

Longevity stabilizes core gene families and reshapes *NLR* architecture in trees

Yuanpeng Fang^{1,2*} ¹ College of Agriculture, Shanxi Agricultural University, Taigu, Jinzhong 030801, China² State Key Laboratory of Sustainable Dryland Agriculture, Shanxi Agricultural University, Taiyuan 030031, China* Correspondence: fyp1290941655@yeah.net (Fang Y)

Abstract

Although plant lifespans vary across six orders of magnitude, the degree to which extended longevity is associated with differences in genome-wide gene repertoire dynamics remains largely obscure. In this study, we employed an alignment-based orthology framework to measure core genome stability within a phylogenetically stratified cohort of 40 species. Our results demonstrate a robust correlation between longevity and the preservation of core orthogroups; long-lived taxa maintain a more comprehensive ancestral gene toolkit, even as dispensable and lineage-specific genes undergo concurrent expansion. This pattern, in which the absolute number of core orthogroups is relatively stable while their proportional contribution decreases, is consistent with a dilution effect associated with variation in non-core gene repertoires. Functional interrogations reveal that these core orthogroups sustain fundamental cellular infrastructure—particularly ribosomal and chromatin-associated mechanisms—while allowing for nuanced lineage-specific diversification. Remarkably, prolonged vegetative longevity is also associated with a qualitative transformation in immune architecture: long-lived species showed higher *NLR* densities in the present dataset, whereas the evidence for lifespan-associated differences in TNL and RNL proportions was statistically uncertain. These patterns are distributed across major lineages of the plant tree of life and suggest that life-history differences may covary with the conservation of housekeeping-related genes and the composition of intracellular immune receptors. Collectively, our findings substantiate a macroevolutionary model wherein longevity is associated with a dual pattern of genomic organization: the conservation of essential cellular machinery coupled with the elaboration of specialized immune repertoires, thereby providing a comparative link between life-history variation and genome composition in land plants.

Keywords: Plant longevity, Genomic stasis, Core genome, Deep-time conservation, Macroevolution, Life history lifespan

Citation: Fang Y. 2026. Longevity stabilizes core gene families and reshapes *NLR* architecture in trees. *Forestry Research Advances* 1: e019 <https://doi.org/10.48130/fra-0026-0017>

Introduction

The emergence of pan-genomics has revolutionized plant genomics, moving the field from single-reference annotations toward a holistic examination of genomic diversity across species and populations^[1]. This shift is especially vital for forest trees, whose intricate life histories—marked by extended generation times, perennial growth patterns, and significant somatic plasticity—introduce unique challenges to genomic stability not seen in annual crops^[2]. Although early genomic studies have established a foundation, progressing from merely cataloging genetic content to deciphering the evolutionary forces that shape it remains a critical frontier in forestry research.

Current literature predominantly examines the evolutionary dynamics of specific gene families, such as Argonaute in immune responses or Cytochrome P450s in secondary metabolism^[3–5]. While these investigations offer high-resolution insights into localized adaptations, they encounter a 'curse of dimensionality' when extrapolated across deep phylogenetic divides. The manual curation and validation of orthologs, though methodologically rigorous, become statistically unfeasible when comparing dozens of species with divergent genome sizes and annotation qualities^[6]. As a result, our comprehension of how entire genomic repertoires—not just isolated gene clusters—adapt to life-history constraints remains fragmented.

Among the defining characteristics of tree biology, longevity represents an important life-history variable that may covary with genome evolution^[7,8]. Theoretical models of life-history evolution

suggest that long-lived organisms must curtail the accumulation of germline and somatic mutations through heightened genomic conservation^[9–11]. In plants, this implies a hypothesis wherein extended lifespans may be associated with reduced turnover or greater retention of core genetic repertoires—the essential genes required for basic cellular viability. In contrast, shorter-lived herbs may show different patterns of gene loss and gain, enabling swift adaptation to transient environments. Validating this macroevolutionary prediction necessitates a shift from reductionist methodologies to a comprehensive evaluation of gene repertoire dynamics across the plant kingdom.

To bridge this knowledge gap, we employed an alignment-based orthology inference framework to systematically analyze homologous gene families within a phylogenetically stratified dataset comprising 40 plant species, spanning from algae and mosses to long-lived gymnosperms and angiosperm trees^[12–14]. This approach emphasizes breadth and objectivity, enabling us to quantify the association between species longevity and core genome stability. Our findings demonstrate a strong negative correlation between organismal lifespan and the turnover rate of core orthogroups, supporting an association between longer maximum recorded lifespan and greater retention of core orthogroups. This study establishes a foundational macro-evolutionary benchmark for elucidating the genetic basis of tree longevity, paving the way for future functional inquiries into the preservation of genomic integrity over centuries.

Materials and methods

Phylogenetic sampling and genome acquisition

To capture the full breadth of the plant kingdom's evolutionary tapestry, we assembled a dataset of 40 species strategically positioned at pivotal phylogenetic junctures. This collection ranges from the ancestral green algae (*Chlamydomonas reinhardtii*) and early land-colonizing bryophytes (*Marchantia polymorpha*, *Sphagnum fallax*) to ferns (*Ceratopteris richardii*), basal angiosperms (*Amborella trichopoda*, *Nymphaea colorata*), and gymnosperms (*Ginkgo biloba*, *Pinus longaeva*), culminating in a diverse spectrum of eudicots that bridges the gap between enduring woody perennials and ephemeral annual herbs (Supplementary Table S1).

Genome assemblies alongside their annotated protein counterparts were sourced from the Genome Warehouse (GWH) within the National Genomics Data Center (NGDC, Beijing, China; <https://ngdc.cncb.ac.cn/gwh>). Selection criteria prioritized 'Chromosome-level' or 'High-quality Draft' assemblies equipped with functional annotations. To safeguard against analytical distortions caused by gene model redundancy or fragmentation, the dataset was refined to retain the primary transcript for each gene locus prior to downstream evaluation.

Genome quality assessment

The integrity and quality of the downloaded proteomes were evaluated using established metrics. We assessed completeness using Benchmarking Universal Single-Copy Orthologs (BUSCO, *viridiplantae_odb10*) v5.47.7 to quantify the proportion of conserved genes present in each assembly. Additionally, we examined the distribution of protein lengths and the total gene count per species to identify potential anomalies, such as gene model over-prediction or fragmentation. Assemblies exhibiting excessive redundancy or low BUSCO scores (< 80%) were excluded.

Inference of orthogroups and quantification of core repertoire stability

Orthogroups were inferred using OrthoFinder v3.1.2^[14]. The presence or absence of each orthogroup was recorded for every species based on the corresponding Orthogroups.tsv output. Orthogroup occupancy was calculated as the number of sampled species containing at least one gene assigned to that orthogroup.

Core orthogroups were defined as those present in at least 95% of the sampled species. For each species, core orthogroup retention was calculated as the proportion of core orthogroups present in that species:

$$\text{Core retention}_i = \frac{\text{Number of core orthogroups present in species } i}{\text{Total number of core orthogroups}}$$

Core repertoire deficit was calculated as:

$$\text{Core loss proportion}_i = 1 - \text{Core retention}_i$$

The associations between $\log_{10}$ -transformed lifespan and core retention or core loss proportion were tested using ordinary linear regression. Total orthogroup number was included as a covariate in adjusted models. To account for non-independence among species caused by shared evolutionary history, phylogenetic generalized least-squares models were fitted using a Brownian correlation structure based on the inferred species tree. Sensitivity analyses were conducted using core occupancy thresholds of 90%, 95%, and 99%.

Phylogenetic analysis and functional annotation

Phylogenomic inference was conducted using 39 near-single-copy orthologous genes identified by OrthoFinder v3.1.2^[14]. Each orthogroup was aligned with MAFFT v7.505 and trimmed with trimAl v1.4 (-automated1). IQ-TREE v2.2.0 was used to infer a maximum-likelihood phylogeny under the partitioned model (one partition per orthogroup), with ModelFinder selecting the best-fit substitution model and 1,000 ultrafast bootstrap replicates^[15].

Functional annotation of protein sequences was performed using eggNOG-mapper v2.1.12 against the eukaryotic orthology database (eggNOG v6.0)^[16]. Gene Ontology (GO, www.geneontology.org)^[17] terms and KEGG Orthology (KO, www.genome.jp/kegg)^[18] identifiers were extracted for downstream analyses. To identify significantly enriched functional categories, we performed hypergeometric tests comparing genes within target orthogroups against the background of all annotated genes in each species. Multiple testing corrections were applied using the Benjamini–Hochberg method, and terms with an adjusted p -value < 0.05 and at least three foreground genes were considered statistically significant.

To complement the broad GO enrichment analysis (based on enrichplot), we performed a focused annotation of core orthogroups for functions potentially relevant to longevity. Five functional modules were defined *a priori*: (1) DNA repair and genome stability, (2) telomere maintenance, (3) oxidative-stress response and redox homeostasis, (4) autophagy and proteostasis, and (5) chromatin and epigenetic regulation. Candidate orthogroups were identified by screening their GO terms, KEGG Orthology assignments, and eggNOG functional descriptions for module-specific annotations.

Longevity data compilation and statistical analysis

Species-specific longevity estimates (maximum recorded lifespan in years) were compiled from authoritative botanical databases, peer-reviewed literature, and institutional records (e.g., *Pinus longaeva* ~5,000 years; *Arabidopsis thaliana* ~1 year; Supplementary Table S2a, S2b). Due to the wide range of values spanning several orders of magnitude, longevity data were $\log_{10}$ -transformed prior to statistical testing.

Because the study used cross-species comparative data, all analyses were interpreted as tests of association rather than causal effects. We evaluated whether the association between maximum recorded lifespan and genome-repertoire traits remained after accounting for phylogenetic relatedness and selected covariates. To test the hypothesis that longevity is associated with reduced variation in core gene turnover, we performed linear regression analyses between $\log_{10}$ -transformed lifespan and the proportion of core orthogroups relative to the total orthogroup count per species. Pearson correlation coefficients (r) and associated p -values were calculated to assess the strength and significance of observed relationships. All statistical analyses and visualizations were conducted in R v4.2.0 using the ggplot2 and tidyverse packages.

To assess whether the observed associations exceeded expectations under random assignment of longevity values, we performed a tip-label permutation test. Longevity values were randomly permuted across the 40 species while the species tree, orthogroup matrix, and core orthogroup definition were held constant. For each of 10,000 permutations, we refitted the corresponding regression model and recorded the slope coefficient. The permutation p -value was calculated as the proportion of permuted absolute slopes equal to or greater than the observed absolute slope.

Identification of *NLR* gene families in the context of plant longevity

To assess whether long-lived plants compensate for constrained core genome turnover through expanded stress defenses, we profiled intracellular immune receptor repertoires across the 40-species dataset. *NLR*-encoding genes were identified using *NLR*-Annotator v2.1b, followed by protein-level validation with Resistify v1.3.0^[19–21]; only genes supported by both tools were retained for downstream analysis. To control for variation in total gene annotation effort, *NLR* repertoire size was normalized as *NLR* genes per 1,000 orthogroups:

$$NLR \text{ density} = \frac{\text{Number of } NLR \text{ genes}}{\text{Total number of orthogroups}} \times 1,000$$

NLR architectures were classified based on N-terminal domain organization: *TNL* (TIR domain present), *CNL* (coiled-coil domain present, TIR absent), *RNL* (RPW8 domain present), and other unclassifiable or truncated architectures. Relative proportions of *TNL* and *RNL* fractions were calculated as the number of genes in each class divided by the total *NLR* count per species.

Phylogenetic generalized least squares (PGLS) models were conducted using the rooted species tree inferred by OrthoFinder v3.1.2 under the STAG algorithm, which accounts for gene-tree conflict through multispecies coalescent-aware inference. A Brownian motion model of trait evolution was assumed, as justified by Pagel's λ estimation ($\lambda = 0.27$). Because the *TNL* and *RNL* fractions included zero values, proportions were transformed using the logit function after adding a pseudocount of 10^{-6} . To assess the sensitivity of the results to this choice, we repeated the PGLS analyses using pseudocounts of 10^{-3} , 10^{-4} , and 10^{-8} . Associations between $\log_{10}$ -transformed lifespan and *NLR* density, *TNL* fraction, and *RNL* fraction were evaluated using the 'phylolm' package in R v4.2.0.

Results

Phylogenetic distribution and genome-wide variation in gene repertoires

The selected flora spanned a vast phylogenetic and life-history landscape, encompassing algae, bryophytes, ferns, gymnosperms, basal angiosperms, monocots, and eudicots (Fig. 1a). Rather than being confined to a single clade, life-history strategies—ranging from annuals and perennials to shrubs and woody trees—were dispersed across numerous independent evolutionary lineages. This phylogenetic architecture established a robust comparative framework for interrogating whether variations in the genome-wide gene repertoire correlate with organismal longevity.

The total gene count fluctuated markedly across species, spanning from under 7,500 to over 15,000 genes (Fig. 1b; Supplementary Table S3a). Although considerable variability existed among taxa, an overarching positive correlation emerged between total gene number and $\log_{10}$ -transformed lifespan, as delineated by the fitted regression line. While several ephemeral species harbored relatively modest gene repertoires, certain long-lived woody taxa displayed more expansive total gene complements. Nevertheless, the scattering of data points around the regression trajectory signaled that lifespan alone could not wholly account for the disparities in total gene abundance.

Further dissection of gene repertoire composition unveiled distinct contributions from core, dispensable, and unique gene categories (Fig. 1c). Core genes comprised the predominant fraction in

most species, aligning with their entrenched role in essential cellular and developmental machinery. Dispensable genes constituted a significant portion across the phylogeny, albeit with varying relative abundances among lineages. Unique genes typically represented only a minor slice of the total repertoire, though notable surges were evident in select taxa. Consequently, variation in total gene numbers was primarily associated with differences in dispensable genes and, secondarily, with unique gene fractions, rather than with substantial shifts in the core component.

The environmental trait heatmap illuminated pronounced inter-specific differences in lifespan, precipitation, elevation, temperature range, and latitude among the sampled species (Fig. 1d). Lifespan and temperature-range metrics exhibited the most striking variation, whereas precipitation and elevation patterns appeared comparatively subdued and heterogeneous. Collectively, these findings affirm that the dataset encapsulates robust diversity in phylogenetic placement, life-history strategies, genome size-associated gene content, and environmental attributes, thereby laying a solid foundation for subsequent investigations into the nexus between longevity and the stability of core gene repertoires.

Longevity is associated with greater conservation of core orthogroup repertoires

To ascertain whether longevity correlates with the preservation of widely shared genomic elements, we measured orthogroup occupancy across the 40 species encompassed by our comparative study. Orthogroups detected in no fewer than 95% of these taxa—specifically, in at least 38 of the 40—were designated as core orthogroups. This occupancy-driven categorization delineated a universally conserved core from both dispensable orthogroups and those unique to individual species (Fig. 2a).

At the 95% occupancy threshold, species with extended lifespans typically preserved a more substantial fraction of their core orthogroup repertoire (Fig. 2b; Supplementary Table S3b). This association was statistically significant in the original regression model, $p = 0.0193$, and remained significant under a 10,000-permutation tip-label test, $p = 0.0279$ (Supplementary Fig. S1). Conversely, the deficit in core orthogroups—defined as one minus retention—exhibited a tendency to diminish as lifespan increased (Fig. 2c; Supplementary Table S3c). Consequently, both metrics, acting as complementary lenses, revealed an identical pattern: longevity correlates with diminished erosion of the core orthogroup set and a more intact representation of the shared genomic foundation.

This relationship proved robust regardless of the specific threshold employed to delineate the core set. Positive correlations between lifespan and core orthogroup retention persisted across occupancy thresholds of 90%, 95%, and 99% (Fig. 2d; Supplementary Fig. S2). While stricter thresholds inevitably narrowed the core set and compressed retention values toward total conservation, the directional link with lifespan remained unwavering. Similarly, the species-level presence-absence matrix indicated that the majority of core orthogroups were maintained across taxa, with observed absences confined to a relatively narrow subset of orthogroups and species (Supplementary Fig. S3).

Acknowledging that species sharing evolutionary history cannot be treated as statistically independent units, we further assessed this association using phylogenetic generalized least-squares models. By accounting for phylogenetic relatedness and total orthogroup counts, these models offered a more conservative evaluation of the link between lifespan and core orthogroup retention or deficit (Supplementary Tables S4, S5). Collectively, these findings

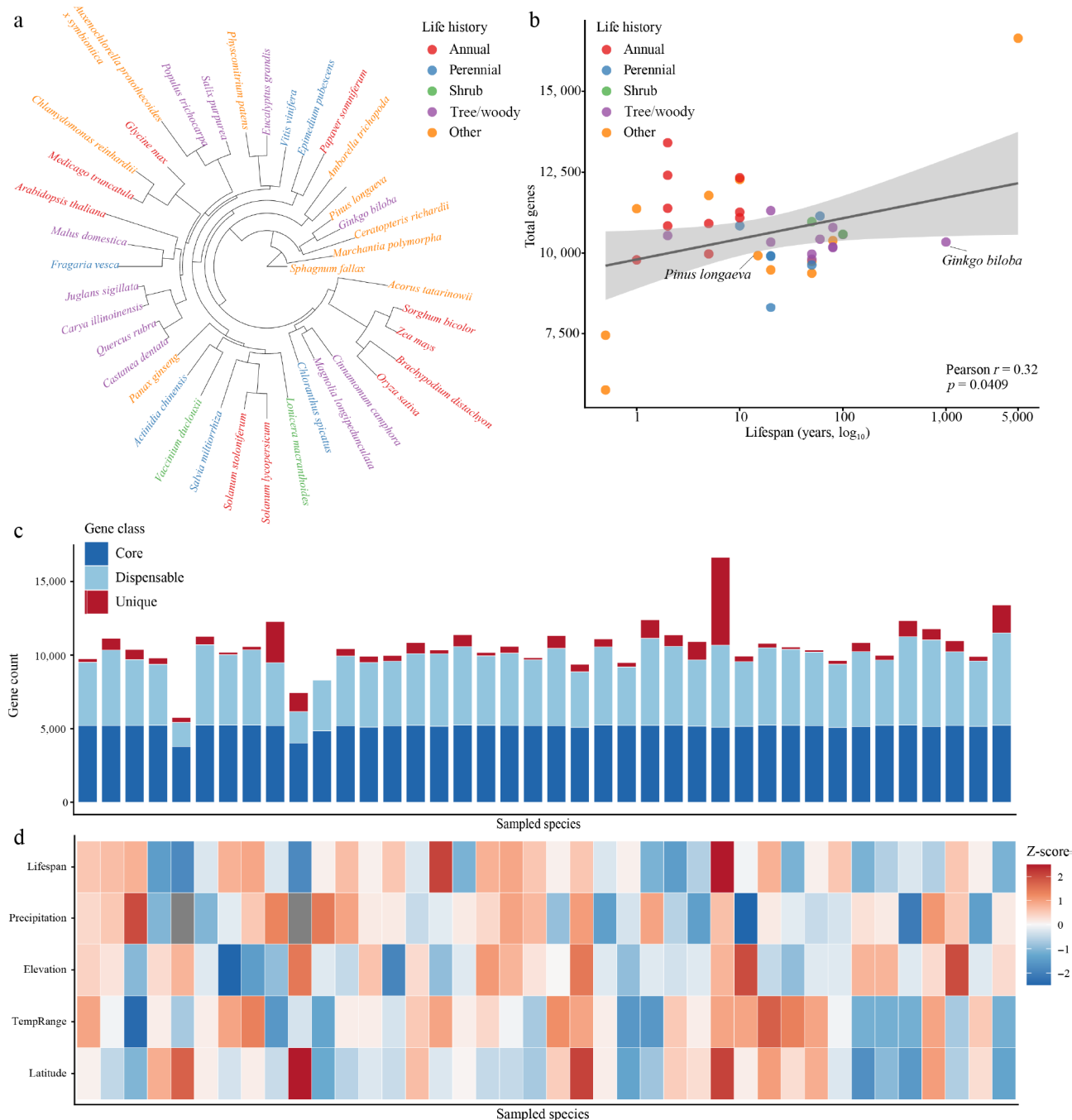


Fig. 1 Phylogenetic distribution, lifespan-associated variation, and gene repertoire composition across the sampled plant species. (a) Circular phylogenetic tree showing the evolutionary relationships among the sampled species. Tip-label colors represent major life-history categories: annual, perennial, shrub, tree/woody, and other. (b) Relationship between species lifespan and total life-history number. Lifespan is plotted on a log₁₀-transformed scale, and point colors indicate life-history categories. The fitted linear regression line is shown in dark gray, with the corresponding 95% confidence interval shown as a gray band. (c) Stacked bar plot showing the composition of gene repertoires in each sampled species. Each bar represents one sampled species. Gene families are classified as core, dispensable, or unique according to their distribution across the sampled species. (d) Heatmap showing standardized values (Z-scores) of lifespan and selected environmental variables for each sampled species. Each column corresponds to the same species order as in panel (c). Red indicates relatively high values, blue indicates relatively low values, and gray indicates missing data.

indicate an association between longevity and heightened conservation of core orthogroup repertoires.

Relationship between core orthogroup proportion and plant longevity

To explore how gene repertoires evolve alongside plant lifespan, we examined the link between the fraction of core orthogroups and

maximum recorded longevity. Phylogenetic generalized least squares (PGLS) regression uncovered a significant negative association between the proportion of core orthogroups and log-transformed lifespan ($\beta = -0.025$, $p = 0.01$; Fig. 3a). Conversely, non-parametric Spearman's rank correlation detected no significant monotonic trend ($\rho = 0.10$, $p = 0.53$), implying that the observed linear phylogenetic signal may stem from outliers or particular clades rather than representing a universal gradient across all taxa.

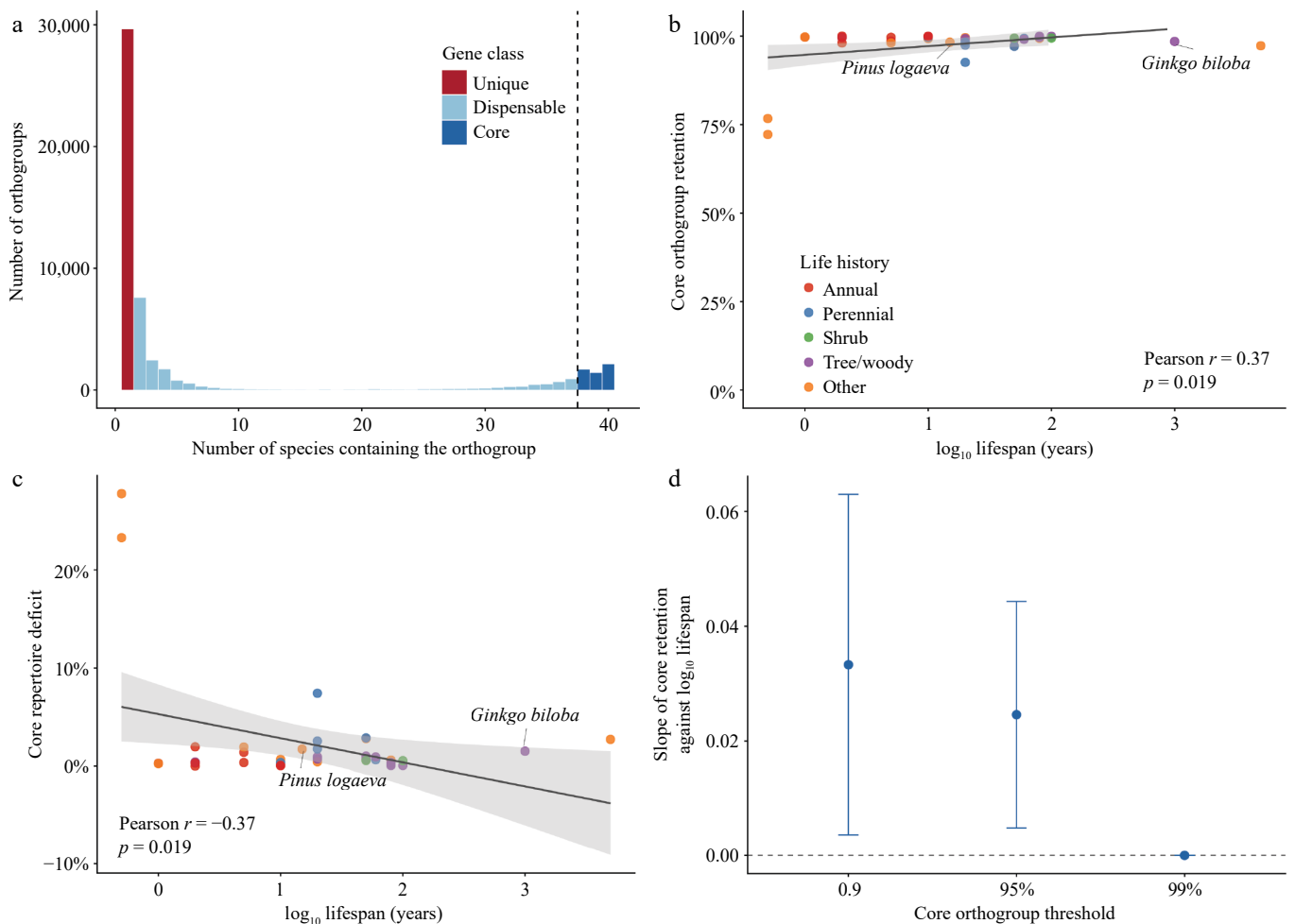


Fig. 2 Lifespan-associated stability of the core orthogroup repertoire across species. (a) Distribution of orthogroup occupancy across the 40 species included in the analysis. The dashed vertical line indicates the 95% core threshold. (b) Relationship between $\log_{10}$ -transformed lifespan and core orthogroup retention. (c) Relationship between $\log_{10}$ -transformed lifespan and core repertoire deficit. Core repertoire deficit was calculated as one minus core orthogroup retention. The solid line and shaded area represent the ordinary least-squares regression and its 95% confidence interval, respectively. (d) Sensitivity of the lifespan association to the definition of the core orthogroup set. Points show the regression slope relating core orthogroup retention to $\log_{10}$ -transformed lifespan under core thresholds of 90%, 95%, and 99%. Error bars indicate 95% confidence intervals. The dashed horizontal line indicates a slope of zero. All analyses were based on the 40 species shared among the lifespan, species information, and orthogroup datasets.

While the absolute number of retained core orthogroups showed a positive association with longevity before multiple-testing correction (slope = 129.08, $p = 0.019$; FDR-adjusted $p = 0.058$; Fig. 2b), their proportional contribution to the total repertoire decreased with longevity (Fig. 3a). This pattern indicates that the lower relative representation of core orthogroups does not reflect an absolute loss of the conserved core. Instead, it may partly reflect expansion of the overall orthogroup repertoire, although the positive associations observed for dispensable and lineage-specific orthogroups were not statistically significant (dispensable: $p = 0.197$; lineage-specific: $p = 0.140$; Table 1). Thus, the evidence for a specific dilution effect driven by non-core orthogroup expansion remains limited.

When stratified by life-history category, the proportions of core orthogroups showed distinct distributions among functional groups (Fig. 3b). Long-lived woody perennials, including trees and shrubs, generally exhibited lower proportions of core orthogroups than annual herbs, although substantial overlap was observed among categories. This pattern is consistent with the possibility that long-lived woody species possess more variable or expanded non-core repertoires, but it does not by itself demonstrate relaxed selection on conserved gene sets or enhanced genome plasticity.

Subsequent examination uncovered a pronounced inverse correlation between the fraction of core orthogroups and that of species-specific orthogroups ($r = -0.67$, $p < 0.001$; Fig. 3c). This reciprocal dynamic underscores a genomic compromise: lineages abundant in conserved, housekeeping gene suites typically harbor fewer lineage-restricted genes, whereas taxa boasting expanded unique repertoires display diminished core genome shares. In Fig. 3c, data point dimensions, scaled to reflect lifespan, further illuminate how the most long-lived species (reaching up to 5,000 years) congregate within the low-core, high-unique quadrant, thereby strengthening the link between longevity and genetic innovation.

When narrowing the lens to angiosperms, the negative association between core orthogroup proportion and lifespan persisted across varied life histories (Fig. 3d). Within this clade, annuals retained the highest core proportions, while perennial woody angiosperms exhibited significant declines. This within-angiosperm analysis indicates that the observed pattern is not explained solely by comparisons among distantly related major plant groups, although residual confounding cannot be excluded.

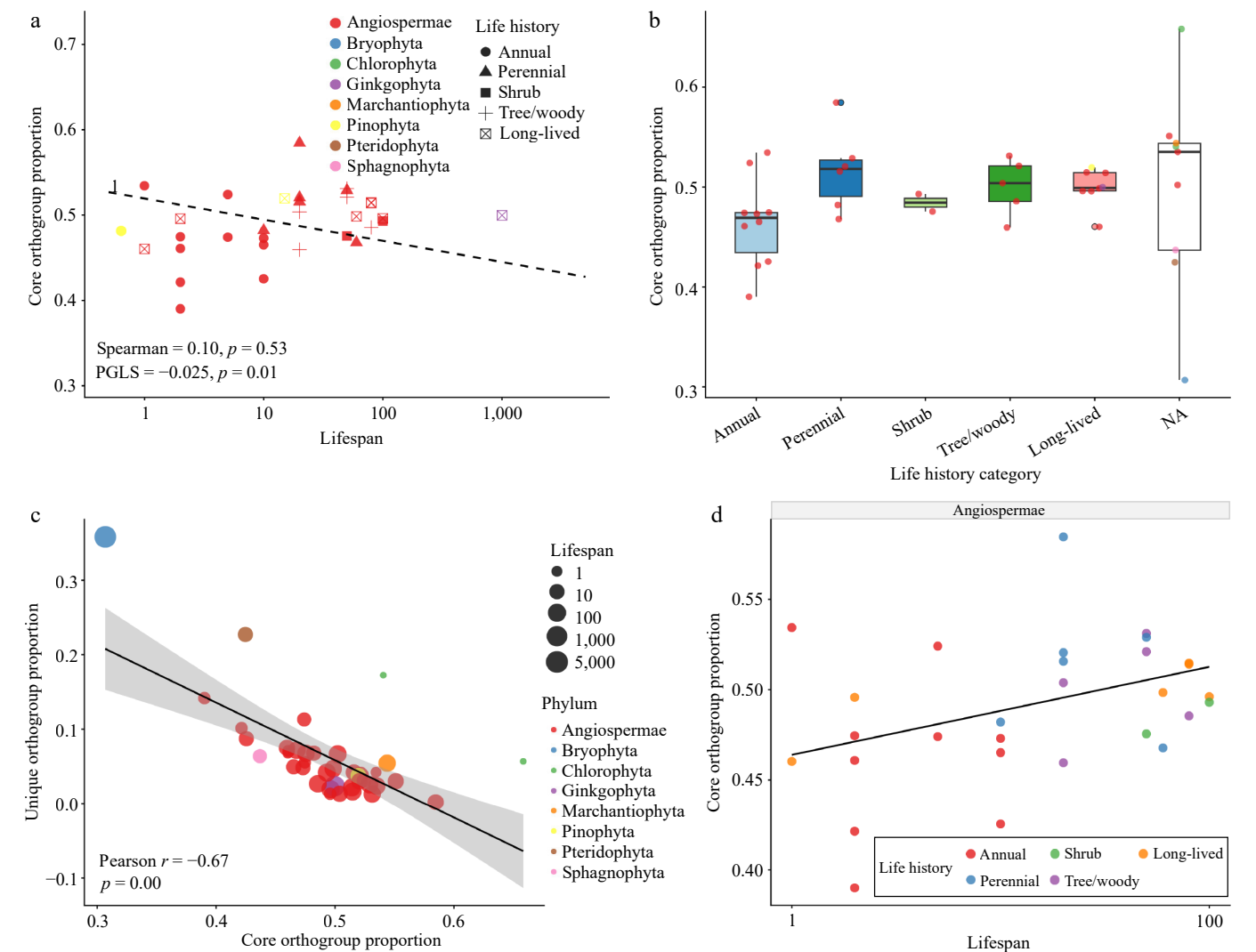


Fig. 3 Core orthogroup proportion in relation to plant lifespan and life history traits across taxonomic groups. (a) Scatter plot of core orthogroup proportion vs lifespan (log₁₀ years), colored by phylum and shaped by growth form (shrub, tree/woody, long-lived). (b) Boxplot of core orthogroup proportion grouped by life history category (annual, perennial, shrub, tree/woody, long-lived, NA). (c) Scatter plot of unique orthogroup proportion vs core orthogroup proportion, with point size indicating lifespan (1–5,000 years) and color indicating phylum. Regression line with 95% CI shaded. (d) Scatter plot of core orthogroup proportion vs lifespan (log₁₀ years) for Angiospermae only, colored by life history category (annual, perennial, shrub, tree/woody, long-lived).

Table 1. Regression analyses of absolute orthogroup counts in relation to plant lifespan.

Orthogroup category	Slope (β)	SE	t	p -value	95% CI	R^2
Core retained	129.079	52.832	2.443	0.019	22.127 to 236.032	0.136
Dispensable	224.448	170.741	1.315	0.197	-121.199 to 570.095	0.043
Unique	283.05	187.577	1.509	0.14	-96.680 to 662.780	0.057

Functional enrichment of core orthogroups

To unravel the biological functions embedded within the 2,139 core orthogroups (OGs) conserved across all 40 land plant species, we conducted a Gene Ontology (GO) enrichment analysis employing a per-species background model. This inquiry revealed a profound enrichment of terms governing structural integrity and cellular organization. Universally, these core genes emerged as the architectural bedrock of the cell, showing overwhelming association with intracellular anatomical structure (GO:0044444), organelle part (GO:0044422), and membrane-bounded organelle (GO:0043226) categories (Fig. 4a; $p_{\text{adj}} < 1 \times 10^{-20}$ in most species). Notably,

signatures linked to the ribonucleoprotein complex (GO:0031974), cytoplasm (GO:0005737), and nuclear part (GO:0044428) consistently dominated the statistical landscape, mirroring the indispensable role of these genes as the custodians of fundamental cellular machinery, from the protein-synthesizing ribosomes to the intricate choreography of chromatin organization (Supplementary Table S6). Notably, although core orthogroups exhibited profound enrichment in Gene Ontology (GO) terms, they displayed only modest significant enrichment within specific KEGG pathways. Rather than clustering into discrete routes, these core genes demonstrated broad representation across fundamental KEGG categories—

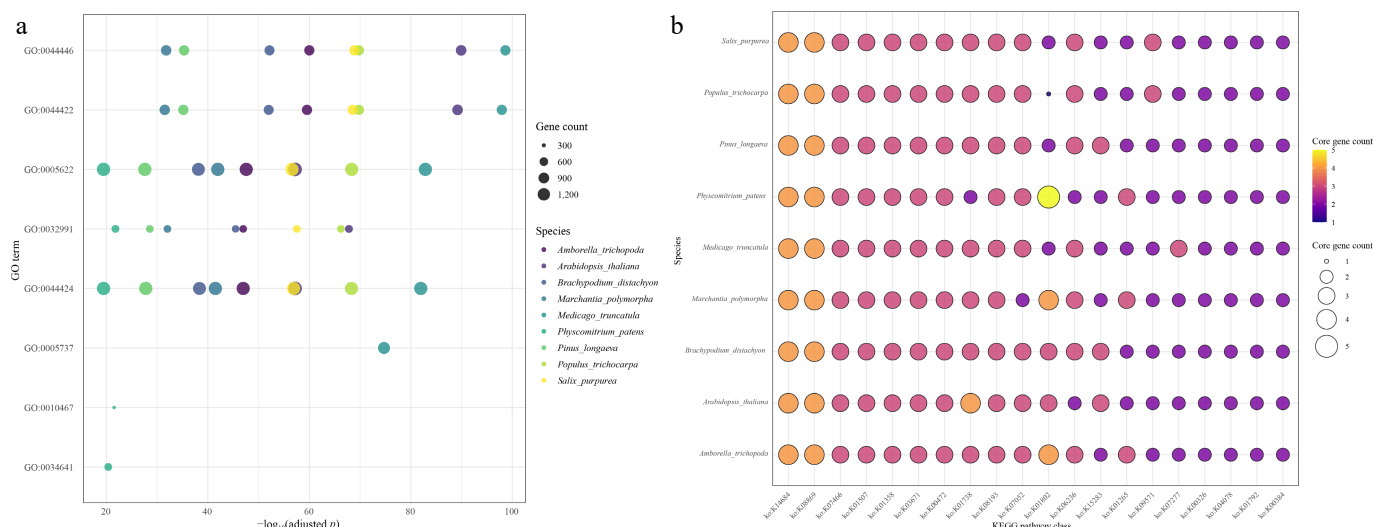


Fig. 4 Evolutionary dynamics of core genes across 40 land plants. (a) Enrichment analysis of core orthogroups (OGs) based on Gene Ontology (GO) terms. (b) Coverage of KEGG pathway classes by core orthogroups. Each bubble indicates the number of core genes associated with a specific KEGG pathway class (e.g., metabolism, genetic information processing) across species.

encompassing metabolism, genetic information processing, and cellular processes (Fig. 4b). This dispersed pattern underscores the 'housekeeping' character of deeply conserved genes, which operate ubiquitously across a web of interconnected networks instead of being confined to specialized biochemical corridors. Taken together, these findings reveal that core orthogroups form a resilient genetic scaffold that sustains essential cellular architecture while accommodating nuanced functional divergence that echoes the evolutionary arcs of major plant lineages.

Evolutionary dynamics of NLR immune receptor repertoires across plant lifespan

PGLS analyses identified a positive phylogenetically structured association between maximum recorded lifespan and NLR gene density, quantified as NLRs per 1,000 orthogroups ($\beta = 0.36 \pm 0.10$, $t = 3.37$, $p < 0.01$, $\lambda = 0.27$, $R^2 = 0.34$; Fig. 5a; Supplementary Tables S7, S8). Once phylogenetic relationships were accounted for, long-lived species retained significantly elevated copy numbers of NLR receptors, exhibiting an approximate threefold surge in NLR density for every tenfold extension of lifespan. This trend remained consistent across divergent lineages—including gymnosperms, basal angiosperms, and derived eudicots—and proved resilient to alternative core-genome definitions (Supplementary Fig. S4).

To ascertain whether this expansion represented merely a numerical accumulation or a qualitative transformation in immune architecture, we stratified NLR repertoires into major structural classes. The relative proportion of TNL (Toll/interleukin-1 receptor nucleotide-binding leucine-rich repeat) receptors rose with lifespan ($\beta = 1.42 \pm 0.74$, $p = 0.064$; Fig. 5b), while RNL (Resistance to Powdery Mildew 8-like) receptors diminished ($\beta = -1.31 \pm 0.77$, $p = 0.096$; Fig. 5c). These trajectories are consistent with a possible change in NLR class composition along the lifespan gradient, but the evidence for the TNL and RNL associations was statistically uncertain. This architectural trade-off was further corroborated by an inverse relationship between TNL and RNL fractions across species (Supplementary Figs. S5, S6).

Broader taxonomic comparisons reinforced these observations. Green algae, microalgae, and bryophytes consistently displayed markedly lower NLR densities compared to vascular plants, whereas fast-growing and long-lived trees harbored the highest NLR copy

numbers (Supplementary Fig. S4). Among angiosperms, perennial growth forms—especially long-lived trees—demonstrated heightened NLR investment relative to annual or short-lived herbaceous lineages. Notably, several early-diverging land plants (e.g., *Sphagnum fallax*, *Marchantia polymorpha*) retained minimal NLR complements, highlighting the innovation of intracellular immune recognition coinciding with terrestrialization.

Using the original phylolm Brownian-motion framework, lifespan was positively associated with NLR density (Fig. 5d; Supplementary Table S8). This association was attenuated by some geographic and climatic covariates but remained statistically supported in the full environmental model. Thus, measured environmental variation does not fully explain the lifespan–NLR-density association. Collectively, these findings support a model wherein extended vegetative longevity is associated with the expansion and qualitative refinement of NLR repertoires, establishing a mechanistic bridge between life history evolution and intracellular immunity in land plants.

Focused annotation of longevity-relevant functions within core orthogroups

All five modules were significantly enriched among core orthogroups relative to annotated non-core orthogroups (Fig. 6a). The strongest enrichment was observed for oxidative-stress response and redox homeostasis (325 core orthogroups; FDR = 2.6×10^{-18}), followed by chromatin and epigenetic regulation (282 orthogroups; FDR = 9.0×10^{-7}), autophagy and protein homeostasis (275 orthogroups; FDR = 3.3×10^{-12}), DNA repair and genome stability (142 orthogroups; FDR = 3.6×10^{-12}), and telomere-associated processes (15 orthogroups; FDR = 4.4×10^{-3}). These results indicate that the conserved genomic component is particularly enriched for functions involved in redox balance, proteome maintenance, genome integrity, and epigenetic regulation.

Functional annotation coverage was generally high across the plant species, although the proportion of core genes assigned GO terms, KEGG orthologs, KEGG pathways, or eggNOG descriptions varied among species (Fig. 6b). The high coverage of at least one annotation type supports the functional interpretation of the conserved orthogroup set, while the differences among annotation resources highlight variation in annotation depth and database

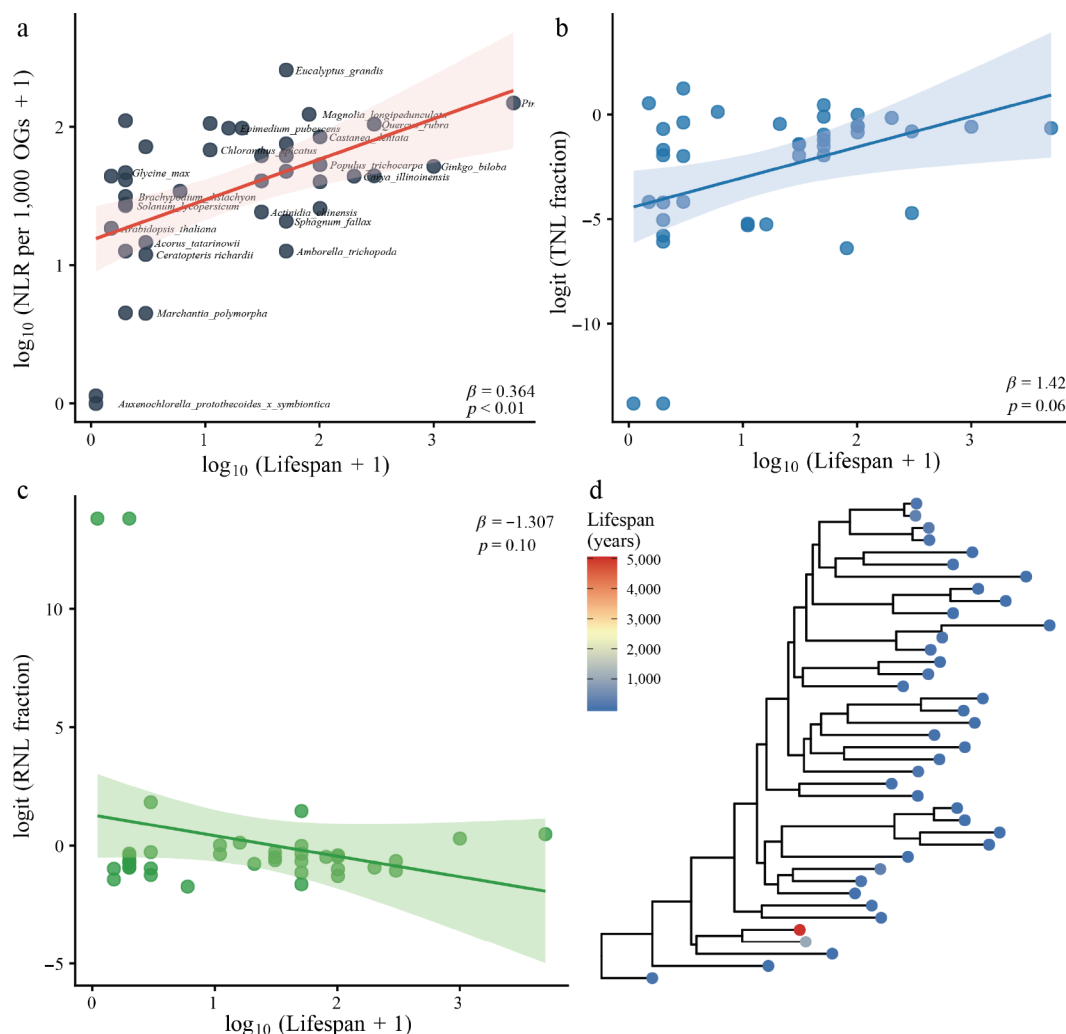


Fig. 5 Evolutionary dynamics of NLR immune receptor repertoires across plant lifespan. (a) Phylogenetic generalized least squares (PGLS) regression of $\log_{10}(\text{NLR gene density; NLRs per 1,000 orthogroups} + 1)$ against $\log_{10}(\text{maximum recorded lifespan} + 1)$ across 40 species spanning algae, bryophytes, ferns, and angiosperms. Species names are labelled adjacent to data points; shading denotes 95% confidence intervals. (b) PGLS regression of TNL (Toll/interleukin-1 receptor nucleotide-binding leucine-rich repeat) receptor proportion against lifespan. (c) PGLS regression of RNL (Resistance to Powdery Mildew 8-like) receptor proportion against lifespan. (d) Left-to-right maximum-likelihood species tree inferred using OrthoFinder (STAG algorithm), with tip colours scaled to species lifespan.

coverage. We further examined representative core orthogroups from these five modules using cross-species copy-number profiles (Supplementary Fig. S7). The selected orthogroups were broadly retained across the sampled species, but their copy numbers varied among lineages. The conserved presence of these orthogroups, together with their enrichment in longevity-relevant functions, suggests that genome maintenance, stress protection, protein homeostasis, chromatin regulation, and telomere-associated processes constitute recurrent components of the conserved plant gene repertoire.

Discussion

Our discovery that long-lived plants preserve a more comprehensive suite of core orthogroups both echoes and expands upon classical life-history theory. The Williams pleiotropy hypothesis, alongside subsequent models of somatic mutation accumulation, posits that organisms with extended lifespans must mitigate germline and somatic genomic instability by bolstering the conservation of essential genes^[7,22]. While earlier comparative genomics in mammals and

birds largely substantiate this perspective—noting reduced gene turnover rates in long-lived lineages^[23]—plant genomes diverge fundamentally from their animal counterparts due to their tolerance for whole-genome duplication, pervasive transposable element activity, and post-zygotic somatic mutations^[2,24]. Our finding that the absolute count of core orthogroups remains constant even as their proportional contribution diminishes with lifespan refines these existing frameworks: longer maximum recorded lifespan was not associated with genome-wide uniformity; instead, it covaried with relatively greater retention of core orthogroups alongside variation in non-core repertoires. This interpretation reconciles an apparent conflict with recent pan-genome studies in *Arabidopsis* and *Glycine*, which revealed extensive structural variation even among closely related accessions^[25,26]. Those investigations, centered on intraspecific diversity, likely captured precisely the expansion of the 'dispensable genome' that we demonstrate is not only compatible with, but perhaps facilitated by, the conservation of core elements in long-lived perennials.

The qualitative recalibration of NLR immune receptor architectures along the longevity continuum presents a striking analogy to,

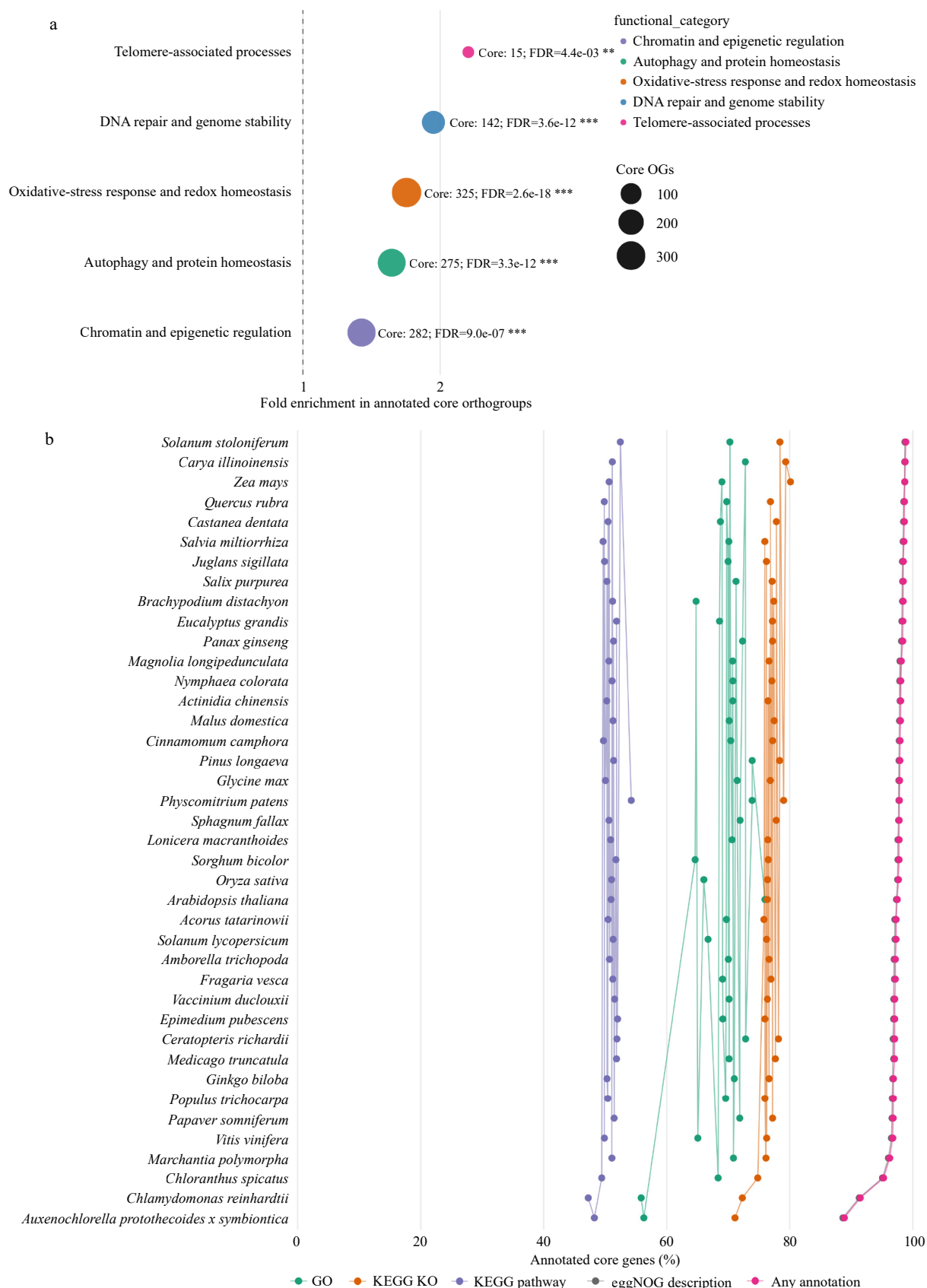


Fig. 6 Enrichment of longevity-relevant functional modules and annotation completeness of core genes. (a) Enrichment of five functional modules in core orthogroups relative to non-core orthogroups. Bubble size indicates the number of core orthogroups, and significance was assessed using one-sided Fisher's exact tests followed by Benjamini–Hochberg correction. (b) Functional annotation completeness of core genes across 40 plant species. Percentages indicate the proportions of core genes with GO, KEGG KO, KEGG pathway, eggNOG description, or any annotation.

yet a distinct departure from, canonical paradigms in plant immunity. While prior scholarship has underscored *NLR* expansions as reactive maneuvers against localized pathogen pressures and coevolutionary duels—typically viewed through the prism of the gene-for-gene hypothesis^[27,28]—and comparative audits across Brassicaceae and Poaceae have illuminated lineage-specific *NLR* surges tied to domestication or host transitions^[29–31], our findings point to a broader, life-history-propelled macroevolutionary current. Across the 40 species, maximum recorded lifespan was associated with variation in the relative representation of *TNL* and *RNL* receptors, although this pattern may also reflect correlated ecological or life-history traits. This trend echoes functional inquiries suggesting that *TNL*s orchestrate more refined effector-triggered immunity, frequently mobilized in long-lived woody perennials besieged by chronic, low-virulence pathogens^[32,33]. Nevertheless, our dependence on maximum recorded lifespans—aggregated from botanical archives rather than longitudinal field expeditions—precludes determining whether this immune topology is causally engendered by longevity itself or merely mirrors correlated attributes such as woodiness, shade tolerance, or sluggish growth rates^[8]. Furthermore, although our orthology framework adeptly seizes canonical *NLR* structures, it may undercount highly divergent or atypical receptors, potentially veiling lineage-specific innovations that transcend the *TNL/RNL* dichotomy.

Integrating these patterns into the wider story of plants colonizing land sheds light on the constant push-and-pull between evolutionary breakthroughs and biological limits. Although early land pioneers such as bryophytes and hornworts display meager *NLR* collections—a trend traditionally blamed on the rudimentary nature of pre-vascular immunity^[34–36]—our results realign this viewpoint. The paucity of *NLR*s in these groups aligns with a striking preservation of core orthogroups, suggesting that early-diverging land plants showed relatively high core orthogroup conservation and low *NLR* abundance in our dataset. This pattern may reflect differences in ecology, life history, genome composition, or *NLR* annotation across these lineages. Conversely, higher *NLR* abundance in several vascular plant lineages coincided with the occurrence of secondary growth and longer vegetative lifespans, but these factors could not be separated in the present dataset. Yet, our species tree, rebuilt using STAG within a multispecies coalescent framework, cannot fully unravel ancient web-like events, such as suspected hybridizations among early seed plant lineages that might have spurred *NLR* expansions regardless of lifespan^[37]. Moreover, while scaling *NLR* counts against orthogroup numbers reduces biases arising from annotation intensity, this method does not account for fluctuations in genome size or repetitive content, factors known to sway *NLR* gene density across particular clades^[38].

Our collective findings recalibrate the prevailing paradigm of plant genome evolution, highlighting lifespan as a potentially important correlate of genome-wide repertoire composition. Within this framework, genomes preserve immutable sequences in core cellular machinery where replication fidelity is paramount, while simultaneously permitting sequence innovation in immune receptors and lineage-specific loci that necessitate adaptive plasticity. This strategic equilibrium resolves the enduring enigma regarding the coexistence of genomic conservation and malleability observed in prior evolutionary scholarship^[39,40]. Future inquiries ought to integrate somatic mutation metrics, population-scale pangenome datasets, and pathogen challenge assays across species with divergent lifespans to disentangle causal mechanisms from spurious correlations. Nevertheless, macroevolutionary analysis identifies a phylogenetically structured association between plant life-history

variation and genome-repertoire composition, offering a robust model to elucidate how plant genomes negotiate the tension between long-term stability and environmental responsiveness over deep time.

Conclusions

This research constructs a macroevolutionary paradigm connecting plant lifespan to the dynamic flux of genome-wide gene repertoires. We reveal that prolonged vegetative existence aligns with a bifurcated genomic approach: the preservation of a steadfast core genetic toolkit coupled with the qualitative honing of intracellular immune receptors. Although species with extended lifespans retain nearly intact sets of ancestral orthogroups, the relative weight of this core is diluted by the proliferation of dispensable and lineage-specific genes, exposing a delicate equilibrium between genomic inertia and adaptability. Critically, this life-history continuum is associated with a transformation in immune architecture—shifting from *RNL*-dominated basal defenses in ephemeral taxa to *TNL*-enriched arsenals capable of sophisticated effector discrimination in long-lived perennials. Deeply embedded within the plant phylogeny, these patterns imply that gene-repertoire composition varies with longevity, although the extent to which this pattern reflects direct selection on longevity-related traits remains unresolved, and that a distinct genomic constitution is fine-tuned for somatic resilience and pathogen pressure across centuries to millennia. Our results offer a mechanistic basis for comprehending genome stability in long-lived entities and pioneer fresh pathways for correlating life history evolution with immune system design in terrestrial plants.

Ethical statements

During the preparation of this work, the author(s) used ChatGPT (Version: 5.6 Terra) for language refinement. The author(s) reviewed and edited all content produced with the assistance of this tool, verified its accuracy, and take full responsibility for the integrity and originality of the final manuscript. This work represents the author's own intellectual contribution, and no AI tool is credited as an author.

Author contributions

The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary information accompanies this paper online at: <https://doi.org/10.48130/fra-0026-0017>.

Dates

Received 2 August 2026; Revised 15 September 2026; Accepted 21 September 2026; Published online 23 October 2026

References

- [1] Li W, Liu J, Zhang H, Liu Z, Wang Y, et al. 2022. Plant pan-genomics: recent advances, new challenges, and roads ahead. *Journal of Genetics and Genomics* 49:833–846
- [2] Neale DB, Kremer A. 2011. Forest tree genomics: growing resources and applications. *Nature Reviews Genetics* 12:111–122
- [3] Fang Y, Wei J, Huang X, Li Y, Pan X. 2023. Identification and analysis of protein family associated with RNA interference pathway in Juglandaceae. *Frontiers in Bioscience (Landmark Edition)* 28:218
- [4] Fang Y, Wei J, Huang X, Qin L, Li Y. 2023. RNA interference-core proteins from the *Actinidiaceae*: evolution, structure, and functional differentiation. *Plant Gene* 34:100419
- [5] Tanveer A, Atif RM, Khan SH, Ahmad Rana I. 2025. Comparative genomics and evolutionary insights in divergence and expression of CYP450 super gene family in citrus (*C. sinensis*, *C. grandis* & *P. trifoliata*). *BMC Plant Biology* 25:1614
- [6] Peng J, Xia L, Wang J, Guo C. 2026. Evolution of plant AIG1-like proteins: different modes of sequence divergence and their contributions to functional diversification. *Plants* 15:301
- [7] Gao J, Fang K, Chen JM, Li J, Rossi S, et al. 2025. Climate-driven patterns of global tree longevity. *Communications Earth & Environment* 6:610
- [8] Piovesan G, Biondi F. 2021. On tree longevity. *New Phytologist* 231:1318–1337
- [9] Foley J, McPherson J, Roger M, Batista C, Mauxion R, et al. 2026. Evolution of increased longevity and slowed ageing in a genus of tropical butterfly. *Nature Communications* 17:5077
- [10] Maklakov AA, Chapman T. 2019. Evolution of ageing as a tangle of trade-offs: energy versus function. *Proceedings of the Royal Society B: Biological Sciences* 286:20191604
- [11] Nettle D, Frankenhuis WE. 2019. The evolution of life-history theory: a bibliometric analysis of an interdisciplinary research area. *Proceedings Biological Sciences* 286:20190040
- [12] Bélanger S, Zhan J, Meyers BC. 2023. Phylogenetic analyses of seven protein families refine the evolution of small RNA pathways in green plants. *Plant Physiology* 192:1183–1203
- [13] Su LY, Liu ZT, Chen PY, Wang XL, Liu H, et al. 2025. Origin and evolution of signaling pathways responsible for ascorbic acid synthesis and catabolism during plant terrestrialization. *Horticulture Research* 12:uhaf184
- [14] Emms DM, Liu Y, Belcher L, Holmes J, Kelly S. 2026. OrthoFinder: improved phylogenetic orthology inference with enhanced accuracy and scalability. *Nature Methods* 23:1327–1333
- [15] Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37:1530–1534
- [16] Hernández-Plaza A, Szklarczyk D, Botas J, Cantalapiedra CP, Giner-Lamia J, et al. 2023. eggNOG 6.0: enabling comparative genomics across 12 535 organisms. *Nucleic Acids Research* 51:D389–D394
- [17] Consortium GO. 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research* 32:D258–D261
- [18] Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. 2023. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research* 51:D587–D592
- [19] Smith M, Jones JT, Hein I. 2025. Resistify: a novel NLR classifier that reveals helitron-associated NLR expansion in Solanaceae. *Bioinformatics and Biology Insights* 19:11779322241308944
- [20] Zhang W. 2020. NLR-annotator: a tool for de novo annotation of intracellular immune receptor repertoire. *Plant Physiology* 183:418–420
- [21] Fang Y, Wang Z, Uddin N, Srivastava AK, Yang S, et al. 2025. Tissue-specific metabolic reprogramming, not anthocyanins, drives chromium tolerance in *Brassica rapa*. *Environmental Science & Technology* 59:20484–20498
- [22] Williams GC. 1957. Pleiotropy, natural selection, and the evolution of senescence. *Evolution* 11:398–411
- [23] Ji Y, Feng S, Wu L, Fang Q, Brüniche-Olsen A, et al. 2022. Orthologous microsatellites, transposable elements, and DNA deletions correlate with generation time and body mass in neoavian birds. *Science Advances* 8:eabo0099
- [24] Schoen DJ, Schultz ST. 2019. Somatic mutation and evolution in plants. *Annual Review of Ecology, Evolution, and Systematics* 50:49–73
- [25] Lian Q, Huettel B, Walkemeier B, Mayjonade B, Lopez-Roques C, et al. 2024. A pan-genome of 69 *Arabidopsis thaliana* accessions reveals a conserved genome structure throughout the global species range. *Nature Genetics* 56:982–991
- [26] Sultan AB, Nawaz H, Saleem F, Nawaz S, Danial M, et al. 2024. Divergent evolution of NLR genes in the genus *Glycine*: impacts of annuals and perennials' life history strategies. *Frontiers in Plant Science* 15:1383135
- [27] Jones JDG, Staskawicz BJ, Dangl JL. 2024. The plant immune system: from discovery to deployment. *Cell* 187:2095–2116
- [28] van der Hoorn RAL, Kamoun S. 2008. From Guard to Decoy: a new model for perception of plant pathogen effectors. *The Plant Cell* 20:2009–2017
- [29] Goritschnig S, Steinbrenner AD, Grunwald DJ, Staskawicz BJ. 2016. Structurally distinct *Arabidopsis thaliana* NLR immune receptors recognize tandem WY domains of an oomycete effector. *New Phytologist* 210:984–996
- [30] Van de Weyer AL, Monteiro F, Furzer OJ, Nishimura MT, Cevik V, et al. 2019. A species-wide inventory of NLR genes and alleles in *Arabidopsis thaliana*. *Cell* 178:1260–1272.e14
- [31] Wang Z, Fan Y, Sun J, Ma S, Wang Z, et al. 2025. Pan-analysis of intra- and inter-species diversity reveals a group of highly variable immune receptor genes in rice. *The Plant Journal* 122:e70163
- [32] Jacob F, Vernaldi S, Maekawa T. 2013. Evolution and conservation of plant NLR functions. *Frontiers in Immunology* 4:297
- [33] Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. 2003. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *The Plant Cell* 15:809–834
- [34] Shi C, Chen S, Wang J, Chen W, Sun C, et al. 2026. A tomato telomere-to-telomere super-pangenome empowers stress resilience breeding. *Nature Genetics* 58:630–642
- [35] Beaulieu C, Libourel C, Mbadinga Zamar DL, El Mahboubi K, Hoey DJ, et al. 2025. The *Marchantia polymorpha* pangenome reveals ancient mechanisms of plant adaptation to the environment. *Nature Genetics* 57:729–740
- [36] Castel B, Baker M, Keller J, Martinez Y, Bonhomme M, et al. 2025. Exploring fern pathosystems and immune receptors to bridge gaps in plant immunity. *BMC Biology* 23:301
- [37] Stull GW, Pham KK, Soltis PS, Soltis DE. 2023. Deep reticulation: the long legacy of hybridization in vascular plant evolution. *The Plant Journal* 114:743–766
- [38] Plomion C, Aury JM, Amselem J, Leroy T, Murat F, et al. 2018. Oak genome reveals facets of long lifespan. *Nature Plants* 4:440–452
- [39] Adams KL, Wendel JF. 2005. Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology* 8:135–141
- [40] Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, et al. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96:336–348



Copyright: © 2026 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.