude that the unbalanced genomic clustering pattern of NLR genes found previously in *P. flexilis* (Liu et al. 2019) occurs in a taxonomically broad range of other conifer species. The absence of this pattern in nonconiferous gymnosperms (Figs. 2A and 3) indicates that NLR gene clusters arose in the ancestors of conifers.

Diversification within these large groups of sequences appears to be variable and largely lineage-specific (such as in the Pinaceae). Partial functional redundancy is likely, in response to similar environmental cues and selection pressures among conifer taxa, thus contributing to the high abundance of NLR genes in conifers. These ancestral clusters may have enabled lineage-specific tandem duplications leading to the high (and variable) abundance of NLR genes observed in conifers (Van Ghelder et al. 2019; this study), as compared with nonconiferous plants such as cycads (Table 1 and Fig. 4A) and many angiosperms. A high diversity of NLR genes may lead to a high population versatility in drought and disease resistance for these woody perennials. These genes are likely to contribute to the ecological dominance of conifers in many boreal and

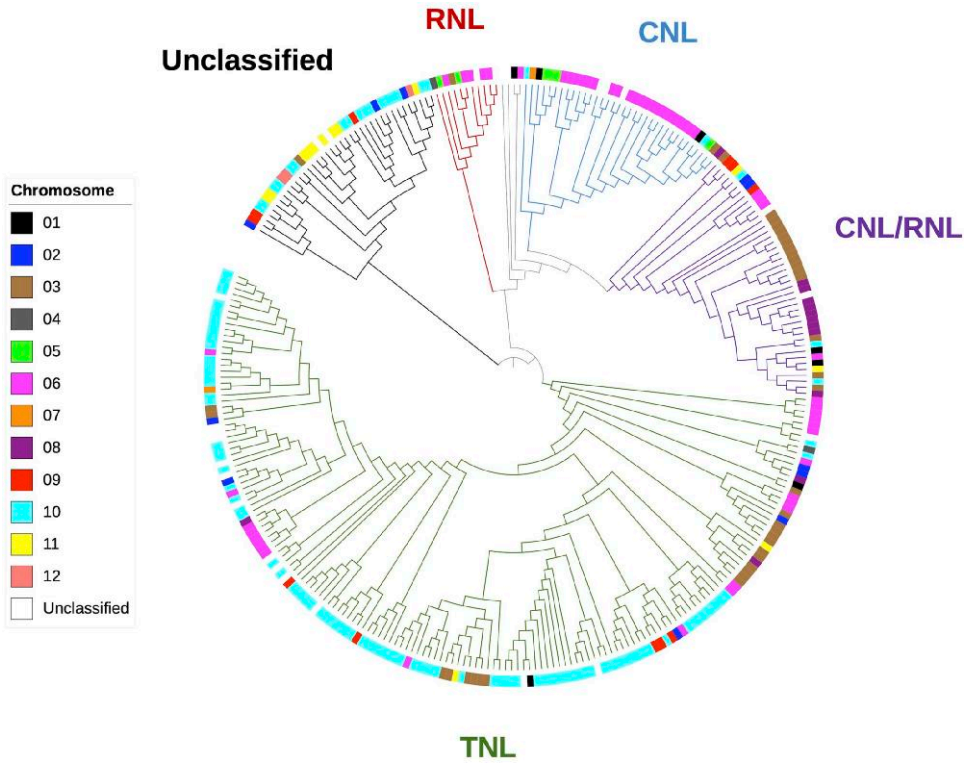

Fig. 5.—Chromosomal structuring in phylogenetic relationships within a conifer species (*T. chinensis*) displayed in an intragenomic framework. The branch colors correspond to the NLR gene subfamily, as indicated in the same colors next to the phylogeny. The colored squares at the tips of branches represent the chromosome on which the respective NLR genes are located. The NLR genes found on scaffolds that were not assembled into chromosomes are indicated with empty color squares ("Unclassified").

temperate forests (Bonello et al. 2006), including in highly inhospitable habitats (Laberge et al. 2000).

Genomic clusters of NLR genes have previously been found in a variety of angiosperm lineages (van Wersch and Li 2019), such as *Arabidopsis* Heinh. (Brassicaceae) (Meyers et al. 2003), lettuce (Asteraceae) (Christopoulou et al. 2015), peach (Rosaceae) (Verde et al. 2013), potatoes (Solanaceae) (Seo et al. 2016), and wheat (Poaceae) (Smith et al. 2007). In contrast to the studied conifers, NLR clustering is frequent but not as ubiquitous in angiosperms. What further sets conifers apart from angiosperms regarding NLRs is the ubiquitous presence and abundance of all three NLR subfamilies (CNLs, RNLs, and TNLs) (Van Ghelder et al. 2019, Table 1 of this study). RNL abundance is rare in angiosperms and TNLs are absent in monocots (Van Ghelder et al. 2019). We found RNLs comprised 4% to 14% of the total NLR diversity in all gymnosperms, indicating that RNL abundance is an ancestral trait of the gymnosperms. RNLs were consistently divided over two phylogenetic clades, one of which comprised mainly CNLs (Fig. 4). This is consistent with an evolutionary divergence between TNLs and CNL/RNLs predating that between CNLs and RNLs (Shao et al. 2019). Interestingly, CNLs and TNLs shared ancestral origins in *G. biloba* (Fig. 4B). This potentially indicates frequent domain swapping between NLR genes, highlighting the dynamic nature of these resistance genes.

Evolution of Genomic Architecture in Conifer Giga-genomes

Conifers have very large genomes (18 to 34 Gb) and harbor large amounts of repetitive DNA sequences (Mackay et al. 2012; Birol et al. 2013; Nystedt et al. 2013; De La Torre et al. 2014; Zimin et al. 2014). Genome evolution is considered to be less dynamic in conifers and other gymnosperms compared with flowering plants (Leitch and Leitch 2012), which may suggest lower rates of gene diversification. Conversely, conifers have suites of rapidly evolving genes (Gagalova et al. 2022) and highly diversified gene families or subfamilies (e.g. Bedon et al. 2010; Stival Sena et al. 2018; Van Ghelder et al. 2019) both related to stimuli and stress response. Several comparative studies in conifers have shown high levels of intergeneric macro-synteny and macro-collinearity among Pinaceae taxa (Pelgas et al. 2006; Ritland et al. 2011; Pavy et al. 2012; Westbrook et al. 2015) and clear chromosomal rearrangements when comparing Pinaceae and Cupressaceae (Moriguchi et al. 2012; de Miguel et al. 2015). These observations

are consistent with a small number of whole-genome duplications early in conifer evolution (Li et al. 2015). By contrast, our study has focused on the genomic architecture of a targeted gene family, showing conserved localized clustering and shedding insights into the evolutionary trajectory of NLR genes. Our investigation was possible due to the following two types of relatively recent genomic resources: (i) highly contiguous genome assemblies, e.g. *S. giganteum* (Lindl.) J.Buchholz (Scott et al. 2020) and *T. chinensis* (Pilg.) Rehder (Xiong et al. 2021), among others, which are developed using proximity ligation (e.g. HiC [Belton et al. 2012]) and long-read sequencing (e.g. PacBio SMRT) and (ii) high-density genetic maps, which are available for *Pinus* L. spp. (e.g. Liu et al. 2019) and *Picea* A.Dietr. spp. (Bernhardsson et al. 2019; Gagalova et al. 2022; Tumas et al. 2024).

A specific feature of conifer genomes probably enabled the accumulation of NLR genes and facilitated the formation of very large gene clusters. Conifers are inefficient at removing extra copies of DNA sequence through proofediting, hence their propensity to accumulate gigabases of repetitive sequences such as the Type I transposable elements (e.g. copia and gypsy sequences) (Nystedt et al. 2013; Zimin et al. 2014). Conifers also retain a high proportion of pseudogenes (Warren et al. 2015) and single-copy sequences that are similar to protein coding genes (Pellicer et al. 2018). Therefore, we could expect a significant proportion of genomic NLR sequences in conifers to represent pseudogenes. However, RNA sequencing has identified between 271 and 725 NLR genes expressed across a suite of Pinaceae and Cupressaceae species (Van Ghelder et al. 2019; Liu et al. 2021b; Ence et al. 2022). Expression of selected NLRs was shown to be responsive to infection by *Phytophthora ramorum* in *Larix* spp. (Dun et al. 2022) or drought in *P. glauca* (Van Ghelder et al. 2019) and to be variable across different seed families in *P. flexilis* (Liu et al. 2021b).

Some neofunctionalization events may occur in NLR genes as several CNLs and TNLs were found to share an ancestral origin with the other subfamily (Fig. 4). These novel NLRs are likely to have formed by the loss of their ancestral *N*-terminal domain followed by the fusion to a new one, as has been reported in other plant systems (Seong et al. 2020). Similarly, we found evidence of RPW8 domains immediately upstream to CNLs across gymnosperms, leading to the birth of new RNLs. Finally, the presence of unclassified genes in both CNL and TNL clades (Fig. 4) indicates that this process is still ongoing for many NLR genes. Together, these findings emphasize the highly dynamic evolutionary nature of NLR genes, providing opportunities for evolvability to increasingly abundant and widespread pathogenic diseases. The genomic sequences identified here may prompt further work to determine which of these are expressed and under what circumstances. Studies of functional sequence divergence and evolutionary rates, aiming at identifying footprints of natural selection (e.g. Guillet-Claude et al. 2004), should also be considered (e.g. Chia and Carella 2023). In particular, the more accurate functional characterization of transcripts through long transcriptome sequencing (e.g. PacBio IsoSeq; Yu et al. 2023) and the adaptive sampling approach of full NLR genes (e.g. Oxford Nanopore Technologies; Belinchon-Moreno et al. 2023) seem promising avenues.

Opportunities for Genomic Breeding Using NLR Genes for Pest, Disease and Drought Resistance

Resistance to viruses, bacteria, oomycetes, fungi, and some insects has been linked to NLR genes in a range of flowering plants (Kourelis and van der Hoorn 2018), but our understanding of their contribution to resistance in conifers is very rudimentary. Dissection of the genetic resistance to fusiform rusts (*Cronartium quercuum* (Berk.) Miyabe ex Shirai f. sp. *fusiforme*) in *P. taeda* L. breeding populations has provided evidence for genetic resistance (Wilcox et al. 1996) and implicated NLR-encoding genes as likely candidates (Quesada et al. 2014; Ence et al. 2022). In *P. taeda*, nine different fusiform rust resistance loci were identified across three linkage groups (Amerson et al. 2015). These are hypothesized to contain NLR sequences, which were found to vary in numbers of genomic sequences when comparing populations from different geographic areas (Ence et al. 2022). Similarly, in *P. flexilis*, fine genetic dissection, evolutionary analysis, and expression profiling have identified two NLR genes as candidates for resistance linked to the Cr4 locus (Liu et al. 2021b) among the 155 NLR genes mapped across 12 linkage groups to date. Interestingly, the major clusters found in *P. flexilis* were on linkage groups distinct to the Cr4 locus. Taken together, these results indicate that resistance-associated NLR genes may be distributed across the genome. To our knowledge, resistance phenotypes have been identified in only a few studies in conifers and these have involved fungal rusts infecting pines, although expression profiling indicated responsiveness to *P. ramorum* (Dun et al. 2022) and drought in other species (Van Ghelder et al. 2019).

NLR-dense genomic regions could act as potential reservoirs for NLR diversity. The NLR clusters found in gymnosperm genomes probably arose by tandem duplications of TNL genes, as indicated by their consistent composition in conifers (Fig. 2B to D) and the close phylogenetic relationships of the TNLs on the same chromosome (Fig. 5). Although tandem duplication inevitably leads to identical gene copies initially (paralogs), it can eventually lead to gene diversification through differential mutation trajectories and domain swapping (Ostermeier and Benkovic 2001). There are documented examples of such neofunctionalizations for resistance genes in different plant lineages (Kong and Ranganathan 2008; Wei et al. 2023) and for

genes encoding transcription factors in the Pinaceae (Guillet-Claude et al. 2004). Given the long evolutionary history of conifer lineages (Leslie et al. 2018) and their considerable NLR diversity, a high degree of noncanonical NLR genes are expected, especially in and around these dense NLR gene clusters. Considering whole-genome datasets, about one in four NLR genes is unclassified in conifers (Table 1), meaning that a distinctive *N*-terminal domain is missing or ambiguous (e.g. fusion of TIR and CC domains). Closer inspection of NLR-dense regions could therefore reveal interesting noncanonical domains in conifer NLR genes. Together with the overall diversity of canonical NLR genes in these clusters, this could further emphasize their potential in genomics-assisted breeding to improve disease and drought resistance.

The NLR gene family is among the most studied in plants due to its agronomic importance (Kourelis and van der Hoorn 2018) with many linked breeding and genetic engineering applications proposed to improve resistance in crops (van Wersch et al. 2020) and to a lesser extent in forest trees (Ence et al. 2022). We have shown how improved genome sequences along with transcriptome data may enhance our understanding of NLR genomic architecture in this understudied group. To develop genetic resources that will help to respond to emerging threats in conifers, the following three components are also needed: (i) populations of phenotypically diverse individuals in which to study resistance traits (Liu et al. 2019; Ence et al. 2022); (ii) efficient and accurate assessment of the susceptibility and resistance phenotypes to relevant pests and diseases, which is difficult to accomplish and is therefore often either lacking or suboptimal in conifers; and (iii) fast and accurate genome scanning methods that are suitable for differentiating among genes and alleles within and among populations. The large size and the variability of the NLR gene family adds to the challenge of linking genes to resistance phenotypes; however, knowledge of the position of clusters of these immune receptor genes paves the way to more focused investigations in conjunction with genome selection and other genome-wide analyses.

Materials and Methods

Genomic Distribution of NLR Genes in Pinaceae and Other Conifer Families

Following the evidence for dense genomic clusters of NLR immune receptor genes in *P. flexilis* (Liu et al. 2019), we tested whether this is a Pinaceae family-wide phenomenon by examining publicly available genomic resources in other members of the family. To characterize the distribution of NLR genes in the speciose *Picea* genus, we deployed a high-density linkage map for *P. abies* (Bernhardsson et al. 2019) and high-quality genome assemblies of *P. glauca* and *P. sitchensis* (Gagalova et al. 2022). Although these assemblies do not have chromosome-level contiguity, they are suitably scaffolded into linkage groups corresponding to an updated version of the original *P. glauca* high-density linkage map (Pavy et al. 2017).

NLR genes were identified using the NLR Annotator pipeline v2.1 (Steuernagel et al. 2020). This actively updated pipeline (https://github.com/steuernb/NLR-Annotator) scans genome sequences for NLR-specific motifs and records the location and motif composition of each discovered gene. NLR Annotator, like its predecessor NLR-Parser, is actively used in highly ranked studies on NLR discovery and annotation (Avni et al. 2022; Li et al. 2023; Lin et al. 2023; Salcedo et al. 2023) and achieves the highest sensitivity and annotation specificity among genomic NLR annotation tools (Kourelis et al. 2021).

Linkage map positions were recorded for the 3 *Picea* species to map NLR gene distribution on the 12 different linkage groups. We utilized the recent linkage map comparison work by Tumas et al. (2024) to determine the syntenic linkage groups. NLR distribution data for *P. flexilis* were taken from the original high-density linkage map publication (Liu et al. 2019).

To compare distribution patterns across main lineages of conifers and gymnosperms, we applied the same pipeline to recently published chromosome-level genome assemblies: *S. giganteum* (Cupressaceae) (Scott et al. 2020), *T. chinensis* (Taxaceae) (Xiong et al. 2021), *G. biloba* (Ginkgoales) (Liu et al. 2021a), *P. tabuliformis* (Niu et al. 2022), and *C. panzhihuaensis* L.Zhou & S.Y.Yang (Cycadales) (Liu et al. 2022). The nucleotide positions of NLR genes were recorded for each chromosome.

To visualize the distribution of NLR genes, histograms were produced in R (R Core Development Team 2021) with ggplot2 v3.4.2 (Wickham 2016) where genes were plotted along the length of chromosomes (linkage groups for linkage maps) using the starting position in Mb (cM for linkage maps). The bin width was chosen to correspond roughly to 1% of the largest chromosome (or linkage group), i.e.: 13 Mb for *C. panzhihuaensis*, 12 Mb for *G. biloba*, 10 Mb for *S. giganteum* and *T. chinensis*, 4 Mb for *P. glauca*, 2 Mb for *P. sitchensis* (Bong.) Carrière, 4 cM for *P. abies* (L.) H.Karst., and 2 cM for *P. flexilis*. We calculated the expected number of NLR genes for each chromosome (or linkage group) to quantify abnormal distribution patterns:

$$\#\,\text{NLR genes}_{\text{expected}} = \frac{\text{Total \# NLR genes observed in species}}{\text{Total length of assembly (Mb) or linkage map (cM)}} \times \text{Length of chromosome (Mb) or linkage group}$$

The anomalies were visualized by plotting the observed values against the expected values for each chromosome (or linkage group) in a scatter plot. To determine whether these anomalies were statistically significant we performed two-character Fisher's exact tests in R (R Core Team 2021) and computed the *P*-values.

NLR Classification

The NLR Annotator pipeline describes the motifs discovered in each NLR based on a curated list of NLR motifs (Jupe et al. 2012). We utilized a custom python script (available on https://github.com/hung-th/NLRmeta), adapted from open-source code by Philipp Bayer (https://gist.github.com/philippbayer/0052f5ad56121cd2252a1c5b90154ed1) and based on the motif table in Jupe et al. (2012), to extract the motif output from the NLR Annotator and convert it into the CNL or TNL subfamily classification.

The third subfamily, RNL, is characterized by the variable *N*-terminal RPW8 domain but is not annotated by the NLR Annotator. In a previous study on conifer NLR genes, Van Ghelder et al. (2019) discovered two RNL-characteristic signatures in conifers: one located in the RNBS-D amino acid motif (CFLDLGxFP) and one in the MHD motif (QHD). We searched for these signatures in the generated NLR datasets by deploying the MAST software from the MEME suite (Bailey et al. 2015), and classified NLR genes as RNLs if they contained both signatures with a sum total of one amino acid mismatch allowed. A further search for RPW8 domains was performed with tblastn (Camacho et al. 2009) (*eV* <0.05) using the conifer RPW8 sequences characterized by Van Ghelder et al. (2019) against the genome assemblies and linkage map loci. We only retained RPW8 hits with ≥1/3 amino acid identity in the reference sequences. NLRs with an RPW8 domain fused to the *N*-terminal side were thereby classified as RNLs regardless of their RNBS-D and MHD motif composition. Separate RPW8 sequences (e.g. not fused to an NLR gene) were also characterized as RNL genes.

A fourth category of "unclassified" NLRs encompasses NLR genes that could not be classified into one of the three subfamilies owing to a lack of characteristic domains and motifs. NLR genes containing motifs characteristic of CNL as well as TNL were also labeled as "unclassified". For *P. flexilis*, we utilized the annotation information from the original linkage map publication (Liu et al. 2019) to divide NLRs into classes. RNLs in *P. flexilis* were classified in the same way as for the other species. NLR class information was used to further annotate the intrachromosomal NLR distribution histograms (see "Genomic Distribution of NLR Genes in Pinaceae and Other Conifer Families" section) to visualize patterns of class distribution.

NLR Annotator further determines whether detected NLR genes are complete, partial (missing domains), or pseudogenes (unexpected stop codon in sequence), which we recorded for each discovered gene, except for the separate RPW8 sequences.

After the discovery of surprising phylogenetic placements (see "NLR Phylogenies" section) for several NLR genes (i.e. a CNL gene placed in a TNL clade), we performed manual annotation checks on these outliers. We performed a tblastn with an expertly annotated set of NLR proteins from 6 different conifers (Van Ghelder et al. 2019) against 18 anomalies (see Supplementary Material). This revealed three CNLs from *S. giganteum* and one from *G. biloba* that were placed in TNL clades, but for which no clear evidence of a *N*-terminal CC domain was found. The annotation for these genes was subsequently updated to "unclassified". One TNL from *P. flexilis* was placed in an RNL clade but lacked an *N*-terminal TIR domain. The tblastn search revealed strong similarities with other conifer RNLs and the annotation of this gene was therefore updated to the status of RNL.

NLR Phylogenies

To determine the evolutionary history of discovered NLR genes, maximum likelihood phylogenies were generated from alignments of the central (conserved) NB-ARC domain. We deployed the field standard of using the amino acid sequence of the NB-ARC domain as it gives an unbiased evolutionary history, irrespective of the (highly variable) class-specific *N*-terminal domains. For *P. flexilis*, NB-ARC sequences were identified through a BlastP search (Camacho et al. 2009) (*eV* <0.05) of the reference NB-ARC sequence used by the NLR Annotator against the translated NLR sequences. All hits with ≥60 amino acid residues were extracted and aligned using MAFFT v7 (Katoh and Standley 2013) using default settings. For the other species, we utilized the "-a" flag in NLR Annotator to obtain NB-ARC domains of all complete NLR genes. Maximum likelihood phylogenetics was performed with IQTree v1.6.12 (Nguyen et al. 2015) using the "GTR20" model for protein evolution and 1,000 ultra-fast bootstrap replicates to calculate node support values.

Phylogenies were visualized in the online Interactive Tree of Life (iTOL) tool (Letunic and Bork 2021) and rerooted at the node separating RNL/CNL and TNL clades. NLR class and chromosome were mapped onto the topologies using the iTOL annotation editor.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We are grateful for feedback to this study and its results from the Sitka Spruced (https://sitkaspruced.web.ox.ac.uk/) and Spruce-Up communities (https://spruce-up.ca/en/). As part of the Sitka Spruced project (BB/P020488/1), this research was funded by BBSRC and a group of forest and wood processing industries.

Author Contributions

Y.W. and J.J.M. designed the study. Y.W. conducted the bioinformatic analyses with input from H.T., C.v.G., and T.H.H. Y.W. and J.J.M. wrote the paper with input from all authors.

Conflict of Interest

The authors declare that they have no competing interests.

Data Availability

Genomic mapping data of NLR gene distributions generated for this study are available in the Figshare digital repository under the following DOI:10.6084/m9.figshare.24412579. Phylogenetic trees presented in this study can be accessed on the interactive Tree of Life (iTOL) platform through the following link: https://itol.embl.de/shared/28ok88a8GMSLQ.

Literature Cited

Amerson HV, Nelson C, Kubisiak T, Kuhlman E, Garcia S. Identification of nine pathotype-specific genes conferring resistance to fusiform rust in loblolly pine (*Pinus taeda* L.). Forests. 2015:6(12): 2739–2761. https://doi.org/10.3390/f6082739.

Avni R, Lux T, Minz-Dub A, Millet E, Sela H, Distelfeld A, Deek J, Yu G, Steuernagel B, Pozniak C, et al. Genome sequences of three *Aegilops* species of the section Sitopsis reveal phylogenetic relationships and provide resources for wheat improvement. Plant J. 2022:110(1):179–192. https://doi.org/10.1111/tpj.15664.

Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. Nucleic Acids Res. 2015:43(W1):W39–W49. https://doi.org/10.1093/nar/gkv416.

Barragan AC, Weigel D. Plant NLR diversity: the known unknowns of pan-NLRomes. Plant Cell. 2021:33(4):814–831. https://doi.org/10.1093/plcell/koaa002.

Bedon F, Bomal C, Caron S, Levasseur C, Boyle B, Mansfield SD, Schmidt A, Gershenzon J, Grima-Pettenati J, Séguin A. Subgroup 4 R2R3-MYBs in conifer trees: gene family expansion and contribution to the isoprenoid- and flavonoid-oriented responses. J Exp Bot. 2010:61(14):3847–3864. https://doi.org/10.1093/jxb/erq196.

Belinchon-Moreno J, Berard A, Canaguier A, Chovelon V, Cruaud C, Engelen S, Feriche-Linares R, Le-Clainche I, Marande W, Rittener-Ruff V, et al. 2023. Nanopore adaptive sampling to identify the NLR-gene family in melon (*Cucumis melo L.*). bioRXiv 572599. https://doi.org/10.1101/2023.12.20.572599, 21 December 2023, preprint: not peer reviewed.

Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi–C: a comprehensive technique to capture the conformation of genomes. Methods. 2012:58(3):268–276. https://doi.org/10.1016/j.ymeth.2012.05.001.

Bernhardsson C, Wang X, Scofield DG, Schiffthaler B, Baison J, Street NR, García-Gil MR, Ingvarsson PK. An ultra-dense haploid genetic map for evaluating the highly fragmented genome assembly of Norway spruce (*Picea abies*). G3 (Bethesda). 2019:9(5): 1623–1632. https://doi.org/10.1534/g3.118.200840.

Birol I, Raymond A, Jackman SD, Pleasance S, Coope R, Taylor GA, Yuen MMS, Keeling CI, Brand D, Vandervalk BP, et al. Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. Bioinformatics. 2013:29(12):1492–1497. https://doi.org/10.1093/bioinformatics/btt178.

Bonello P, Gordon TR, Herms DA, Wood DL, Erbilgin N. Nature and ecological implications of pathogen-induced systemic resistance in conifers: a novel hypothesis. Physiol Mol Plant Pathol. 2006:68(4-6):95–104. https://doi.org/10.1016/j.pmpp.2006.12.002.

Bousquet J, Gérardi S, De Lafontaine G, Jaramillo-Correa JP, Pavy N, Prunier J, Lenz P, Beaulieu J. Spruce population genomics. In: Rajora OP, editor. Population genomics. Springer Nature; 2021. p. 1–64.

Brar S, Tabima JF, McDougal RL, Dupont P-Y, Feau N, Hamelin RC, Panda P, LeBoldus JM, Grünwald NJ, Hansen EM, et al. Genetic diversity of Phytophthora pluvialis, a pathogen of conifers, in New Zealand and the west coast of the United States of America. Plant Pathol. 2018:67(5):1131–1139. https://doi.org/10.1111/ppa.12812.

Brasier C, Webber J. Sudden larch death. Nature. 2010:466(7308): 824–825. https://doi.org/10.1038/466824a.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009:10(1):421. https://doi.org/10.1186/1471-2105-10-421.

Chen J, Zhao Y, Luo X, Hong H, Yang T, Huang S, Wang C, Chen H, Qian X, Feng M, et al. NLR surveillance of pathogen interference with hormone receptors induces immunity. Nature. 2023:613(7942): 145–152. https://doi.org/10.1038/s41586-022-05529-9.

Chia KS, Carella P. Taking the lead: NLR immune receptor N-terminal domains execute plant immune responses. New Phytol. 2023:240(2):496–501. https://doi.org/10.1111/nph.19170.

Christopoulou M, Wo SR-C, Kozik A, McHale LK, Truco M-J, Wroblewski T, Michelmore RW. Genome-wide architecture of disease resistance genes in lettuce. G3 (Bethesda). 2015:5(12): 2655–2669. https://doi.org/10.1534/g3.115.020818.

De La Torre AR, Birol I, Bousquet J, Ingvarsson PK, Jansson S, Jones SJM, Keeling CI, MacKay J, Nilsson O, Ritland K, et al. Insights into conifer giga-genomes. Plant Physiol. 2014:166(4): 1724–1732. https://doi.org/10.1104/pp.114.248708.

de Miguel M, Bartholomé J, Ehrenmann F, Murat F, Moriguchi Y, Uchiyama K, Ueno S, Tsumura Y, Lagraulet H, de Maria N. Evidence of intense chromosomal shuffling during conifer evolution. Genome Biol Evol. 2015:evv185. https://doi.org/10.1093/gbe/evv185.

Dun HF, Hung TH, Green S, MacKay JJ. Comparative transcriptomic responses of European and Japanese larches to infection by *Phytophthora ramorum*. BMC Plant Biol. 2022:22(1):480. https://doi.org/10.1186/s12870-022-03806-3.

Duxbury Z, Wu C, Ding P. A comparative overview of the intracellular guardians of plants and animals: NLRs in innate immunity and beyond. Annu Rev Plant Biol. 2021:72(1):155–184. https://doi.org/10.1146/annurev-arplant-080620-104948.

Ence D, Smith KE, Fan S, Gomide Neves L, Paul R, Wegrzyn J, Peter GF, Kirst M, Brawner J, Nelson CD, et al. NLR diversity and candidate fusiform rust resistance genes in loblolly pine. G3 (Bethesda). 2022:12(2):jkab421. https://doi.org/10.1093/g3journal/jkab421.

Farjon A, Filer D, editors. 2013. An atlas of the world's conifers: an analysis of their distribution, biogeography, diversity and conservation status. Leiden: Brill.

Gagalova KK, Warren RL, Coombe L, Wong J, Nip KM, Yuen MMS, Whitehill JGA, Celedon JM, Ritland C, Taylor GA, et al. Spruce giga-genomes: structurally similar yet distinctive with differentially expanding gene families and rapidly evolving genes. Plant J. 2022:111(5):1469–1485. https://doi.org/10.1111/tpj.15889.

Guillet-Claude C, Isabel N, Pelgas B, Bousquet J. The evolutionary implications of knox-I gene duplications in conifers: correlated evidence from phylogeny, gene mapping, and analysis of functional divergence. Mol Biol Evol. 2004:21(12):2232–2245. https://doi.org/10.1093/molbev/msh235.

Jakoby O, Lischke H, Wermelinger B. Climate change alters elevational phenology patterns of the European spruce bark beetle (*Ips typographus*). Glob Change Biol. 2019:25(12):4048–4063. https://doi.org/10.1111/gcb.14766.

Jia YX, Yuan Y, Zhang Y, Yang S, Zhang X. Extreme expansion of NBS-encoding genes in Rosaceae. BMC Genet. 2015:16(1):48. https://doi.org/10.1186/s12863-015-0208-x.

Jupe J, Pritchard L, Etherington GJ, MacKenzie K, Cock PJA, Wright F, Sharma SK, Bolser D, Bryan GJ, Jones JDG, et al. Identification and localisation of the NB-LRR gene family within the potato genome. BMC Genomics. 2012:13(1):75. https://doi.org/10.1186/1471-2164-13-75.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013:30(4):772–780. https://doi.org/10.1093/molbev/mst010.

Kong L, Ranganathan S. Tandem duplication, circular permutation, molecular adaptation: how Solanaceae resist pests via inhibitors. BMC Bioinformatics. 2008:9(S1):S22. https://doi.org/10.1186/1471-2105-9-S1-S22.

Kourelis J, Sakai T, Adachi H, Kamoun S. RefPlantNLR is a comprehensive collection of experimentally validated plant disease resistance proteins from the NLR family. PLoS Biol. 2021:19(10):e3001124. https://doi.org/10.1371/journal.pbio.3001124.

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. 2018:30(2):285–299. https://doi.org/10.1105/tpc.17.00579.

Laberge M-J, Payette S, Bousquet J. Life span and biomass allocation of stunted black spruce clones in the subarctic environment. J Ecol. 2000:88(4):584–593.

Leitch IJ, Leitch AR. Genome size diversity and evolution in land plants. In: Greilhuber J, Dolezel J, Wendel J, editors. Plant genome diversity volume 2. Springer; 2012. p. 307–322. https://doi.org/10.1007/978-3-7091-1160-4_19.

Lenz PRN, Nadeau S, Mottet M-J, Perron M, Isabel N, Beaulieu J, Bousquet J. Multi-trait genomic selection for weevil resistance, growth, and wood quality in Norway spruce. Evol Appl. 2020:13(1):76–94. https://doi.org/10.1111/eva.12823.

Leslie AB, Beaulieu J, Holman G, Campbell CS, Mei W, Raubeson LR, Mathews S. An overview of extant conifer evolution from the perspective of the fossil record. Am J Bot. 2018:105(9):1531–1544. https://doi.org/10.1002/ajb2.1143.

Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021:49(W1):W293–W296. https://doi.org/10.1093/nar/gkab301.

Li Z, Baniaga AE, Sessa EB, Scascitelli M, Graham SW, Rieseberg LH, Barker MS. Early genome duplications in conifers and other seed plants. Sci Adv. 2015:1(10):e1501084. https://doi.org/10.1126/sciadv.1501084.

Li X, Ma L, Wang Y, Ye C, Guo C, Li Y, Mei X, Du F, Huang H. PlantNLRatlas: a comprehensive dataset of full- and partial-length NLR resistance genes across 100 chromosome-level plant genomes. Front Plant Sci. 2023:14:1178069. https://doi.org/10.3389/fpls.2023.1178069.

Lin X, Jia Y, Heal R, Prokchorchik M, Sindalovskaya M, Olave-Achury A, Makechemu M, Fairhead S, Noureen A, Heo J, et al. Solanum americanum genome-assisted discovery of immune receptors that detect potato late blight pathogen effectors. Nat Genet. 2023:55(9):1579–1588. https://doi.org/10.1038/s41588-023-01486-9.

Lin X, Zhang Y, Kuang H, Chen J. Frequent loss of lineages and deficient duplications accounted for low copy number of disease resistance genes in Cucurbitaceae. BMC Genomics. 2013:14(1):335. https://doi.org/10.1186/1471-2164-14-335.

Liu JJ, Schoettle AW, Sniezko RA, Williams H, Zamany A, Rancourt B. Fine dissection of limber pine resistance to Cronartium ribicola using targeted sequencing of the NLR family. BMC Genomics. 2021b:22(1):567. https://doi.org/10.1186/s12864-021-07885-8.

Liu JJ, Schoettle AW, Sniezko RA, Yao F, Zamany A, Williams H, Rancourt B. Limber pine (Pinus flexilis James) genetic map constructed by exome-seq provides insight into the evolution of disease resistance and a genomic resource for genomics-based breeding. Plant J. 2019:98(4):745–758. https://doi.org/10.1111/tpj.14270.

Liu Y, Wang S, Li L, Yang T, Dong S, Wei T, Wu S, Liu Y, Gong Y, Feng X, et al. The Cycas genome and the early evolution of seed plants. Nat Plants. 2022:8(4):389–401. https://doi.org/10.1038/s41477-022-01129-7.

Liu H, Wang X, Wang G, Cui P, Wu S, Ai C, Hu N, Li A, He B, Shao X, et al. The nearly complete genome of Ginkgo biloba illuminates gymnosperm evolution. Nat Plants. 2021a:7(6):748–756. https://doi.org/10.1038/s41477-021-00933-x.

Liu Y, Zeng Z, Zhang Y-M, Li Q, Jiang X-M, Jiang Z, Tang J-H, Chen D, Wang Q, Chen J-Q, et al. An angiosperm NLR Atlas reveals that NLR gene reduction is associated with ecological specialization and signal transduction component deletion. Mol Plant. 2021c:14(12):2015–2031. https://doi.org/10.1016/j.molp.2021.08.001.

Mackay JJ, Dean JFD, Plomion C, Peterson DG, Cánovas FM, Pavy N, Ingvarsson PK, Savolainen O, Guevara MÁ, Fluch S, et al. Towards decoding the conifer giga-genome. Plant Mol Biol. 2012:80(6):555–569. https://doi.org/10.1007/s11103-012-9961-7.

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis. Plant Cell. 2003:15(4):809–834. https://doi.org/10.1105/tpc.009308.

Mitton JB, Ferrenberg SM. Mountain pine beetle develops an unprecedented summer generation in response to climate warming. Am Nat. 2012:179(5):E163–E171. https://doi.org/10.1086/665007.

Moriguchi Y, Ujino-Ihara T, Uchiyama K, Futamura N, Saito M, Ueno S, Matsumoto A, Tani N, Taira H, Shinohara K. The construction of a high-density linkage map for identifying SNP markers that are tightly linked to a nuclear-recessive major gene for male sterility in Cryptomeria japonica D. Don. BMC Genomics. 2012:13(1). https://doi.org/10.1186/1471-2164-13-95.

Mota MM, Braasch H, Bravo MA, Penas AC, Burgermeister W, Metge K, Sousa E. First report of Bursaphelenchus xylophilus in Portugal and in Europe. Nematology. 1999:1(7):727–734. https://doi.org/10.1163/156854199508757.

Mullin TJ, Lee SJ. Best practices for conifer breeding in Europe. Uppsala: Skogforsk; 2013.

Neale DB, Kremer A. Forest tree genomics: growing resources and applications. Nat Rev Genet. 2011:12:111–122. https://doi.org/10.1038/nrg2931.

Neale DB, Zimin AV, Zaman S, Scott AD, Shrestha B, Workman RE, Puiu D, Allen BJ, Moore ZJ, Sekhwal MK, et al. Assembled and annotated 26.5 Gbp coast redwood genome: a resource for estimating evolutionary adaptive potential and investigating hexaploid origin. G3 (Bethesda). 2022:12(1):jkab380. https://doi.org/10.1093/g3journal/jkab380.

Ngou BPM, Ding P, Jones JDG. Thirty years of resistance: zig-zag through the plant immune system. Plant Cell. 2022:34(5):1447–1478. https://doi.org/10.1093/plcell/koac041.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood

phylogenies. Mol Biol Evol. 2015:32(1):268–274. https://doi.org/10.1093/molbev/msu300.
Niu S, Li J, Bo W, Yang W, Zuccolo A, Giacomello S, Chen X, Han F, Yang J, Song Y, et al. The Chinese pine genome and methylome unveil key features of conifer evolution. Cell. 2022:185(1):204–217.e14. https://doi.org/10.1016/j.cell.2021.12.006.
Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, et al. The Norway spruce genome sequence and conifer genome evolution. Nature. 2013:497(7451):579–584. https://doi.org/10.1038/nature12211.
Ostermeier M, Benkovic SJ. Evolution of protein function by domain swapping. Adv Protein Chem. 2001:55:29–77. https://doi.org/10.1016/S0065-3233(01)55002-0.
Pavy N, Lamothe M, Pelgas B, Gagnon F, Birol I, Bohlmann J, Mackay J, Isabel N, Bousquet J. A high-resolution reference genetic map positioning 8.8 K genes for the conifer white spruce: structural genomics implications and correspondence with physical distance. Plant J. 2017:90(1):189–203. https://doi.org/10.1111/tpj.13478.
Pavy N, Pelgas B, Laroche J, Rigault P, Isabel N, Bousquet J. A spruce gene map infers ancient plant genome reshuffling and subsequent slow evolution in the gymnosperm lineage leading to extant conifers. BMC Biol. 2012:10(1):84. https://doi.org/10.1186/1741-7007-10-84.
Pelgas B, Beauseigle S, Acheré V, Jeandroz S, Bousquet J, Isabel N. Comparative genome mapping among Picea glauca, P. mariana × P. rubens and P. abies, and correspondence with other Pinaceae. Theor Appl Genet. 2006:113(8):1371–1393. https://doi.org/10.1007/s00122-006-0354-7.
Pellicer J, Hidalgo O, Dodsworth S, Leitch IJ. Genome size diversity and its impact on the evolution of land plants. Genes (Basel). 2018:9(2):88. https://doi.org/10.3390/genes9020088.
Pellicer J, Leitch IJ. The plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. New Phytol. 2020:226(2):301–305. https://doi.org/10.1111/nph.16261.
Quesada T, Resende M Jr, Muñoz P, Wegrzyn J, Neale D, Kirst M, Peter G, Gezan S, Nelson C, Davis J et al. Mapping fusiform rust resistance genes within a complex mating design of loblolly pine. Forests. 2014:5(2):347–362. https://doi.org/10.3390/f5020347.
R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2021. https://www.R-project.org/.
Ritland K, Krutovsky KV, Tsumura Y, Pelgas B, Isabel N, Bousquet J. Genetic mapping in conifers. Genetics, genomics and breeding of conifers. Taylor-Francis; 2011. p. 196–238.
Salcedo A, Parada-Rojas CH, Guerrero R, Stahr M, D'Arcangelo KN, McGregor C, Kousik C, Wehner T, Quesada-Ocampo LM. The NLR family of disease resistance genes in cultivated watermelon and other cucurbits: opportunities and challenges. In: Dutta SK, Nimmakayala P, Reddy UK, editors. The watermelon genome. Compendium of plant genomes. Cham: Springer; 2023. p. 37–67.
Scott AD, Zimin AV, Puiu D, Workman R, Britton M, Zaman S, Caballero M, Read AC, Bogdanove AJ, Burns E, et al. A reference genome sequence for giant sequoia. G3 (Bethesda). 2020:10(11):3907–3919. https://doi.org/10.1534/g3.120.401612.
Seo E, Kim S, Yeom SI, Choi D. Genome-wide comparative analyses reveal the dynamic evolution of nucleotide-binding leucine-rich repeat gene family among Solanaceae plants. Front Plant Sci. 2016:7:01205. https://doi.org/10.3389/fpls.2016.01205.
Seong K, Seo E, Witek K, Li M, Staskawicz B. Evolution of NLR resistance genes with noncanonical N-terminal domains in wild tomato species. New Phytol. 2020:227(5):1530–1543. https://doi.org/10.1111/nph.16628.
Shao ZQ, Xue JY, Wang Q, Wang B, Chen J-Q. Revisiting the origin of plant NBS-LRR genes. Trends Plant Sci. 2019:24(1):9–12. https://doi.org/10.1016/j.tplants.2018.10.015.
Smith PH, Hadfield J, Hart NJ, Koebner RMD, Boyd LA. STS markers for the wheat yellow rust resistance gene Yr5 suggest a NBS–LRR-type resistance gene cluster. Genome. 2007:50(3):259–265. https://doi.org/10.1139/G07-004.
Steuernagel B, Witek K, Krattinger SG, Ramirez-Gonzalez RH, Schoonbeek H-J, Yu G, Baggs E, Witek AI, Yadav I, Krasileva KV, et al. The NLR-annotator tool enables annotation of the intracellular immune receptor repertoire. Plant Physiol. 2020:183(2):468–482. https://doi.org/10.1104/pp.19.01273.
Stival SJ, Giguère I, Rigault P, Bousquet J, Mackay J. Expansion of the dehydrin gene family in the Pinaceae is associated with considerable structural diversity and drought-responsive expression. Tree Physiol. 2018:38(3):442–456. https://doi.org/10.1093/treephys/tpx125.
Stocks JJ, Metheringham CL, Plumb WJ, Lee SJ, Kelly LJ, Nichols RA, Buggs RJA. Genomic basis of European ash tree resistance to ash dieback fungus. Nat Ecol Evol. 2019:3:1686–1696. https://doi.org/10.1038/s41559-019-1036-6.
Teshome DT, Zharare GE, Naidoo S. The threat of the combined effect of biotic and abiotic stress factors in forestry under a changing climate. Front Plant Sci. 2020:11:601009. https://doi.org/10.3389/fpls.2020.601009.
Tumas H, Ilska JJ, Girardi S, Laroche J, A'Hara S, Boyle B, Janes M, McLean P, Lopez G, Lee SJ, et al. High-density genetic linkage mapping in Sitka spruce advances the integration of genomic resources in conifers. G3 (Bethesda). 2024:14:jkae020. https://doi.org/10.1101/2023.08.21.554184.
van der Biezen EA, Jones JDG. The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. Curr Biol. 1998:8(7):R226–R228. https://doi.org/10.1016/S0960-9822(98)70145-9.
van Ghelder C, Parent GJ, Rigault P, Prunier J, Giguère I, Caron S, Stival Sena J, Deslauriers A, Bousquet J, Esmenjaud D, et al. The large repertoire of conifer NLR resistance genes includes drought responsive and highly diversified RNLs. Sci Rep. 2019:9(1):11614. https://doi.org/10.1038/s41598-019-47950-7.
van Wersch S, Li X. Stronger when together: clustering of plant NLR disease resistance genes. Trends Plant Sci. 2019:24(8):688–699. https://doi.org/10.1016/j.tplants.2019.05.005.
van Wersch S, Tian L, Hoy R, Li X. Plant NLRs: the whistleblowers of plant immunity. Plant Commun. 2020:1(1):100016. https://doi.org/10.1016/j.xplc.2019.100016.
Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F, et al. The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet. 2013:45(5):487–494. https://doi.org/10.1038/ng.2586.
Warren RL, Keeling CI, Yuen Macaire Man S, Raymond A, Taylor GA, Vandervalk BP, Mohamadi H, Paulino D, Chiu R, Jackman SD. Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. The Plant Journal. 2015:83(2):189–212. https://doi.org/10.1111/tpj.12886.
Wei W, Wu X, Garcia A, McCoppin N, Viana JPG, Murad PS, Walker DR, Hartman GL, Domier LL, Hudson ME, et al. An NBS-LRR protein in the Rpp1 locus negates the dominance of Rpp1-mediated resistance against Phakopsora pachyrhizi in soybean. Plant J. 2023:113(5):915–933. https://doi.org/10.1111/tpj.16038.

Westbrook JW, Chhatre VE, Wu L-S, Chamala S, Neves LG, Muñoz P, Martínez-García PJ, Neale DB, Kirst M, Mockaitis K. A Consensus Genetic Map for *Pinus taeda* and *Pinus elliottii* and Extent of Linkage Disequilibrium in Two Genotype-Phenotype Discovery Populations of *Pinus taeda*. G3 Genes|Genomes|Genetics. 2015:5(8):1685–1694. https://doi.org/10.1534/g3.115.019588.

White TL, Adams WT, Neale DB. Forest Genetics. Wallingford: Cabi; 2007.

Wickham H. Ggplot2: elegant graphics for data analysis. R package version 3.4.2. New York: Springer; 2016.

Wilcox PL, Amerson HV, Kuhlman EG, Liu BH, O'Malley DM, Sederoff RR. Detection of a major gene for resistance to fusiform rust disease in loblolly pine by genomic mapping. Proc Natl Acad Sci USA. 1996:93(9):3859–3864. https://doi.org/10.1073/pnas.93.9.3859.

Xiong Y, Gou J, Liao Q, Li Y, Zhou Q, Bi G, Li C, Du R, Wang X, Sun T, et al. The Taxus genome provides insights into paclitaxel biosynthesis. Nat Plants. 2021:7(8):1026–1036. https://doi.org/10.1038/s41477-021-00963-5.

Yu Z, Chen Y, Zhou Y, Zhang Y, Li M, Ouyang Y, Chebotarov D, Mauleon R, Zhao H, Xie W, et al. Rice Gene Index: a comprehensive pan-genome database for comparative and functional genomics of Asian rice. Mol Plant. 2023:16(5):P798–P801. https://doi.org/10.1016/j.molp.2023.03.012.

Zimin A, Stevens KA, Crepeau MW, Holtz-Morris A, Koriabine M, Marçais G, Puiu D, Roberts M, Wegrzyn JL, de Jong PJ, et al. 2014. Sequencing and assembly of the 22-Gb loblolly pine genome. Genetics. 196: 875–890. https://doi.org/10.1534/genetics.113.159715.

Associate editor: Rob Lanfear