Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Does climate shape plant immune gene repertoires? A phylogenetically controlled test across angiosperms

Stanley SX Tan

stanley.tan@yale.edu

Yale University

Research Article

Keywords: Plant NLRs, climate, phylogeny, functional traits

Posted Date: July 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10157914/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

The number of immune receptor genes - cell-surface pattern recognition receptors (PRRs) and intracellular NLRs - varies by more than an order of magnitude across flowering plants, but the ecological and evolutionary correlates of this variation remain unclear. A widely held expectation is that plants in warmer, wetter, lower-latitude environments experience greater pathogen pressure and therefore maintain larger immune repertoires. Because immune gene counts and climatic niche are both inherited along the phylogeny, testing this expectation requires explicit phylogenetic correction. I related per-species counts of all 39 annotated immune receptor families to climate, geography, and functional traits across angiosperms, and tested whether ambient fungal pathogen exposure explains any residual association.

Results

Bivariate screening across 584 non-special-lifestyle species identified 56 of 702 climate-receptor correlations as significant, but all effect sizes were modest ($|\rho| \leq 0.24$) and none survived phylogenetic generalised least squares (PGLS) correction. Among functional traits, eleven of 585 tests survived correction, linking woody and tree growth forms to elevated LysM-RLK, Malectin-RLK, and several LRR-RLK subclasses. Under phylogenetic correction across all 39 families, however, no family survived joint FDR correction for either the woody/non-woody or the tree-versus-herb contrast (LysM-RLK closest, FDR $p = 0.13$). LysM-RLK, a pre-specified focal family, remained the strongest and most consistent signal - elevated in woody and tree species under both non-parametric (FDR $p = 0.004$) and phylogenetic tests, significant when evaluated within the 12 principal receptor families ($p = 0.007$) and robust to branch-length scaling, taxon subsampling, domestication, and genome size, but not significant under the more conservative joint correction across all 39 annotated families (FDR $p = 0.13$).

Conclusion

Across angiosperms, climate and geography are not robust predictors of immune receptor gene content once shared ancestry is taken into account. The more consistent correlate of immune gene number is growth form: woody and tree species tend to carry more LysM-RLK and other cell-surface receptor genes than herbs, a pattern not explained by ambient pathogen exposure. Whether this reflects pathogen immunity or mycorrhizal symbiosis cannot be resolved from aggregate gene counts and is a priority for future work.

Background

Plants cannot escape their pathogens. Lacking mobility and an adaptive immune system, they defend themselves through an innate immune system built on two classes of receptor (1). Cell-surface pattern recognition receptors (PRRs) - predominantly receptor-like kinases (RLKs) and receptor-like proteins

(RLPs) - detect conserved microbial molecules such as chitin, flagellin, and peptidoglycan, initiating pattern-triggered immunity (2–4). Intracellular nucleotide-binding leucine-rich repeat receptors (NLRs) recognise pathogen effectors delivered into the host cell and trigger a stronger, often localised defence response (5,6). Together these receptor families form the molecular interface between a plant and its microbial environment, and the size of each repertoire can reflect the spectrum of pathogens a lineage has encountered over evolutionary time (7).

The number of immune receptor genes varies enormously across flowering plants, ranging from a few dozen to well over a thousand NLRs per genome (8). This variation is generated by rapid birth-and-death evolution: NLR and many RLK families expand through tandem and segmental duplication and contract through pseudogenisation and deletion, producing copy numbers that can diverge substantially even between close relatives (9,10). Such lability makes immune gene counts a plausible target of selection imposed by the local biotic environment, and raises a long-standing question in plant macroecology: do plants exposed to greater pathogen pressure maintain larger immune repertoires?

A widely invoked expectation is that pathogen pressure increases toward warmer, wetter, lower-latitude environments, mirroring the broader latitudinal gradient in species diversity and biotic interaction intensity (11–13). If immune investment tracks pathogen pressure, plants of tropical and mesic climates might be expected to carry more NLR and PRR genes than those of cold or arid regions. Evidence for this prediction, however, is mixed: studies of immune gene variation across species have reported associations with climate and latitude in some systems but not others, and the direction of any such effect remains contested (14,15). Critically, immune gene counts and climatic niche are both inherited along the plant phylogeny, so that closely related species tend to resemble one another in both (15,16). Any apparent correlation between climate and immune repertoire may therefore reflect shared ancestry rather than a direct adaptive response - a confound that can only be addressed with explicit phylogenetic comparative methods.

Pathogen pressure is not the only axis along which immune demand may vary. Plant functional traits, particularly the distinction between woody and herbaceous growth, impose contrasting ecological lifespans and tissue economies that could shape immune investment independently of climate (17–19). Long-lived woody tissues persist for years to centuries and can be expected to withstand sustained and accumulating pathogen exposure, which may select for larger or differently composed receptor repertoires than the comparatively ephemeral tissues of herbs. Whether immune gene content covaries with growth form, and whether any such association is separable from climate and phylogeny, has received little systematic attention.

A handful of studies have begun to place immune receptor gene variation in an explicitly comparative and ecological frame, but the picture they paint is fragmentary. Early analyses established that plant resistance-gene clusters evolve by a birth-and-death process driven by tandem duplication and unequal crossing-over (20), and species-wide surveys have since shown that NLR repertoires differ dramatically even within a single species or genus and are difficult to anchor to any single reference genome (21).

With the accumulation of sequenced genomes, broader work has related receptor diversity to ecological context: subsets of NLR genes show signatures of adaptation as lineages colonise new habitats (22), and NLR number has been linked to genome size, whole-genome duplication history, and to life-history axes such as annual versus perennial habit (23–25). Where a climatic signal in immune gene content has been reported, it has come largely from intraspecific gradient studies within single species: disease-resistance gene number and diversity increase with precipitation in the prairie perennial *Silphium integrifolium* (26) and in the dominant tallgrass *Andropogon gerardii* (27), leaving open whether such within-species patterns generalise across angiosperms once shared ancestry is taken into account.

The strength of biotic interactions, including herbivory and pathogen pressure, is itself widely argued to increase toward the equator (28), so climate, pathogen pressure, and phylogeny are deeply entangled in any cross-species comparison. In parallel, the dual role of LysM-type receptors in both chitin-triggered immunity and the establishment of mycorrhizal and rhizobial symbioses is well established at the molecular level in model systems (29), yet neither this functional duality nor the woody habit has been tested as a macroecological correlate of immune gene number across angiosperms. The recent availability of a uniformly annotated immune receptor catalogue spanning hundreds of angiosperm genomes (8) makes it possible to evaluate these competing climatic and trait-based hypotheses on a phylogenetic basis.

Here I use a recently published genome-wide annotation of immune receptor genes across more than 800 angiosperm species (8) to test these questions at macroecological scale. I set out to evaluate a leading hypothesis: that plants in warmer and moister climates, and at lower latitudes, experience greater pathogen pressure and therefore carry larger immune receptor repertoires, in particular more NLR genes. I relate per-species counts of all 39 annotated immune receptor families to climate, geography, and functional traits; apply phylogenetic comparative methods throughout to distinguish genuine ecological associations from shared ancestry; and use independent estimates of ambient fungal pathogen exposure, derived from global occurrence and soil-metabarcoding databases, to test whether the pathogen environment accounts for any residual signal. As shown, the growth-form contrast (woody/tree species carrying more LysM-RLK and related receptor genes) is the strongest trait signal. LysM-RLK was identified *a priori*, among the principal receptor families, as significantly higher in trees ($p = 0.007$), but this signal did not retain significance under the conservative correction across all 39 families (FDR $p = 0.13$). The association is robust to confounding variables yet modest in effect-size. Growth form, rather than climate, is therefore the clearest correlate of immune gene content in these data, though the signal is modest and does not reach genome-wide significance

Methods

Immune receptor gene counts

Per-species counts of pattern recognition receptors (PRRs) and intracellular NBS-LRR receptors (NLRs) were taken from a published genome-wide annotation of immune receptor genes across 808

angiosperm species (8), which annotated receptor genes from publicly available whole-genome assemblies. All 39 immune receptor gene families available in the dataset were analysed: six NLR families (NLR total, CNL, TNL, RNL, ADR1, NRG1); eight top-level RLK families (IR-RLK, Duf26-RLK, G-lectin-RLK, L-lectin-RLK, LRR-RLK aggregate, LysM-RLK, Malectin-RLK, WAK-RLK); the 21 LRR-RLK subclasses (I-XV) defined in the source classification; and four RLP families (IR-RLP, LRR-RLP, LRR-RLP-ID4LRR, LysM-RLP). The LRR-RLK subclasses are non-independent of the LRR-RLK aggregate and are interpreted in that light. All families had complete data across the 176 non-SLH trait-dataset species and 584 non-SLH climate-dataset species. All counts were log(x + 1)-transformed prior to analysis to address right skew.

Exclusion of special-lifestyle/habitat species

Species annotated in the source data as having a special lifestyle or habitat (SLH; aquatic, carnivorous, desert, mangrove, parasitic, or mycoheterotrophic) were excluded from primary analyses, as such species show convergent reduction of immune receptor repertoires that would confound macroecological comparisons (Li et al., 2025). SLH exclusion was applied separately to each analysis dataset, and SLH inclusion was tested as a sensitivity analysis.

Geographic and climate data

Georeferenced occurrence records were retrieved per species from the Global Biodiversity Information Facility (GBIF) API. Up to 500 records per species were retained, restricted to a basis of record of human observation, machine observation, or preserved specimen, with coordinate precision better than 0.01°, and excluding records flagged by GBIF for zero-degree coordinates, country centroids, or invalid coordinates. Species with fewer than 20 valid records were excluded. Four geographic variables were computed per species: absolute centroid latitude, latitudinal range, longitudinal range, and the proportion of records within 23.5° of the equator.

Climate values were extracted from WorldClim v2.1 bioclimatic rasters at 2.5 arc-minute resolution (30) at every occurrence point using rasterio. For each of the 19 bioclimatic variables, the per-species median (niche central tendency) and standard deviation (niche breadth) were computed. The climate predictor set comprised 18 continuous variables: 4 geographic, 8 WorldClim medians (annual mean temperature, mean diurnal range, isothermality, temperature seasonality, maximum temperature of warmest month, minimum temperature of coldest month, temperature annual range, and annual precipitation), and 6 WorldClim standard deviations. The merged climate dataset contained 660 species with both gene counts and geographic/climate data; 584 were retained after SLH exclusion. Plant lineage (monocot, eudicot, or basal angiosperm) was assigned from family-level taxonomy using the V.PhyloMaker2 reference species list (Jin & Qian, 2022).

Plant functional traits

Trait data were merged into a per-species table of 192 species with both gene counts and trait data; 176 were retained after SLH exclusion, of which 162 were represented in the phylogeny. Three categorical traits were retrieved from the Global Inventory of Floras and Traits (GIFT): woodiness (woody / non-woody / variable), growth form (herb / shrub / tree), and Raunkiaer life form. Species were matched by exact name. Missing values for woodiness and growth form in the original datasets were imputed using data from online botanical repositories and knowledge sources (Supplementary Information – Condensed Growth Form).

Twelve continuous traits were obtained from the TRY Plant Trait Database (31,32) as per-species means pooled across all contributor datasets, restricted to traits with at least 40 non-missing species-level observations: specific leaf area, leaf nitrogen per dry mass, leaf dry matter content, leaf phosphorus per dry mass, leaf dry mass, stem specific density, leaf thickness, leaf length, leaf carbon per dry mass, leaf tissue density, leaf nitrogen per area, and leaflet number.

Genome assembly quality (BUSCO completeness), domestication status, and assembled genome size were taken from the source paper's supplementary data. BUSCO was included as a covariate in all phylogenetic models; domestication status and genome size were used in sensitivity analyses.

Species Filtering

Immune receptor gene counts were obtained for 808 angiosperm species from the source annotation (Li et al., 2025), and this set was filtered separately for the climate and trait analyses according to data availability. For the climate analysis, 660 species had sufficient georeferenced occurrence records (GBIF) and extractable WorldClim climate data; of these, 76 were classified as special-lifestyle/habitat (SLH) species and initially excluded (until sensitivity analysis), retaining 584 non-SLH angiosperms for bivariate screening. Phylogenetic analyses were further restricted to the 460 of these species that could be placed on the GBOTB backbone with V.PhyloMaker2, the remainder belonging to genera absent from the reference phylogeny. For the trait analysis, 192 species had matching functional-trait data from the TRY and GIFT databases; 16 SLH species were excluded, retaining 176 non-SLH angiosperms, of which 162 were placed on the phylogeny for comparative analyses.

Statistical framework

All analyses were conducted in R v4.4.1 (R Core Team, 2024). Two parallel analyses were performed: a climate analysis (584 non-SLH species; 39 families × 18 continuous predictors) and a trait analysis (176 non-SLH species, 162 with phylogenetic data; 39 families × 3 categorical and 12 continuous predictors). Both used non-parametric bivariate screening with joint FDR correction, followed by PGLS for significant associations and a common suite of sensitivity tests.

Spearman rank correlations were used for continuous predictors and Kruskal-Wallis (KW) tests for categorical predictors, with Dunn's post-hoc tests (BH-adjusted) for significant KW results. The climate

analysis comprised 702 Spearman tests (39 × 18), jointly BH-corrected. The trait analysis comprised 585 tests (117 KW + 468 Spearman), jointly BH-corrected. Associations with FDR-adjusted $p < 0.05$ were considered significant.

Phylogenetic comparative analyses

A dated supertree was generated separately for each dataset with V.PhyloMaker2 (33), placing species onto the GBOTB backbone (34) under Scenario 3; 460 tips for the climate dataset and 162 for the trait dataset. Phylogenetic signal was quantified with Blomberg's K (999 permutations; phytools).

PGLS was fitted with caper, estimating Pagel's λ by maximum likelihood, with BUSCO completeness as a covariate. Each model was fitted on the complete-case subset for its predictors, with the supertree trimmed to match and the data frame restricted to formula-relevant columns to avoid spurious NA-omission.

Multiple-testing correction for the trait PGLS was applied in two tiers, reflecting the distinction between confirmatory and exploratory tests. The focal hypothesis concerned LysM-RLK, identified a priori, and was first evaluated within the 12 principal immune receptor families - the major NLR classes (NLR total, CNL, TNL, RNL, ADR1, NRG1) and the major top-level cell-surface receptor families (IR-RLK, LRR-RLK, LysM-RLK, WAK-RLK, LRR-RLP, LysM-RLP) - together with the full suite of sensitivity analyses. The analysis was then extended to all 39 annotated families, of which the 27 additional families are predominantly fine-grained LRR-RLK subclasses nested within the LRR-RLK aggregate; because these sub-partitions are not statistically independent of one another or of the aggregate, joint Benjamini-Hochberg FDR correction across all 39 families is conservative. We therefore report both the within-12-family result for the confirmatory LysM-RLK test and the joint 39-family result for the exploratory sweep, applying the same two-tier scheme to the growth-form predictor (tree vs herb). Families significant only in the 39-family non-parametric sweep are treated as hypothesis-generating.

Lineage analysis

Because monocots essentially lack TNL genes and lineage composition covaries with latitude, the CNL proportion [CNL / (CNL + TNL)] was compared across lineages by KW test, and the CNL-absolute latitude Spearman correlation was recomputed within each lineage, to test whether any latitudinal CNL signal reflected a taxonomic confound.

Sensitivity analyses

The primary trait model was refit under three Pagel κ branch-length transformations (κ = 0.5, 1.0, 1.5) with λ co-estimated. All key models were rerun with SLH species included. Four further tests were applied to the primary model log(LysM-RLK + 1) ~ woodiness + BUSCO; a family-level subsampling bootstrap (100 iterations, one species per plant family); a domestication check (excluding domesticated

species and, separately, adding domestication as a covariate); a genome-size/polyploidy check (adding log assembled genome size as a covariate); and ancestral state reconstruction (phytools::fastAnc, Brownian motion), with per-tip branch changes compared between woody and non-woody species by Wilcoxon rank-sum test. As a quality control, BUSCO completeness was correlated with all predictors and with each receptor family count.

Fungal pathogen exposure

To test whether ambient fungal pathogen pressure explains any woodiness-LysM-RLK association, two per-species exposure estimates were derived. Occurrence records for 33 fungal and oomycete genera containing common plant pathogens were retrieved from GBIF (257,426 records); for each plant species, pathogen genus richness, total records, and record density were computed within the species' bounding box. Soil fungal community data were obtained from GlobalFungi v5 (35); 84,972 georeferenced samples × 4,424 genera); 28 of 33 target genera were present (the five absent were oomycetes, undetectable by ITS metabarcoding). Per-sample pathogen richness, relative abundance, and Shannon diversity were averaged across GlobalFungi samples within each species' bounding box (expanded 0.5°). All metrics were log-transformed.

Three sequential tests were performed on non-SLH angiosperms: Wilcoxon rank-sum tests comparing each metric between woody and non-woody species (n = 112; 45 woody, 67 non-woody); PGLS mediation tests adding each metric to log(LysM-RLK + 1) ~ woodiness + BUSCO (n = 112), with log(NLR + 1) and log(CNL + 1) as secondary responses; and within-woody PGLS and Spearman correlations (n = 45; 9 tests jointly BH-corrected).

Results

Climate and geographic associations do not survive phylogenetic correction

Across 584 non-SLH angiosperms and all 39 receptor families, bivariate screening identified 56 of 702 Spearman correlations as significant after joint FDR correction. The most significant correlations based on p values are shown in Table 1. Effect sizes were uniformly modest ($|\rho| \leq 0.24$). The strongest associations were negative correlations between temperature variables and the helper-NLR families TNL, NRG1, and RNL (e.g. annual mean temperature vs TNL, $\rho = -0.24$; minimum temperature of coldest month vs NRG1, $\rho = -0.24$), and positive correlations of absolute latitude with RNL and NRG1 ($\rho = +0.18$, + 0.17). Extending the analysis to all families added a scatter of significant correlations among individual LRR-RLK subclasses (e.g. isothermality vs LRR-RLK VII-1, $\rho = +0.20$), but these were of similar modest magnitude and did not form a coherent climatic pattern.

These associations did not survive phylogenetic correction. In PGLS models with maximum-likelihood Pagel's λ between roughly 0.5 and 0.9, none of the key climate associations remained significant (all $p > 0.14$), including NLR ~ annual mean temperature, CNL ~ absolute latitude, and NLR ~ absolute latitude,

shown in Table 2. The pattern indicates that apparent climate signals are attributable to shared ancestry rather than to a direct environmental effect on immune gene number. Notably, the direction of the strongest raw correlations was opposite to the leading hypothesis: warmer climates were associated with fewer, not more, helper-NLR genes.

Table 1
Representative significant Spearman correlations between climate/geographic predictors and receptor families (56 of 702 tests significant at FDR $p < 0.05$; $n = 584$).

Predictor	Family	ρ	p (raw)	p (FDR)
Annual mean temp (median)	TNL	−0.237	< 0.001	< 0.001
Min temp coldest month (median)	NRG1	−0.240	< 0.001	< 0.001
Annual mean temp (median)	RNL	−0.232	< 0.001	< 0.001
Isothermality (median)	LRR-RLK VII-1	+ 0.199	< 0.001	< 0.001
Absolute latitude	RNL	+ 0.178	< 0.001	< 0.001
Absolute latitude	NRG1	+ 0.170	< 0.001	0.001
Absolute latitude	CNL	−0.168	< 0.001	0.002
Absolute latitude	LysM-RLK	−0.126	0.002	0.035

Table 2
PGLS models for key climate associations (Pagel's λ by ML; $n = 460$). All associations attenuate to non-significance.

Model	λ	R^2	p (main)
NLR ~ annual mean temp + BUSCO	0.685	0.022	0.51
CNL ~ absolute latitude	0.703	0.001	0.64
LysM-RLK ~ annual mean temp + BUSCO	0.532	0.128	0.21
NLR ~ absolute latitude + BUSCO	0.686	0.021	0.97
TNL ~ absolute latitude + BUSCO	0.774	0.011	0.15

A latitudinal CNL signal reflects a monocot confound

The negative CNL-latitude correlation in the full dataset ($\rho = -0.17$, FDR $p < 0.01$) was driven by lineage composition. As seen in Fig. 1, Monocots, essentially lack TNL genes (median TNL = 0 vs 36 in eudicots) and have a median CNL proportion of 1.00. Monocots were also disproportionately represented at lower latitudes. The CNL proportion differed strongly across lineages (KW $\chi^2 = 149.2$, df = 2, $p < 10^{-32}$). Within eudicots the CNL-latitude correlation was negative ($\rho = -0.19$, $p < 0.001$), but within monocots it reversed sign ($\rho = +0.36$, $p = 0.002$), confirming the full-dataset signal as a lineage composition artefact.
[Fig. 1: CNL proportion by plant lineage.]

Figure 1 Legend: CNL proportion [CNL / (CNL + TNL)] by plant lineage across non-SLH angiosperms (eudicot n = 387, basal angiosperm n = 1, monocot n = 72). Monocots lack TNL genes (median proportion = 1.00), whereas eudicots show a broad spread. The proportion differs strongly among lineages (Kruskal–Wallis $p = 3.95 \times 10^{-33}$). Boxes show median and interquartile range, whiskers extend to 1.5× IQR, and points are individual species (jittered). This lineage structure underlies the apparent CNL–latitude correlation, which reverses sign within lineages.

Trait associations centre on LysM-RLK and Malectin-RLK

Of 585 trait-receptor tests, eleven survived joint FDR correction. The strongest were Kruskal-Wallis associations of LysM-RLK with woodiness and growth form, seen in Table 3. The same two growth traits also linked to Malectin-RLK and to several LRR-RLK subclasses (LRR-RLK VII.3, XI, Xa, V). These links, as depicted in Figs. 2 and 3, show that woody and tree species carry more receptor genes except LRR-RLK V, which was higher in non-woody species (LysM-RLK with woodiness FDR $p = 0.004$). No continuous leaf or stem trait was significantly associated with any receptor family. Post-hoc Dunn tests localised the categorical effects to woody > non-woody for LysM-RLK ($z = 4.48$, $p < 0.001$) and tree > herb for both LysM-RLK ($z = 4.27$, $p < 0.001$) and Malectin-RLK ($z = 3.92$, $p < 0.001$).

I evaluated the LysM-RLK woodiness association under PGLS in two tiers. In the *a priori* test within the 12 principal immune receptor families, LysM-RLK was significantly elevated in woody species ($\lambda = 0.41$, woody coefficient + 0.21, $p = 0.007$), and this signal held across the full suite of sensitivity analyses (below). I then extended the model to all 39 annotated families to ask whether any additional family also associated with woodiness. Several did at the raw level, predominantly fine-grained LRR-RLK subclasses (LRR-RLK XI $p = 0.004$, VII.3 $p = 0.017$, Xa $p = 0.021$), with LysM-RLK among the strongest. Under joint Benjamini-Hochberg correction across all 39 families, however, no family retained significance (LysM-RLK and LRR-RLK XI closest, FDR $p = 0.13$), as seen in Table 4. Because the 27 added families are largely LRR-RLK sub-partitions nested within the LRR-RLK aggregate, this 39-family correction is conservative; the LysM-RLK association is therefore robust as a pre-specified confirmatory test but does not survive genome-wide correction once the full, partly redundant family set is included.

The growth-form (tree vs herb) PGLS gave a similar result: LysM-RLK showed the strongest tree-versus-herb association (tree coefficient + 0.23, raw $p = 0.007$; Table 4), but no family survived FDR correction across all 39 families (LysM-RLK closest, FDR $p = 0.13$). Neither the woody/non-woody nor the tree-versus-herb contrast therefore reaches genome-wide significance for any family after phylogenetic correction and multiple-testing control. Because trees are a subset of woody plants, the two predictors describe overlapping biology and behave near-identically. LysM-RLK is the strongest and most consistent signal under both, but the failure of either contrast to clear the stringent all-family correction means the growth-form association, while directionally robust, is modest at the genome-wide scale.

Table 3
Significant Kruskal-Wallis trait associations (11 of 585 tests) and Dunn post-hoc results.

Family	Predictor	KW stat	p (FDR)	n
LysM-RLK	Woodiness	15.78	0.018	113
LysM-RLK	Growth form	19.53	0.018	131
Malectin-RLK	Growth form	18.64	0.018	131

Dunn post-hoc: woody > non-woody (LysM-RLK) $z = 4.48$, (Malectin-RLK) $z = 4.13$, $p < 0.001$; tree > herb (LysM-RLK) $z = 4.27$, (Malectin-RLK) $z = 3.92$, both $p < 0.001$.

Table 4
All-family PGLS for the leading families (Pagel's λ by ML; n = 162; FDR across all 39 families). No family survives joint FDR correction (p (FDR) > 0.05).

Family	Predictor	coefficient	p (raw)	p (FDR)
LysM-RLK	woody	+ 0.212	0.0066	0.13 (ns)
Malectin-RLK	woody	+ 0.201	0.0777	0.34 (ns)
LRR-RLK XI	woody	+ 0.260	0.0040	0.13 (ns)
LysM-RLK	tree	+ 0.230	0.0073	0.13 (ns)
Malectin-RLK	tree	+ 0.124	0.3151	0.43 (ns)

Although it fell short of the genome-wide threshold, the woody/non-woody association for LysM-RLK was internally robust to every confound examined. As shown in Table 5, the raw PGLS significance was retained under all three branch-length transformations ($\kappa = 0.5–1.5$; $p = 0.009 – 0.003$), with SLH species included ($p = 0.011$), after excluding domesticated species ($p = 0.007$), and after adding log genome size as a covariate ($p = 0.010$; genome size itself was also significant, $p < 0.001$); a family-level bootstrap (one species per family) returned $p < 0.05$ in 84 of 100 iterations. BUSCO completeness was a strong positive predictor of gene counts across families, supporting its inclusion as a covariate. Ancestral state reconstruction was inconclusive as to whether the pattern reflects gene gains in woody lineages or losses in herbaceous lineages (tip mean 2.105 vs reconstructed ancestral mean 2.106). The LysM-RLK signal is therefore not attributable to branch-length scaling, taxon sampling, domestication, or ploidy; rather, it reflects a genuine but modest growth-form effect that does not reach genome-wide significance once tested against all 39 receptor families.

Table 5
Sensitivity analyses for the LysM-RLK woodiness model log(LysM-RLK + 1) ~ woodiness + BUSCO.

Sensitivity test	Result	p (woodiness)
κ = 0.5 (speciational)	significant	0.0092
κ = 1.0 (standard)	significant	0.0066
κ = 1.5 (gradual)	significant	0.0031
SLH species included (n = 177)	significant	0.011
Family bootstrap (100 iter.)	84% significant	median 0.0027
Domesticated species excluded	significant	0.0068
+ log(genome size) covariate	significant	0.0100

[Fig. 2: LysM-RLK and Malectin-RLK counts by woodiness]

Figure 2 legend: log(LysM-RLK + 1) gene counts by woodiness (non-woody n = 102, woody n = 74) across non-SLH angiosperms. Woody species carry significantly more LysM-RLK genes than non-woody species (Dunn post-hoc woody > non-woody, $p < 0.001$). Boxes show median and interquartile range, whiskers extend to 1.5× IQR, and points are individual species (jittered). The effect is robust to confounds and significant as an *a priori* test within the 12 principal families, but modest and sub-threshold under conservative correction across all 39 families.

[Fig. 3: LysM-RLK and Malectin-RLK counts by growth form]

Figure 3 legend: log(LysM-RLK + 1) gene counts by growth form (tree n = 54, shrub n = 20, herb n = 102) across non-SLH angiosperms. Tree and shrub species carry more LysM-RLK genes than herbs (Dunn post-hoc tree > herb, $p < 0.001$). Boxes show median and interquartile range, whiskers extend to 1.5× IQR, and points are individual species (jittered). The association is significant in non-parametric tests but does not survive phylogenetic correction across all 39 receptor families (see text).

Fungal pathogen exposure does not explain the LysM-RLK signal

Ambient fungal pathogen exposure, estimated from 257,426 GBIF pathogen occurrence records and 84,972 GlobalFungi soil samples, did not account for the woodiness-LysM-RLK association. As shown in Fig. 4, woody plants did not experience greater pathogen pressure: GlobalFungi and GBIF pathogen richness did not differ by woodiness, and the only significant difference (relative abundance of soil pathotrophic fungi) was lower in woody plants ($p = 0.026$), the opposite of the direction required for mediation. Adding any pathogen metric to the baseline PGLS left the woodiness coefficient essentially unchanged and all pathogen terms non-significant (all $p > 0.41$). Within woody species alone (n = 45), no pathogen metric predicted LysM-RLK, NLR, or CNL counts after FDR correction. The LysM-RLK signal therefore appears to reflect an intrinsic property of woody or tree growth rather than differential exposure to the ambient fungal pathogen environment.

[Fig. 4: fungal pathogen exposure by growth form]

Figure 4 legend: Ambient fungal pathogen exposure by growth form (tree n = 52, shrub n = 20, herb n = 102) for three metrics: GlobalFungi pathogen richness (log), GlobalFungi pathogen relative abundance (log), and GBIF pathogen richness (log). Where differences occur, herbs show higher, not lower, exposure than trees (e.g., richness and relative abundance), opposite to the direction required if pathogen pressure explained elevated receptor counts in trees. Boxes show median and interquartile range; points are individual species (jittered). Brackets give pairwise Wilcoxon comparisons (ns, not significant; *p < 0.05; **p < 0.01).

Discussion

This study tested a long-standing macroecological expectation - that plants under greater pathogen pressure, indexed by warm, wet, low-latitude climates, carry larger immune receptor repertoires - across angiosperms and all 39 annotated receptor families. The expectation was not supported. Although bivariate screening recovered dozens of significant climate-receptor correlations, every association was modest and none survived phylogenetic correction, indicating that the raw signals reflect shared ancestry among climatically similar relatives rather than a direct environmental effect on immune gene number. The direction of the strongest correlations was itself contrary to the hypothesis: warmer climates were associated with fewer, not more, helper-NLR (TNL, NRG1, RNL) genes. These results add to a growing recognition that apparent macroecological gradients in genomic traits must be considered with appropriate phylogenetic controls, and they caution against interpreting cross-species correlations between immune gene counts and climate as adaptive without relevant phylogenetic controls (36–38).

The latitudinal CNL signal provides a concrete example of how taxonomic structure can masquerade as an environmental gradient. Monocots essentially lack TNL genes (instead carrying TIR-only and TIR-NBARC-TPR proteins) and were over-represented at lower latitudes in the dataset (39–41); the negative CNL-latitude correlation observed across all species reversed sign within eudicots and monocots considered separately. The full-dataset pattern was therefore a composition effect, not evidence that latitude shapes the CNL/TNL balance. This mirrors classic concerns about Simpson's paradox in comparative data and reinforces the need to examine within-lineage patterns before attributing immune variation to climate.

Against this largely negative backdrop, the most consistent ecological correlate of immune gene content was growth form rather than climate. Woody and tree species carried more LysM-RLK genes than herbs under both non-parametric and phylogenetic tests, and in the *a priori* test within the 12 principal receptor families this signal was internally robust - retained under branch-length transformations, taxon subsampling, exclusion of domesticated crops, and control for genome size. Its statistical standing depends, however, on how multiple testing is handled. Evaluated within the principal families, or as the single a priori hypothesis it was, LysM-RLK is significant ($p = 0.007$); corrected jointly across all 39 annotated families it is not (FDR $p = 0.13$). Because the 27 additional families are largely LRR-RLK sub-partitions that are not independent of one another, the 39-family correction is conservative, and thus neither framing is decisive on its own. With the addition of further species the effect size also declined

(woody coefficient + 0.29 to + 0.21), suggesting the association concentrates in the clearest woody and herbaceous cases. I therefore interpret the result conservatively: there is a real but modest tendency for arborescent, long-lived growth forms to carry expanded LysM-RLK repertoires, statistically robust as a focal hypothesis but not at genome-wide scale. Notably, several other cell-surface receptor families - Malectin-RLK and several LRR-RLK subclasses - were non-parametrically elevated in woody and tree species in the same direction, while one (LRR-RLK V) ran inversely, being higher in non-woody plants. The coherence of direction across multiple structurally distinct surface-receptor families, set against this single inversion, argues against a trivial explanation in which woody plants simply carry more of every gene, and tentatively suggests a broader pattern of expanded cell-surface immune investment in woody tissues. These additional families did not survive phylogenetic correction and are reported as hypothesis-generating rather than confirmatory.

Why might tree growth covary with LysM-RLK and Malectin-RLK content? One possibility is that the persistent, long-lived tissues of trees impose sustained surveillance demands that select for expanded surface-receptor repertoires (42,43). Long-lived trees can rely on more dynamic methods to recognize and counter pathogens through larger gene numbers, greater mutation loads accumulated over longer lifespans that are passed down to offspring, and having multiple colinear clustered R-genes that facilitate those random mutations that give rise to diversity in defense genes (44). An alternative explanation, however, is simply that trees live where more pathogens are. Using two exposure datasets, I found that woody species did *not* experience greater ambient fungal pathogen pressure; the only significant difference ran in the opposite direction, and adding pathogen exposure to the model left the growth-form coefficient unchanged. Differential exposure to the ambient soil fungal environment therefore does not explain the association. I note two limitations: soil metabarcoding underrepresents obligate aerial pathogens (rusts, mildews, smuts) that are especially relevant to trees, and the within-woody sample was small (n = 45), so I cannot exclude a role for host-specific or aerial pathogen pressure not captured here.

A central ambiguity remains unresolved. LysM-RLK genes function both in chitin-triggered immunity (via LYK-type receptors) and in establishing mycorrhizal and rhizobial symbioses (via LYR-type receptors) (45,46). An expansion of LysM-RLK in trees is therefore consistent with either greater investment in fungal immunity or greater investment in symbiotic signalling, and the aggregate counts available here cannot distinguish them. I was unable to resolve this with the present data for two structural reasons: the source annotation does not separate LYK from LYR subfamilies, and a species-level mycorrhizal-type analysis is confounded in this dataset because ectomycorrhizal associations are almost entirely restricted to woody lineages while the dataset is overwhelmingly arbuscular-mycorrhizal. Distinguishing the immunity and symbiosis hypotheses will require LYK/LYR subfamily annotation from primary protein sequences, or targeted sampling that decouples mycorrhizal type from growth form. The direction of change is also unresolved: ancestral state reconstruction could not determine whether trees have gained LysM-RLK genes or herbs have lost them, which would require denser phylogenetic sampling.

Several constraints bound the conclusions. Immune gene counts depend on assembly and annotation quality; I controlled for BUSCO completeness throughout and found it to be a strong predictor of gene counts, underscoring that cross-species count comparisons are sensitive to genome quality. Gene counts are also a coarse proxy for immune capacity, ignoring sequence diversity, expression, and functional specialisation. Finally, the trait analyses rested on complete-case subsets of a few hundred species, limiting power for families with sparse counts or unevenly distributed traits. Within these limits, the central message is twofold: climate and geography are not robust predictors of angiosperm immune receptor content once phylogeny is accounted for, and growth form - specifically the trees - is a more consistent, if still modest, correlate whose mechanistic basis, immune or symbiotic, is now the key open question.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The author declares that he has no competing interests.

Funding

The author is supported by the Lewis B Cullman Fellowship, awarded jointly by Yale University and New York Botanical Garden.

Author Contribution

ST conceptualized the study and performed the analysis. ST wrote the manuscript.

Acknowledgements

Not applicable

Data Availability

All data generated or analysed during this study are included in this published article and described in the supplementary information files. Occurrence retrieval and climate extraction required Python 3.12 (requests, rasterio 1.3.8, pandas, numpy). Statistical analysis required R 4.4.1 (tidyverse, corrplot, rstatix, ape, caper, phytools, V.PhyloMaker2). Data sources include WorldClim v2.1, GBIF, GIFT, TRY, GlobalFungi v5. Access to analysis scripts is explained in the Supplementary Information.

References

1. Yuan M, Jiang Z, Bi G, Nomura K, Liu M, Wang Y, et al. Pattern-recognition receptors are required for NLR-mediated plant immunity. Nature. 2021 Mar 10;592(7852):105–109.
2. Čerekovic N, Poltronieri P. Plant signaling pathways activating defence response and interfering mechanisms by pathogen effectors, protein decoys and bodyguards. AIMS Molecular Science. 2017 Jan 1;4(3):370–388.
3. Saijo Y, Loo EP, Yasuda S. Pattern recognition receptors and signaling in plant-microbe interactions. The Plant Journal. 2018 Feb 1;93(4):592–613.
4. Couto D, Zipfel C. Regulation of pattern recognition receptor signalling in plants. Nat Rev Immunol. 2016 Aug 1;16(9):537–552.
5. Kourelis J, Adachi H. Activation and Regulation of NLR Immune Receptor Networks. Plant and Cell Physiology. 2022 Aug 9;63(10):1366–1377.
6. Shepherd S, Yuen ELH, Carella P, Bozkurt TO. The wheels of destruction: Plant NLR immune receptors are mobile and structurally dynamic disease resistance proteins. Current Opinion in Plant Biology. 2023 May 10;74:102372.
7. Yuan M, Cai B, Xin XF. Plant immune receptor pathways as a united front against pathogens. PLoS Pathog. 2023 Feb 9;19(2):e1011106.
8. Li SX, Liu Y, Zhang YM, Chen JQ, Shao ZQ. Convergent reduction of immune receptor repertoires during plant adaptation to diverse special lifestyles and habitats. Nat Plants. 2025 Jan 16;11(2):248–262.
9. Holland PWH, Marlétaz F, Maeso I, Dunwell TL, Paps J. New genes from old: asymmetric divergence of gene duplicates and the evolution of development. Phil Trans R Soc B. 2017 Feb 5;372(1713):20150480.
10. Reams AB, Roth JR. Mechanisms of gene duplication and amplification. Cold Spring Harb Perspect Biol. 2015 Feb 1;7(2):a016592.
11. Milici VR, Dalui D, Mickley JG, Bagchi R. Responses of plant–pathogen interactions to precipitation: Implications for tropical tree richness in a changing world. Journal of Ecology. 2020 Mar 17;108(5):1800–1809.
12. Suh DC, Lance SL, Park AW. Abiotic and biotic factors jointly influence the contact and environmental transmission of a generalist pathogen. Ecology and Evolution. 2024 Aug 1;14(8):e70167.

13. Hiremath SS, Kelkar GD, Kurhade KC, Kumar A, Dhiman S, Nigam R, et al. Impact of Climate Change on Plant Disease Dynamics and Pathogen Distribution: A Review. J Adv Microbiol. 2026 Mar 2;26(2):176–194.
14. Downie AE, Tung J. Evolutionary genetics meets ecological immunology: insights into the evolution of immune systems. Current Opinion in Genetics & Development. 2025 Dec 1;95:102411.
15. Lazzaro BP, Little TJ. Immunity in a variable world. Phil Trans R Soc B. 2008 Oct 16;364(1513):15–26.
16. Jacob F, Vernaldi S, Maekawa T. Evolution and Conservation of Plant NLR Functions. Front Immunol. 2013 Jan 1;4:297.
17. Coley PD, Bryant JP, Chapin FS. Resource Availability and Plant Antiherbivore Defense. Science. 1985 Nov 22;230(4728):895–899.
18. Moreira X, Zas R, Solla A, Sampedro L. Differentiation of persistent anatomical defensive structures is costly and determined by nutrient availability and genetic growth-defence constraints. Tree Physiology. 2015 Jan 15;35(2):112–123.
19. Herms DA, Mattson WJ. The Dilemma of Plants: To Grow or Defend. The Quarterly Review of Biology. 1992 Sept 1;67(3):283–335.
20. Michelmore RW, Meyers BC. Clusters of Resistance Genes in Plants Evolve by Divergent Selection and a Birth-and-Death Process. Genome Res. 1998 Nov 1;8(11):1113–1130.
21. Van De Weyer AL, Monteiro F, Furzer OJ, Nishimura MT, Cevik V, Witek K, et al. The *Arabidopsis thaliana* pan-NLRome. bioRxiv (Cold Spring Harbor Laboratory). Royal Society; 2019.
22. Stam R, Silva-Arias GA, Tellier A. Subsets of NLR genes show differential signatures of adaptation during colonization of new habitats. New Phytologist. 2019 Aug 8;224(1):367–379.
23. Teasdale LC, Murray KD, Collenberg M, Contreras-Garrido A, Schlegel T, Van Ess L, et al. Pangenomic context reveals the extent of intraspecific plant NLR evolution. Cell Host & Microbe. 2025 Aug 1;33(8):1291-1305.e9.
24. Liang X, Dong J. Comparative-genomic analysis reveals dynamic NLR gene loss and gain across Apiaceae species. Front Genet. 2023 Mar 2;14(7):1141194.
25. Hao Y, Pan Y, Chen W, Rashid MAR, Li M, Che N, et al. Contribution of Duplicated Nucleotide-Binding Leucine-Rich Repeat (NLR) Genes to Wheat Disease Resistance. Plants. 2023 July 27;12(15):2794.
26. Keepers K, Peterson K, Raduski A, Turner KM, Van Tassel D, Smith K, et al. Disease resistance gene count increases with rainfall in Silphium integrifolium. Ecology and Evolution. 2024 Sept 1;14(9):e11143.
27. Rouse MN, Saleh AA, Seck A, Keeler KH, Travers SE, Hulbert SH, et al. Genomic and resistance gene homolog diversity of the dominant tallgrass prairie species across the U.S. Great Plains precipitation gradient. PLoS ONE. 2011 Apr 12;6(4):e17641.
28. Schemske DW, Mittelbach GG, Cornell HV, Sobel JM, Roy K. Is There a Latitudinal Gradient in the Importance of Biotic Interactions? Annu Rev Ecol Evol Syst. 2009 Dec 1;40(1):245–269.

29. Desaki Y, Miyata K, Suzuki M, Shibuya N, Kaku H. Plant immunity and symbiosis signaling mediated by LysM receptors. Innate Immun. 2017 Nov 6;24(2):92–100.
30. Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Intl Journal of Climatology. 2017 May 15;37(12):4302–4315.
31. Gliesch M, Dahlin K, Berdugo-Lattke M, Blonder B, Munoz F, Aubin I, et al. TRY plant trait database - enhanced coverage and open access. Global Change Biology. 2019 Dec 31;26(1):119–188.
32. Kattge J, Boenisch G, Diaz S, Lavorel S, Prentice C, Leadley P, et al. The TRY Plant Trait Database - enhanced coverage and open access. 2020 Mar 23;
33. Jin Y, Qian H. V.PhyloMaker2: An updated and enlarged R package that can generate very large phylogenies for vascular plants. Plant Diversity. 2022 May 27;44(4):335–339.
34. Smith SA, Brown JW. Constructing a broadly inclusive seed plant phylogeny. American J of Botany. 2018 Feb 14;105(3):302–314.
35. Větrovský T, Morais D, Kohout P, Lepinay C, Algora C, Awokunle Hollá S, et al. GlobalFungi, a global database of fungal occurrences from high-throughput-sequencing metabarcoding studies. Sci Data. 2020 July 13;7(1):228.
36. Freckleton RP, Harvey PH, Pagel M. Phylogenetic Analysis and Comparative Data: A Test and Review of Evidence. The American Naturalist. 2002 Dec 1;160(6):712–726.
37. Bureš P, Elliott TL, Veselý P, Šmarda P, Forest F, Leitch IJ, et al. The global distribution of angiosperm genome size is shaped by climate. bioRxiv (Cold Spring Harbor Laboratory). Royal Society; 2022.
38. Cacho NI, Mcintyre PJ, Kliebenstein DJ, Strauss SY. Genome size evolution is associated with climate seasonality and glucosinolates, but not life history, soil nutrients or range size, across a clade of mustards. Annals of Botany. 2021 Mar 1;127(7):887–902.
39. Liu Y, Zeng Z, Zhang YM, Li Q, Jiang XM, Jiang Z, et al. An angiosperm NLR Atlas reveals that NLR gene reduction is associated with ecological specialization and signal transduction component deletion. Molecular Plant. 2021 Aug 4;14(12):2015–2031.
40. Guo BC, Zhang YR, Liu ZG, Li XC, Yu Z, Ping BY, et al. Deciphering Plant NLR Genomic Evolution: Synteny-Informed Classification Unveils Insights into TNL Gene Loss. Molecular biology and evolution. 2025 Jan 21;42(2).
41. Tang CQ, Orme CDL, Bunnefeld L, Jones FA, Powell S, Chase MW, et al. Global monocot diversification: geography explains variation in species richness better than environment or biology. Bot J Linn Soc. 2016 Oct 1;183(1).
42. Fiorucci AS. To Grow or Defend? More on the Plant Cornelian Dilemma. Plant Physiol. 2020 June 1;183(2):437–438.
43. Batalova AY, Krutovsky KV. Genetic and Epigenetic Mechanisms of Longevity in Forest Trees. IJMS. 2023 June 20;24(12):10403.
44. Tobias PA, Guest DI. Tree immunity: growing old without antibodies. Trends in Plant Science. 2014 Feb 18;19(6):367–370.

45. Gough C, Cottret L, Lefebvre B, Bono JJ. Evolutionary History of Plant LysM Receptor Proteins Related to Root Endosymbiosis. Front Plant Sci. 2018 July 4;9(e26114):923.
46. Ding Y, Lesterps Z, Gasciolli V, Fuchs AL, Gaston M, Medioni L, et al. Several groups of LysM-RLKs are involved in symbiotic signal perception and arbuscular mycorrhiza establishment. Nat Commun. 2025 July 1;16(1):5999.

Figures

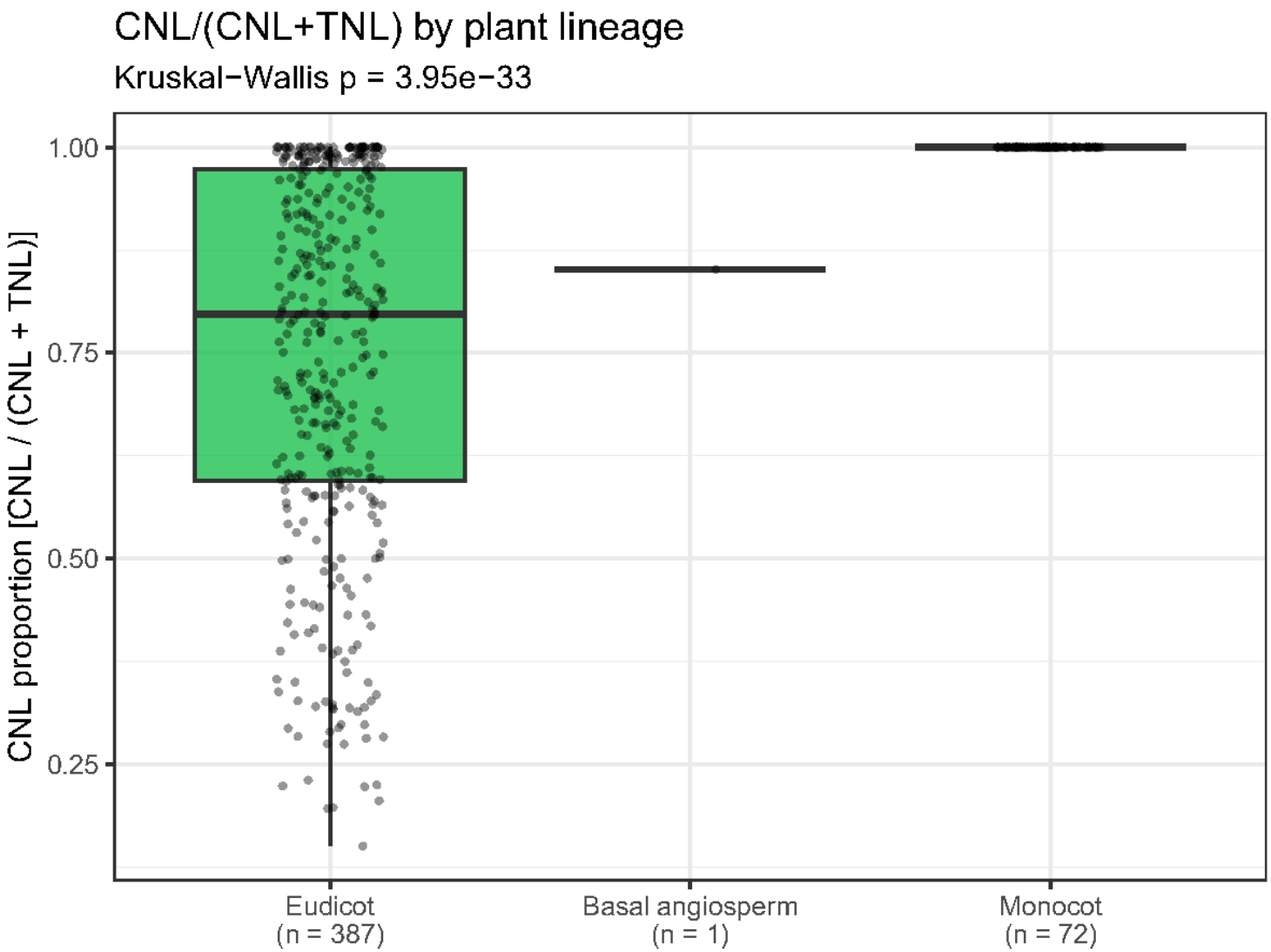


Figure 1

CNL proportion by plant lineage.

Figure 1 Legend: CNL proportion [CNL / (CNL + TNL)] by plant lineage across non-SLH angiosperms (eudicot n = 387, basal angiosperm n = 1, monocot n = 72). Monocots lack TNL genes (median proportion = 1.00), whereas eudicots show a broad spread. The proportion differs strongly among lineages (Kruskal–Wallis p = 3.95×10^{-33}). Boxes show median and interquartile range, whiskers extend

to 1.5× IQR, and points are individual species (jittered). This lineage structure underlies the apparent CNL–latitude correlation, which reverses sign within lineages.

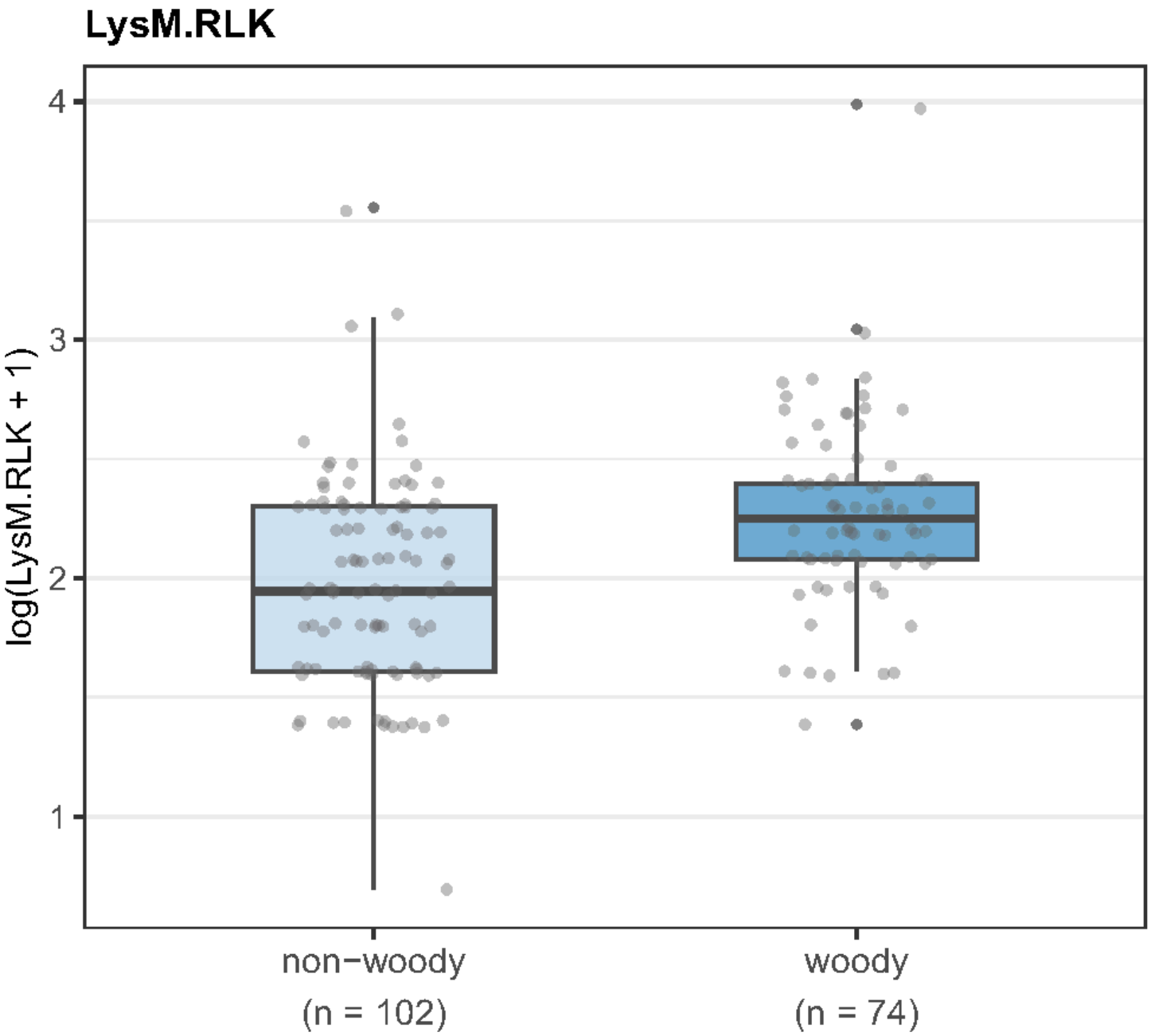


Figure 2

LysM-RLK and Malectin-RLK counts by woodiness

Figure 2 legend: log(LysM-RLK + 1) gene counts by woodiness (non-woody n = 102, woody n = 74) across non-SLH angiosperms. Woody species carry significantly more LysM-RLK genes than non-woody species (Dunn post-hoc woody > non-woody, $p < 0.001$). Boxes show median and interquartile range, whiskers extend to 1.5× IQR, and points are individual species (jittered). The effect is robust to confounds and

significant as an *a priori* test within the 12 principal families, but modest and sub-threshold under conservative correction across all 39 families.

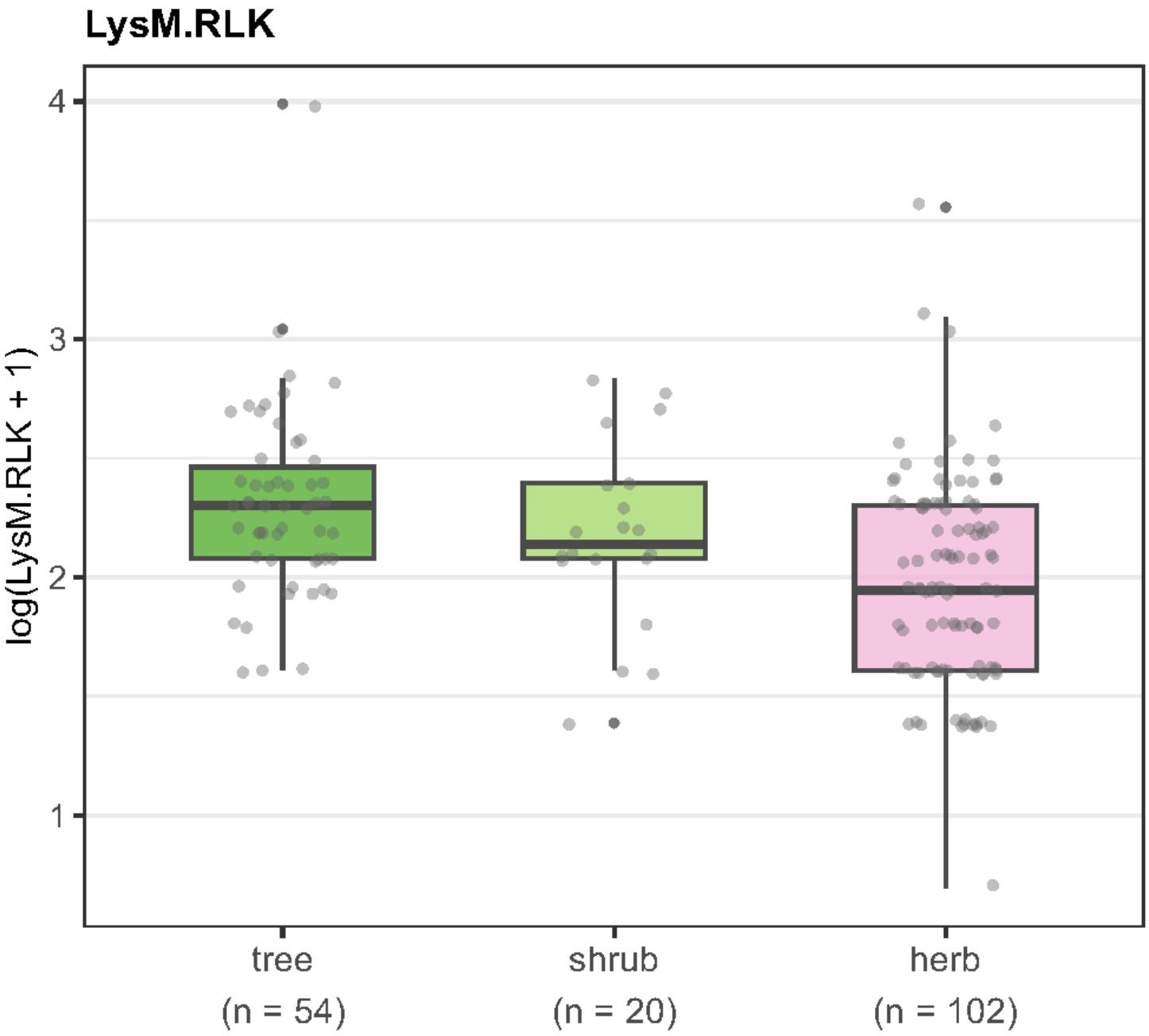


Figure 3

LysM-RLK and Malectin-RLK counts by growth form

Figure 3 legend: log(LysM-RLK + 1) gene counts by growth form (tree n = 54, shrub n = 20, herb n = 102) across non-SLH angiosperms. Tree and shrub species carry more LysM-RLK genes than herbs (Dunn post-hoc tree > herb, $p < 0.001$). Boxes show median and interquartile range, whiskers extend to 1.5× IQR, and points are individual species (jittered). The association is significant in non-parametric tests but does not survive phylogenetic correction across all 39 receptor families (see text).

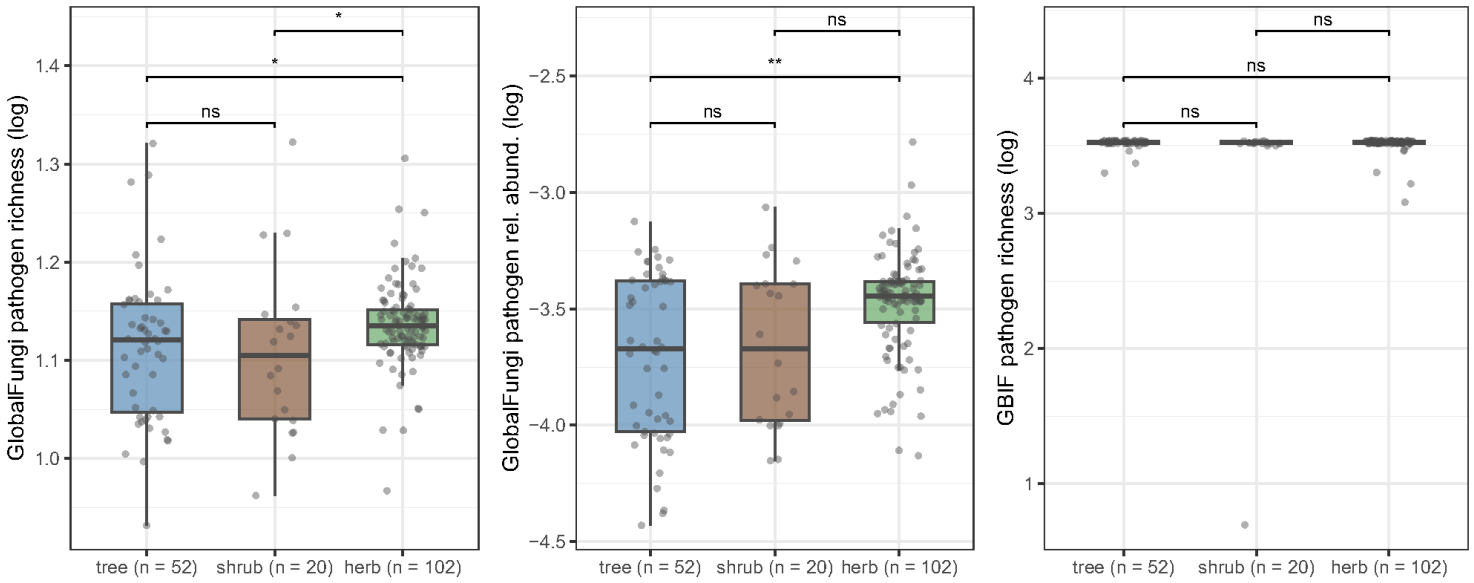


Figure 4

fungal pathogen exposure by growth form

Figure 4 legend: Ambient fungal pathogen exposure by growth form (tree n = 52, shrub n = 20, herb n = 102) for three metrics: GlobalFungi pathogen richness (log), GlobalFungi pathogen relative abundance (log), and GBIF pathogen richness (log). Where differences occur, herbs show higher, not lower, exposure than trees (e.g., richness and relative abundance), opposite to the direction required if pathogen pressure explained elevated receptor counts in trees. Boxes show median and interquartile range; points are individual species (jittered). Brackets give pairwise Wilcoxon comparisons (ns, not significant; *p < 0.05; **p < 0.01).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- CondensedGrowthForm.csv