BMC Genomics

RESEARCH

Open Access

Genomic structure of yellow lupin (*Lupinus luteus*): genome organization, evolution, gene family expansion, metabolites and protein synthesis

J. Eduardo Martinez-Hernandez[1,2†], Haroldo Salvo-Garrido[1*†], Daniela Levicoy[1†], Peter D. S. Caligari[1], Annally Rupayán[1], Tomas Moyano[3], Makarena Carrasco[1], Sebastián Hernandez[1], Grace Armijo-Godoy[1], Fernando Westermeyer[1] and Giovanni Larama[4]

Abstract

Yellow lupin (*Lupinus luteus*) gives valuable high-quality protein and has good sustainability due to its ability in nitrogen fixation and exudation of organic acids, which reduces the need for chemical-based phosphate fertilization in acid soils. However, the crop needs further improvements to contribute in a major way to sustainable agriculture and food security.

In this study, we present the first chromosome-level genome assembly of *L. luteus*. The results provide insights into its genomic organization, evolution, and functional attributes. Using integrated genomic approaches, we unveil the genetic bases governing its adaptive responses to environmental stress, delineating the intricate interplay among alkaloid biosynthesis, mechanisms of pathogen resistance, and secondary metabolite transporters. Our comparative genomic analysis of closely related species highlights recent speciation events within the *Lupinus* genus, exposing extensive synteny preservation alongside notable structural alterations, particularly chromosome translocations. Remarkable expansions of gene families implicated in terpene metabolism, stress responses, and conglutin proteins were identified, elucidating the genetic basis of *L. luteus'* superior nutritional profile and defensive capabilities. Additionally, a diverse array of disease resistance-related (R) genes was uncovered, alongside the characterization of pivotal enzymes governing quinolizidine alkaloid biosynthesis, thus shedding light on the molecular mechanisms underlying "bitterness" in lupin seeds.

This comprehensive genomic analysis serves as a valuable resource to improve this species in terms of resilience, yield, and seed protein levels to contribute to food and feed to face the worldwide challenge of sustainable agriculture and food security.

Keywords *Lupinus luteus*, Genome assembly, Chromosomal rearrangements, Protein and metabolites biosynthesis

†J. Eduardo Martinez-Hernandez, Haroldo Salvo-Garrido and Daniela Levicoy contributed equally to this work.

*Correspondence:
Haroldo Salvo-Garrido
Haroldo.salvo@cgna.cl

Full list of author information is available at the end of the article

Introduction

Lupins are valuable in agriculture because of their substantial seed protein content [1] but generally underexploited. They are found in various regions of the world and are grouped into Old World lupins (Mediterranean) and New World lupins (American) [2] species. The four principal species grown, with the highest protein content, are *Lupinus albus*, *Lupinus luteus*, *Lupinus angustifolius* and *Lupinus mutabilis* [3]. They have different nutritional profiles but high average protein, making them suitable for producing high-quality food and feed in both human and animal diets [4, 5]. In addition, they contribute to the sustainability of cropping systems owing to their low fertiliser requirements and positive impact on soil fertility, including their ability to exude organic acids, fix atmospheric nitrogen in symbiosis with beneficial bacteria and efficiently take up phosphorus from soils [6, 7].

L. luteus has high dehulled seed protein content, 60% dry matter (DM) [8] and twice the cysteine and methionine content of most other lupins [9, 10]. In contrast to soybean, it can be cultivated in regions with mild climates, being a valuable food source in climatic conditions unfavourable for soybean cultivation [11]. *L. luteus* has been studied as an alternative protein source for countries that import soybeans [12, 13, 14]. However, despite its agronomic importance, genomic and molecular studies in this species remain limited [9, 15, 16]. *L. luteus* has emerged as a promising candidate in the dynamic interplay between agricultural productivity and environmental challenges, offering potential solutions for resilient and sustainable crop production [17].

The intricate balance between biotic and abiotic stress factors significantly shapes the adaptive responses of plants, influencing their genetic landscape and biochemical composition [3, 18, 19]. The synthesis of alkaloids, particularly quinolizidine alkaloids (QAs), represents a complex response to these stress stimuli, as well as contributing to the plant's defence mechanisms against pests and pathogens [20]. However, the regulation of alkaloids becomes important due to challenges associated with bitterness and potential toxicity [20, 21], highlighting a nuanced aspect of lupin biology [20]. Recent studies have provided insights into the complex biosynthesis of QAs, revealing that their production is tightly regulated by both environmental stresses and genetic factors [22, 23, 24]. For instance, research in *L. albus* has identified key genes involved in QA synthesis, enhancing our understanding of the genetic control over alkaloid production [21]. Additionally, expression profiling in *L. angustifolius* has shown that alkaloid-related genes are differentially expressed across various plant organs and in response to anthracnose infection, indicating a dynamic regulation of QA biosynthesis in response to biotic factors [22]. Furthermore, mechanistic studies have advanced our understanding of the enzymatic pathways involved in QA formation, offering prospects for pathway elucidation and potential manipulation [25]. Understanding the regulation of alkaloid biosynthesis is essential for developing lupin varieties with optimized alkaloid content, balancing plant defence and nutritional quality.

Regarding biotic stress, candidate genes associated with disease resistance in *Lupinus* species have been identified [26, 27, 28]. For instance, in *L. angustifolius*, a key resistance gene linked to anthracnose and grey leaf spot infection has been mapped, providing valuable insights for breeding programs aimed at enhancing disease resistance in this species [28, 29]. Similarly, genetic mapping and comparative analyses in *L. luteus* have highlighted syntenic regions containing major orthologous genes controlling anthracnose resistance, offering a solid foundation for future improvement strategies [15, 27]. These genetic insights align with the crucial role played by resistance (R) proteins in plant defence [30]. R proteins serve as sentinels of the cellular defence system, recognizing pathogenic invaders and triggering molecular response cascades that enhance disease resistance [31, 32]. The identification of resistance-related genes in *Lupinus* suggests that R proteins may be key components of these defence mechanisms, mediating plant responses to biotic stresses through intricate regulatory networks.

These advances in understanding disease resistance and alkaloid biosynthesis in *Lupinus* provide a strong foundation for future research and breeding strategies aimed at enhancing resilience and nutritional value in sustainable agricultural systems. In the current work, a chromosome-level genome assembly for yellow lupin is presented, offering insights into its structural organisation, synteny patterns with others in the Order Fabales, and the abundance and diversity of repetitive elements. Furthermore, our high-quality genome assembly provides novel insights into the recent divergence between the lupin species, shedding light on the evolutionary dynamics of this agriculturally important genus. Additionally, a diverse array of resistance genes has been identified, indicating the plant's sophisticated defence mechanisms. This comprehensive genome analysis of *L. luteus* serves as a valuable resource for guiding efforts to enhance the resilience and productivity of this crop, which is poised to play a crucial role in ensuring global food security in the future.

Results

Genome assembly and validation

L. luteus C195 (commercial cultivar Alu*Prot*-CGNA ®; 2n = 2x = 52) was chosen for genome sequencing and assembly (Fig. 1A). The chromosome number of *L. luteus* (2n = 52) has been previously reported in cytogenetic studies of Old World lupins [33, 34, 35]. To

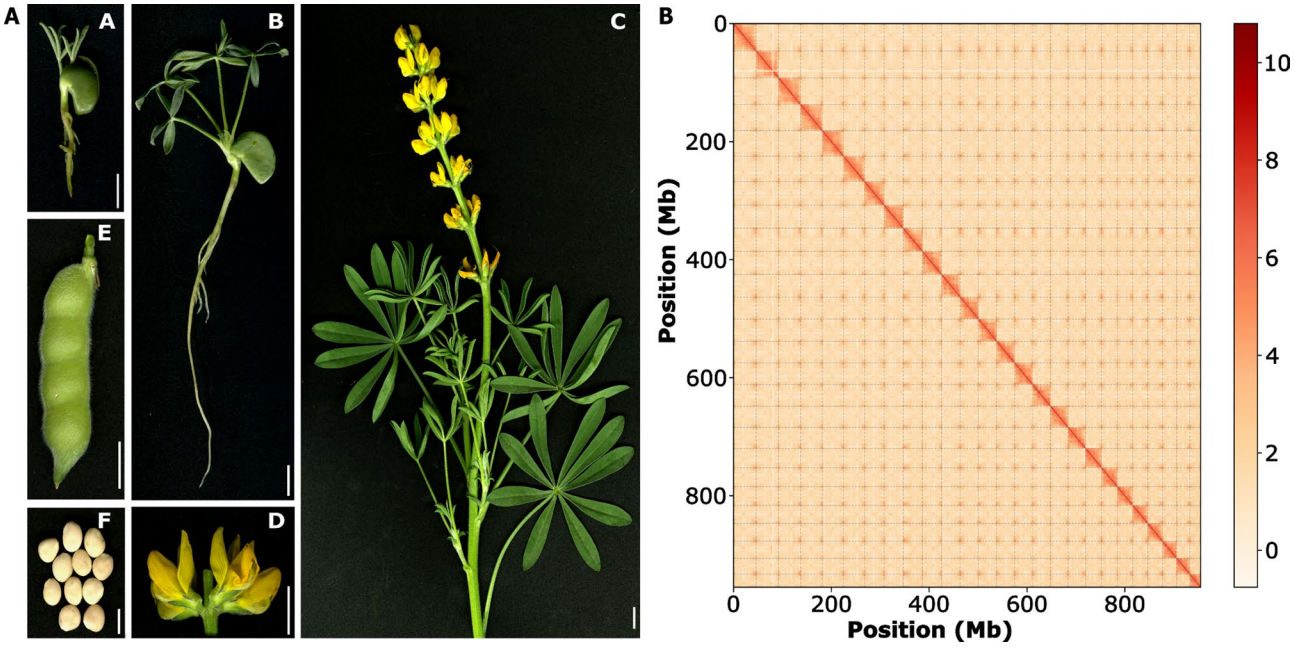

Fig. 1 Genome assembly of yellow lupin. (**A**) Morphological characteristics of a *L. luteus* plant, (**A**) Emergence growth stage; first pair of leaves protruding beyond upright cotyledons; (**B**) Growth stage, bases of several leaves separated from each other; (**C**) Reproductive development growth stages; showing the different stages of flower development; (**D**) Flowering; diverging standard petal stage (anthesis); (**E**) Green pod, septa between seeds, slight bulging of walls, seeds filling 50% of the space between the septa, and (**F**) Seed hard but dentable, mottling of pale fawn coat. Scale bar = 1 cm. (**B**) Contact map of scaffolding of 26 pseudomolecules for *L. luteus* genome assembly based on Hi-C sequencing. Resolution 250 Kb

Table 1 Statistics of L. luteus C195 genome assembly and annotation

Assembly features	
Estimated genome size (Mb)	1,024.49
Total assembly size (Mb)	962.97
GC content (%)	35.40
Repeat content (%)	76.20
Contig number	1,053
Contig N50 (Mb)	15.38
Longest contig (Mb)	26.27
Scaffold number	111
Scaffold N50 (Mb)	37.36
Longest scaffold (Mb)	46.32
Complete BUSCO (%):	
Embryophyta	98.90
Fabales	96.20
Annotation features	
No. of protein-coding genes	36,895
Average gene length (bp)	5,370
Average CDS length (bp)	1,191
Complete BUSCO (%):	
Embryophyta	93.90
Fabales	91.10
Protein-coding genes with annotation	35,940

obtain a high-quality assembly, genome sequencing was performed by integrating HiFi, NGS, and Hi-C data. From this sequencing protocol, were obtained 87.34 Gb (~130x) of HiFi clean data, 78.66 Gb (~119x) of Illumina short reads clean data, and 115.36 Gb (~175x) Hi-C clean data (Supplementary Table S1). A genome survey was performed to assess the genome size and heterozygosity of *L. luteus* using Illumina short-read data. *K*-mer analysis indicated that the estimated size of the genome was 1024.44 Mb (Table 1) with 0.076% heterozygosity (Supplementary Figure S1 and Supplementary Table S2). The primary assembly obtained with the HifiASM assembler [36] resulted in a genome assembly size of 1.01 Gb, consisting of 1,053 contigs with a contig N50 value of 15.38 Mb (Table 1, Supplementary Table S3). The assembly was polished for error correction using Illumina short reads from the same cultivar with Nextpolish [37] and the resulting polished contigs were used for the scaffolding step.

The contigs were scaffolded using Hi-C-assisted assembly, based on High-throughput Chromosome Conformation Capture (Hi-C) data. A final assembly of 111 scaffolds was obtained composed of 245 contigs with a total length of 962.91 Mb. 160 contigs containing approximately 955.60 Mb were arranged into 26 pseudo chromosomes, covering 99.24% of the assembly, with an N50 of 37.36 Mb (Table 1; Supplementary Table S4) with sizes ranging from 21.91 to 46.32 Mb (Supplementary Table S4). The Hi-C interaction map showed a matrix with non-obvious assembly errors comprising 26 clusters, indicating that the *L. luteus* genome was nearly complete (Fig. 1B). BUSCO assessment of the final assembly

revealed that the completeness of the assembly was approximately 98.9 and 96.2% against Embryophyta and Fabales lineages, respectively (Table 1, Supplementary Figure S2). The LTR Assembly Index (LAI) was calculated, showing an average of 14.27 which indicates a reference level assembly according to the classification score proposed by Ou et al., (2018) [38] (Supplementary Figure S3). Then, short DNA reads were mapped to the genome assembly using the BWA software, and 99.83% of Illumina DNA short reads were mapped to the final assembly (Supplementary Table S5). Additionally, the mapping ratio of the RNA-seq data was measured for nine different tissues downloaded from public databases and generated by our laboratory. The ratio of mapped reads to the genome varies between 76 and 94%, depending on the tissue of origin (Supplementary Table S6).

Genome annotation

Repetitive elements were identified and masked in the genome assembly. Both *de novo* and homology-based annotation revealed a highly repetitive genome (76.15%, 733.29 Mb) (Supplementary Table S7). Long Terminal Repeat (LTR) retrotransposons represent the most abundant class of transposable elements (TEs), with 64.03% of all repeats. Within LTR elements, the most abundant class was Gypsy, with 47.91%. The genome also is composed of 4.25% long interspersed nuclear elements (LINEs), 10.08% simple repeats, and 0.27% short interspersed nuclear elements (SINEs) (Supplementary Table S8). Within the LINEs repeat class, L1 was the dominant repeat type. Moreover, the *L. luteus* genome is also constituted of 9.98% DNA transposable elements, of which Class II DNA elements, such as the DNA/CMC-EnSpm (~21%) family and MULE-MuDR (~16%) were the dominant DNA repeat classes.

The repeat-masked assembly was used as an input for gene model prediction and functional annotation. To visualize the genomic characteristics, a Circos plot was generated (Fig. 2), summarizing multiple genomic features across the assembled chromosomes.

Following a combined strategy of *de novo*, homology-based, and transcriptome-based methods, 36,895 protein-coding genes were identified, with an average length of 5,370 bp (Table 1). From these genes, 50,351 CDS were identified which were translated and functionally annotated. Interproscan, EggNOG, KEGG, COG, and Swissprot databases were used for functional annotation of all peptides. Through this strategy, potential functions were assigned to 35,940 (97.44%) protein-coding genes in the genome (Supplementary Table S8, and Supplementary Table S9). Among these genes, 2,492 Transcription Factors belonging to 58 families were identified (Supplementary Table S10). BUSCO annotation of predicted genes showed 93.9% and 91.7% completeness using Embryophyta and Fabales lineages (Table 1, Supplementary Figure S3).

Non-coding RNAs (ncRNAs) were predicted in the *L. luteus* genome as well. 23 microRNA (miRNA) families with putative functions based on homology with other miRNA families in other plants (Supplementary Table S11). In addition, 6,008 ribosomal RNAs (rRNAs), 1,194 transfer RNAs (tRNAs), and 1,282 small nuclear RNAs (snRNAs) were identified (Supplementary Table S12).

Comparative genomic analysis and gene family expansion

The genome was compared with other phylogenetically related Fabales (*G. max*, *L. albus*, *L. angustifolius*, and *M. truncatula*) (Fig. 3A). In addition, the *A. thaliana* genome was used as an outgroup in the phylogenetic analysis (Fig. 3B). Among the Fabales species, 26,956 orthologous groups were identified, of which 13,076 are shared (Fig. 3A). Additionally, 7,561 genes were classified in the 483 species-specific orthologous groups of yellow lupins (Supplementary Table 13). Unique gene families in the *L. luteus* genome were enriched for biological processes such as lactate transport, antibiotic transport, arsenite transport, response to arsenic, cellular response to phosphate starvation, or zeatin and trans-zeatin metabolic process (Supplementary Figure S4 and Supplementary Table S14). Next, 1,938 single-copy orthologs were used for phylogenetic reconstruction between Fabales and *A. thaliana*. Compared with these species 676/3,407 gene families were significantly expanded/contracted in *L. luteus* (Fig. 3B). Expanded gene families were enriched in several biological processes, including terpenes metabolic processes (monoterpenes, diterpenes, and sesquiterpenes), and response to biotic and abiotic stress such, flooding, herbicide, defence to bacterium and defence to insect, to jasmonic acid stimulus, and to salicylic acid stimulus (Fig. 3C, Supplementary Table S15). In addition, 30 members of MYB-related and 10 NIN-like transcription factor families were significantly expanded in the *L. luteus* genome (Supplementary Table S16). Functional analysis of these TFs indicates that they participate in the regulation of secondary metabolites, as well as response to biotic and abiotic stress and regulation of nitrogen uptake and nodulation (Supplementary Table S16). On the other hand, contracted gene families were involved in e.g. regulation of catalytic activity, protein modification, and cold acclimation (Supplementary Table S17, Supplementary Figure S5).

Orthologous groups across *L. albus*, *L. angustifolius*, and *L. luteus* were analysed. We identified 24,012 gene families in the genome of *L. luteus*, surpassing the counts of 23,640 and 23,544 gene families observed in *L. albus* and *L. angustifolius*, respectively (Fig. 3D). Notably, 20,198 gene families were found to be shared among all three *Lupinus* species, underscoring a core genomic

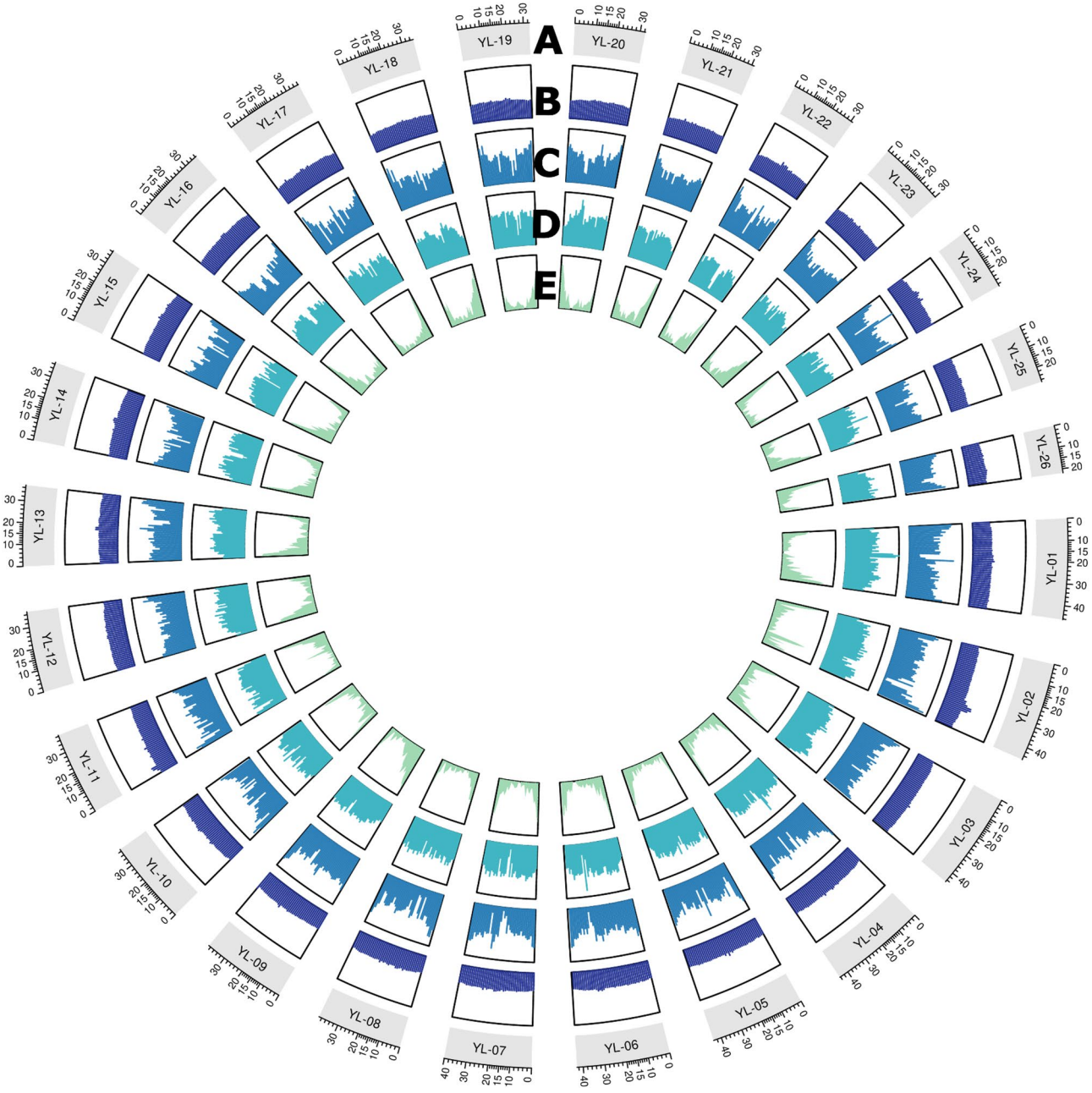

Fig. 2 Circos plot depicting the ***L. luteus*genomic features**. From outer to inner circles: (**A**) chromosome length in Mb; (**B**) GC content; (**C**) LTR repeat density; (**D**) tandem repeat density, and (**E**) gene density. All the data is shown with a window of 1 Mb

repertoire conserved across the most cultivated lupins (Fig. 3D). Remarkably, 442 unique gene families were found to be exclusive to the genome of *L. luteus*. Functional enrichment analysis of this set of unique gene families disclosed compelling associations with biological processes crucial for secondary metabolism and biotic stress response, highlighting the adaptive genomic repertoire of *L. luteus* (Supplementary Figure S6, Supplementary Table S18).

Furthermore, we detected significant expansions (362) and contractions (306) of families within the shared gene families, respectively, in *L. luteus* (Fig. 3E). Strikingly, comparative genomic analyses revealed that *L. luteus* has amplified gene families implicated in diverse biological processes, including terpene metabolism, tryptophan metabolism, defence, and innate immune responses (Fig. 3F and Supplementary Table S19), illuminating its evolutionary adaptations to environmental challenges. Notably, we elucidated the genetic basis underlying

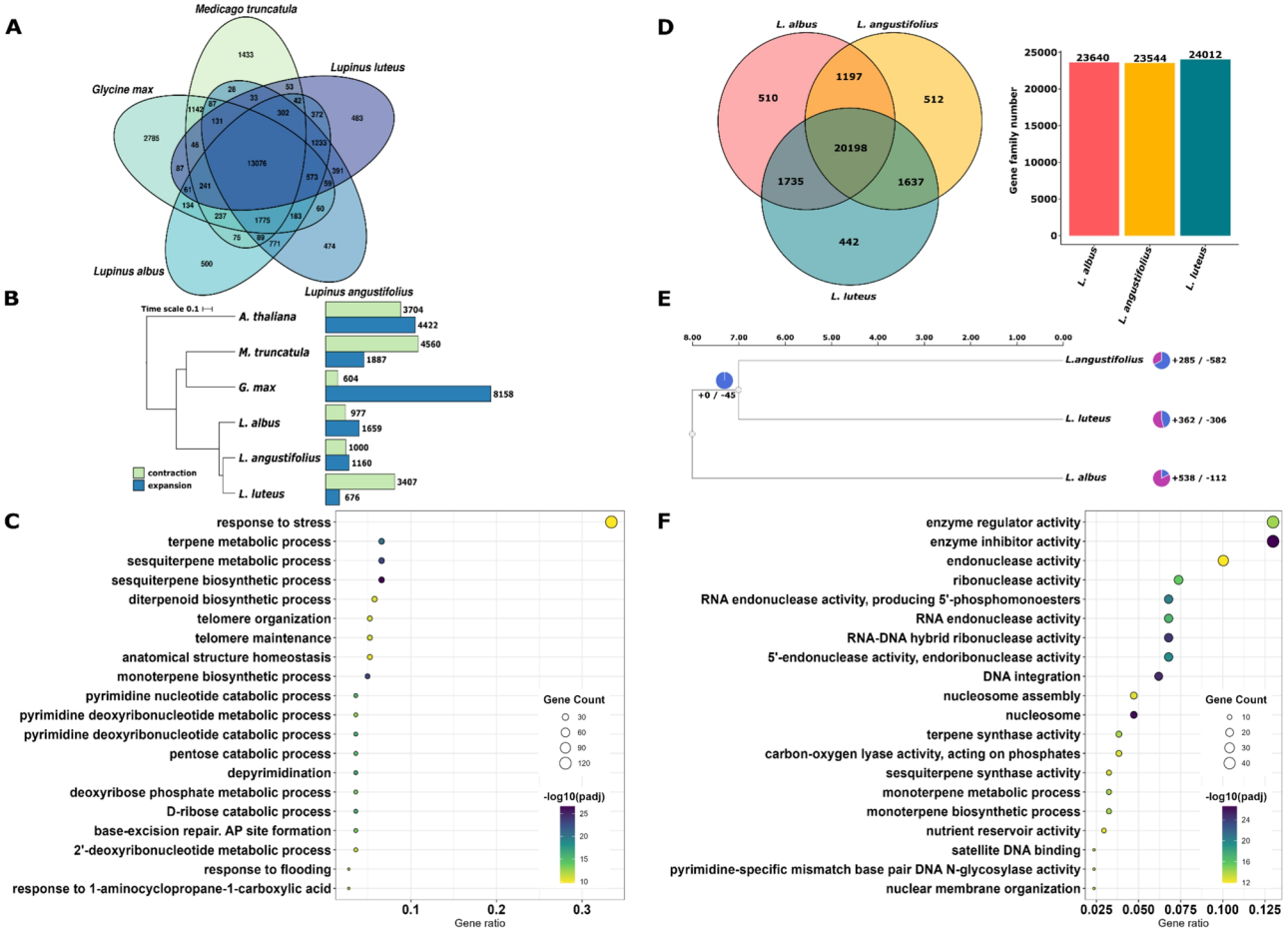

Fig. 3 Comparative Genomic Analysis of Gene Families in Lupinus Species and Related Legumes. (**A**) Venn diagram showing the number of shared and unique gene families among *L. luteus*, *L. albus*, *L. angustifolius*, *M. truncatula*, and *G. max*. (**B**) Phylogenetic analysis of six species using *Arabidopsis thaliana* as an outgroup, highlighting the number of expanded (blue) and contracted (green) gene families in each species. (**C**) Gene Ontology (GO) enrichment analysis of biological processes associated with the expanded gene families in the *L. luteus* genome. (**D**) Venn diagram displaying the shared and unique gene families among *L. luteus*, *L. albus*, and *L. angustifolius*, along with a bar plot showing the total number of gene families in each species. (**E**) Phylogenetic tree of *Lupinus* species, illustrating the number of significantly expanded (magenta) and contracted (blue) gene families. (**F**) GO enrichment analysis associated with the expanded gene families in the *L. luteus* genome

the superior protein content of *L. luteus* seeds, a trait of considerable nutritional/techno functional importance. Our comparative genomic analysis identified 30 genes encoding conglutin proteins in the genome of *L. luteus* (Supplementary Table S19 and Supplementary Table S20), surpassing the counts of 19 and 20 observed in *L. albus* and *L. angustifolius*, respectively. Of note was the significant expansion of the conglutin delta 2 family (GO:0045735; nutrient reservoir activity) in *L. luteus*, shedding light on the genetic underpinnings of its enhanced nutritional profile.

Syntenic analysis

Synteny analysis was conducted on the *L. luteus* genome in comparison to two closely related lupin species, *L. albus* and *L. angustifolius*. The analysis revealed a 1:1 synteny depth among all lupins (Fig. 4A and B and Supplementary Figure S7). However, more significant chromosomal rearrangements were evident between *L. luteus* and *L. angustifolius* compared to *L. luteus* and *L. albus* (Fig. 4A **and** Fig. 4B).

These chromosomal rearrangements include synteny disruptions, reflecting differences in synteny between *L. luteus* and its related species. Synteny analysis indicated that *L. luteus* chromosome 8 partially aligns with *L. albus* chromosomes 1 and 15 (Fig. 4C and Supplementary Figure S8A). Similarly, it showed partial alignment of *L. luteus* chromosome 8 with *L. angustifolius* chromosomes 4 and 9 (Fig. 4D and Supplementary Figure S8B.

This synteny pattern suggests that *L. luteus* chromosome 8 might have originated from the partial translocation and fusion of two ancestral chromosomes during the formation of this species (Fig. 4D and F and Supplementary Figure S8A). Further examination mapped

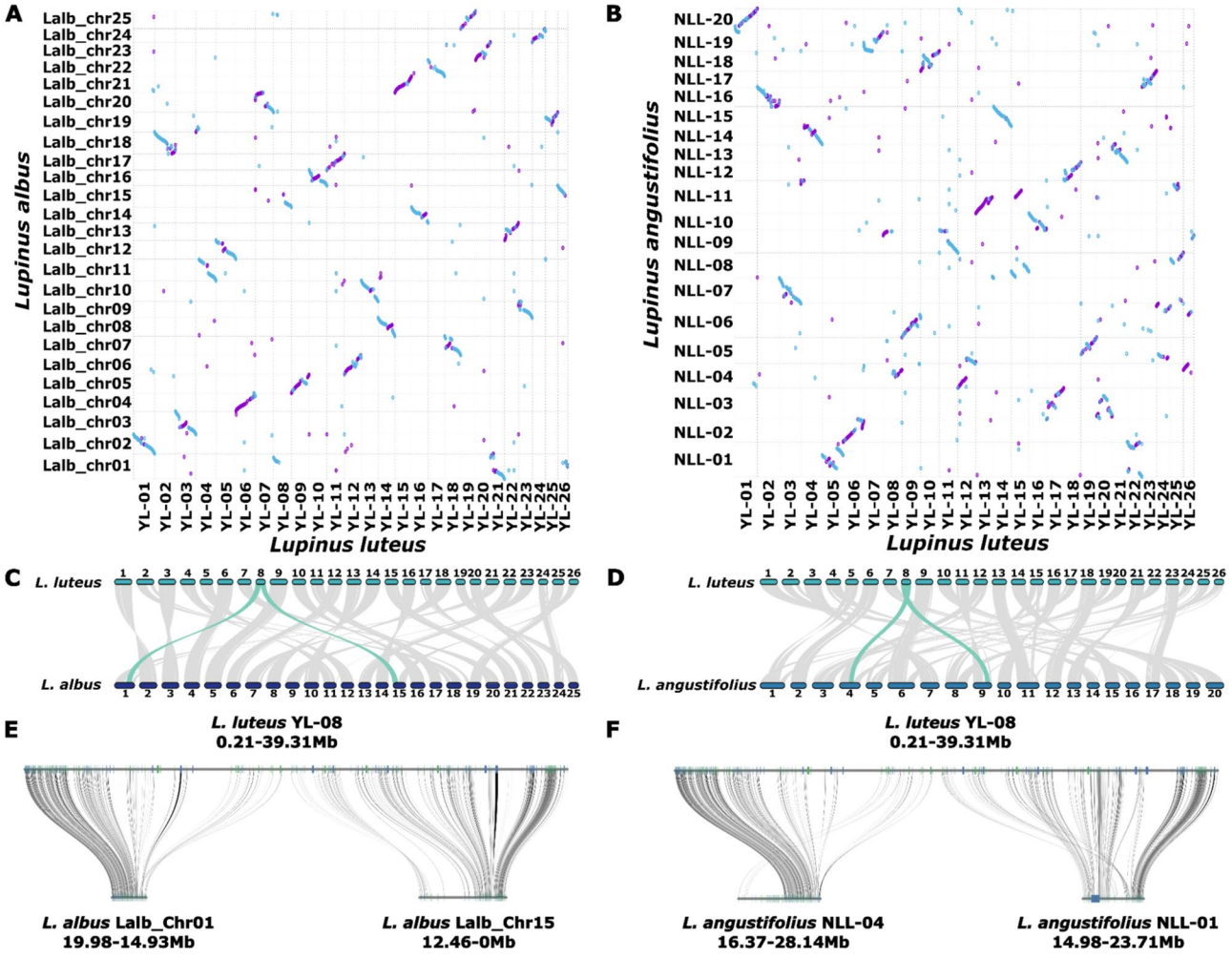

Fig. 4 Comparative synteny analysis of *L. luteus*, *L. albus*, and *L. angustifolius*. (**A**) Synteny map between *L. luteus* and *L. albus* genomes, and (**B**) between *L. luteus* and *L. angustifolius* genomes. Each dot represents a syntenic region, with purple and blue colours indicating synteny disruptions. (**C**) Macrosynteny analysis between *L. luteus* and *L. albus*, and (**D**) between *L. luteus* and *L. angustifolius*. Chromosome numbers are labelled, and each connecting line represents a syntenic block. Green lines highlight an ancestral fusion event leading to the formation of chromosome 8 in *L. luteus*. (**E**) Local synteny analysis of *L. luteus* chromosome 8 (YL-08) with its homologous regions in *L. albus* (Lalb_Chr01 and Lalb_Chr15), and (**F**) with its homologous regions in *L. angustifolius* (NLL-04 and NLL-09). Gray lines indicate conserved syntenic regions associated with protein-coding genes, while chromosomal positions are provided for reference

328 synteny blocks within *L. luteus* chromosome 8 and *L. albus* chromosomes 1 and 15, sharing an identity of 87–96% (Fig. 4E, Supplementary Figure S7A and Supplementary Table S21). Similarly, 500 blocks were identified between chromosome 8 of *L. luteus* and chromosomes 4 and 9 of *L. angustifolius*, with sequence identity ranging from 87 to 97% (Fig. 4F, Supplementary Figure S8B and Supplementary Table S22).

We found that 792 of the 945 genes encoded on chromosome 8 of *L. luteus* have orthologs on the chromosomes of origin in the genomes of *L. albus* and *L. angustifolius*. Supplementary Table S23 presents the functional annotations of chromosome 8 genes and their orthologs across different lupin species, highlighting significant genomic conservation within the *Lupinus* genomes.

Additionally, a synteny analysis between *L. albus* and *L. angustifolius* was conducted to confirm whether the synteny blocks were related to *L. luteus* chromosome 8 (Supplementary Figure S9, Supplementary Table S24). It was observed that chromosome 1 and chromosome 9, as well as chromosome 15 and chromosome 4 of *L. albus* and *L. angustifolius* respectively, exhibit partial synteny in the same coordinates shared with chromosome 8 of *L. luteus* (Supplementary Figure S9B). This observation suggests that chromosome 8 in *L. luteus* might have resulted from the partial translocation of two chromosomes during the formation of this species.

Identification of disease resistance-related (R) genes

A genome-wide analysis was conducted to investigate the presence of R genes in the genome of *L. luteus*. Following a consensus methodology using two databases and HMM profiles, we found 911 different R genes encoded in the genome. The R genes were divided into 4 major classes based on their domain structure: nucleotide binding site (NBS)-encoding proteins; receptor-like proteins (RLP); receptor-like kinases (RLK); and transmembrane coiled-coil proteins (TM-CC) [32]. 80 NBS-type, 580 RLK-type, 61 RLP, and 190 TM-CC-type proteins were identified (Supplementary Figure S10A). In addition, all R gene types were unevenly located in all the twenty-six chromosomes (Supplementary Figure S10B). Then, NBS-type R genes were analysed by their key role in host resistance to diseases [39, 40, 41]. From 80 NBS-type genes encoded in the *L. luteus* genome, the following domains were identified: 22 CC (coiled-coil)-NBS-LRR (leucine-rich repeat) (CNL); 3 NBS; 20 NBS-LRR (NL); 3 RPW8-NOD-like receptor (RNL); 5 RPW8-NBS (RPW8); 2 TIR (Toll/Interleukin-1 receptor)-NBS (TN); 11 TIR-NBS-LRR (TNL); and 14 TIR-unknown domain (TX) distributed over all chromosomes (Supplementary Table S25), except for chromosomes 02, 15, 18 and 20 (YL-02, YL-15, YL-18 and YL-20) (Fig. 5B). Here 7 clusters were observed with more than 3 genes, these clusters were present in chromosome 3 (1), chromosome 7 (3), chromosome 11 (1), chromosome 13 (1), and chromosome 19 (YL-19) (1). Transcriptome analysis of NBS-type genes in nine tissues showed differences in expression patterns (Fig. 5A). The pod wall had 22 NBS-type genes with higher expression levels than the other tissues, followed by 19 in the pedicel and 17 in the root (Fig. 5A, Supplementary Table S26).

Phylogenetic analysis of NBS-type R genes shows that the RPW8 class clustered in a clade separated from the other classes, in addition it was identified that CNL, NBS, and NL clustered into the same phylogenetic clade; on the other hand, the TN, TNL, TX classes were clustered in the same clade, and, it was concluded that gene clusters located in the same clade were more phylogenetically related (Supplementary Figure S11).

Alkaloid biosynthesis-related genes and secondary metabolite transporters

In lupins, the content of alkaloids is an important factor that affects their nutritional impact [42]. Quinolizidine alkaloids (QAs) have a role in protecting plants from pests and fungi, however, they are the principal antinutritional compounds, mainly associated with bitterness and toxicity [20]. In addition to QAs, in *L. luteus* the indole alkaloid gramine is also present and accumulated in the seeds along with the other alkaloids. A genome-wide search was performed to identify the genes related to biosynthesis, transport, and regulation of QA and gramine in the *L. luteus* genome. 15 genes were distinguished that were associated with QA biosynthesis and regulation (Table 2). Also, a 3-aminomethylindole N-methyltransferase encoding gene was identified in the genome of *L. luteus*. The expression levels of all these 16 genes were evaluated. It was observed that two crucial enzymes involved in the QA biosynthesis pathway, lysine decarboxylase (LDC) and copper amine oxidase (CAO) present higher expression in leaves and pedicels (Supplementary Table S27 and Supplementary Figure S12). The 3-aminomethylindole N-methyltransferase (NMT) gene presents higher expression in the roots and pod walls.

The transport of secondary metabolites in plants is critically important to various cellular processes, such as growth, development, survival, defence, and homeostasis. Plants possess numerous transporters that facilitate the movement of secondary metabolites within their cells and tissues. Among these transporters, four major families have been identified: ATP Binding Cassette (ABC) transporters, multidrug and toxic compound extrusion (MATE) transporters, purine uptake permease (PUP) transporters, and nitrate and peptide transporter family (NPF) transporters. Each of these transporter families plays a significant role in facilitating the transport of secondary metabolites in plants [43]. Based on a sequence homology and domain search analysis, the following were identified: 134 putative ABC transporter genes, 44 MATE transporters, 29 PUP transporters, and 69 putative NPF transporter genes (Supplementary Table S28); they are unevenly distributed along the 26 chromosomes (Supplementary Figure S13). It has been reported that purine permease transporter 1 (PUP1) encoding genes are associated with the biosynthesis, transport, and regulation of alkaloids in *Nicotiana tabacum* [44] and *L. angustifolius* [45]. An orthologous gene of *LanPUP1* was identified in the *L. luteus* genome. The expression patterns of *LluPUP1* (Llu07983) present ones similar to *LluLDC* and *LluCAO*. However, in a correlation expression analysis, a significant correlation was not found between these genes (Supplementary Figure S14 and Supplementary Table S29).

Discussion

The *Lupinus* genus includes species such as white lupin (*L. albus*), narrow-leaf lupin (*L. angustifolius*) and yellow lupin (*L. luteus*), economically important for commercial protein production, and valuable functional food crops [46]. Despite their economic and nutritional value, genomic information on *Lupinus* species remain limited, with complete genome sequences available only for *L. angustifolius* [6, 47, 48], and *L. albus* [49]. The present study provides the first chromosome-scale genome assembly for *L. luteus*, generated using a combination of Illumina, PacBio, and Hi-C sequencing platforms.

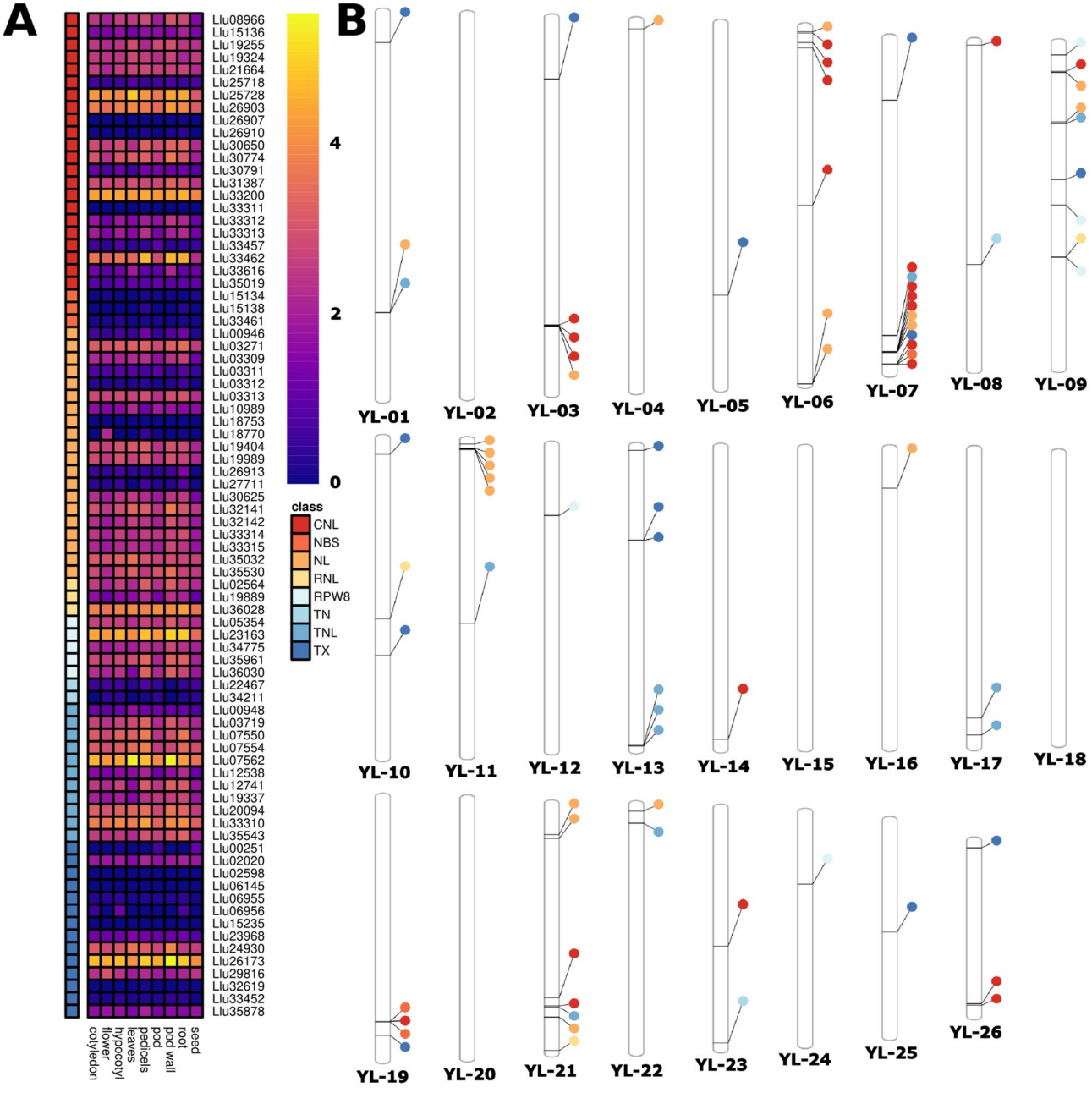

Fig. 5 Expression analysis and chromosomal location of NBS-type R genes in the *L. luteus* genome. (**A**) Expression patterns of NBS-type genes in nine different tissues of *L. luteus*. The scale colour represents log-transformed normalized TMM counts. (**B**) Chromosome location of NBS-type R genes

This high-quality genome assembly exhibits a high level of completeness in protein-coding genes and repetitive sequences, along with improved continuity in contigs and scaffolds. In comparison with other lupin species, we found that the assembled genome of *L. luteus* (962.97 Mb) was larger than the two previously reported lupin assemblies *L. angustifolius* (653 Mb) [47], and *L. albus* (451 Mb) [49]. In addition, the *L. luteus* genome had a higher repetition rate (76.15%; 733.29 Mb) than the genomes of *L. angustifolius* (34.6%; 227.6 Mb), and *L. albus* (60%; 270.58 Mb). Further, *L. luteus* has a higher scaffold N50 (33.87 Mb) than the latest version of *L. angustifolius* assembly (scaffold N50 = 30.7 Mb) [47], and *L. albus* (scaffold N50 = 17.4 Mb) [49]. The BUSCO assessment revealed 98.9% complete genes in the assembled genome, which represents a more contiguous and higher-quality genome assembly than that recently published for other lupin species. Its large genome size and high repeat content highlight the structural complexity of *L. luteus*, consistent with observations from recent genome evolution studies in legumes [50, 51]. This implies that *L. luteus* assembly will serve

Table 2 Genes related to biosynthesis, transport, and regulation of QA and Indole (gramine) in the L. luteus genome

Alkaloid	Gene	Name	Chr	Description
Quinolizidine	Llu01933	*F3H*	YL-10	Flavanone 3-hydroxylase
	Llu02941	*LDOX*	YL-10	Leucoanthocyanidin dioxygenase
	Llu07632	*LDC*	YL-14	Lysine decarboxylase
	Llu07983	*PUP1*	YL-14	Purine permease 1
	Llu08556	*MYB*	YL-14	Transcription factor MYB106
	Llu08676	*CAO*	YL-14	Copper amine oxidase
	Llu12483	*DFR1*	YL-17	Flavanone 4-reductase
	Llu16423	*AT*	YL-02	Acetyltransferase-like gene
	Llu18612	*DHDPS*	YL-21	4-hydroxy-tetrahydrodipicolinate synthase
	Llu18957	*RAP2-7.1*	YL-21	APETALA2/ethylene response transcription factor
	Llu23703	*CES1L*	YL-25	Carboxylesterase 1
	Llu25717	*CCR*	YL-26	Cinnamoyl-reductase 2
	Llu26357	*RAP2-7*	YL-03	APETALA2/ethylene response transcription factor
	Llu26798	*HMT/HLT*	YL-03	Acyltransferase tigloyl-CoA: (–)-13α-hydroxymultiflorine/(+)-13α-hydroxylupanine O-tigloyltransferase
	Llu35155	*Bet_v_I/MLP*	YL-09	Bet v I/Major latex protein domain-containing protein
Idole Gramine	Llu28618	*NMT*	YL-04	3-aminomethylindole N-methyltransferase

as a high-quality reference genome according to the LAI classification score proposed by Ou et al. [38], and offer a robust genomic platform to explore structural and functional genomics in *Lupinus*.

Comparative genomics and chromosomal evolution

Recent comparative genomic studies within the Fabales order have provided deeper insights into the evolutionary divergence and genome architecture of *Lupinus* species. Our analysis supports the view that *L. luteus* diverged relatively recently within the *Lupinus* clade. This finding is consistent with previous phylogenetic studies suggesting that *L. luteus*, *L. angustifolius*, and *L. albus* diverged approximately 8 million years ago [52]. Notably, our findings confirm that *L. angustifolius* is more closely related to *L. luteus* than to *L. albus*, consistent with recent studies demonstrating a greater degree of shared genomic features and recent common ancestry between these species [47]. Comparative synteny analyses further support this close relationship, revealing extensive chromosomal collinearity between *L. luteus* and *L. angustifolius*, in agreement with previous comparative mapping studies, which identified conserved syntenic blocks related to flowering time and anthracnose resistance [15]. However, our findings also indicate extensive chromosomal rearrangements, such as translocations and inversions, between *L. luteus* and *L. angustifolius*, which highlight the evolutionary plasticity of lupin genomes. These structural variations may underlie species-specific adaptations and influence agronomically relevant traits such as pathogen resistance and environmental resilience [53, 54].

Our study also corroborates previous findings obtained using linkage mapping and BAC-FISH (Bacterial Artificial Chromosomes Fluorescence In Situ Hybridization), which demonstrated substantial genome conservation among Old World lupins [15, 53, 54, 55]. For example, Książkiewicz et al. (2017) demonstrated that *L. albus* linkage groups Lalb01 and Lalb15 exhibited partial synteny with NLL-04 and NLL-09 in *L. angustifolius* [53]. Consistently, our data reveal a similar pattern of synteny between *L. luteus* chromosome 8 and the corresponding regions in *L. albus* (chromosomes 1 and 15) and *L. angustifolius* (chromosomes 4 and 9).

These findings suggest a shared evolutionary event, likely an historical chromosomal translocation, that led to the formation of chromosome 8 in *L. luteus*. Such event may have contributed to the functional diversification of gene families involved in responses to environmental stress and defence. Structural genome variations have recently been associated with the evolution of adaptive traits in legumes [48], and studies in *L. luteus* have shown that chromosomal rearrangements may affect resistance gene positioning and expression, further highlighting the role of genome architecture in species-specific adaptation [15, 27].

Conglutins

Lupin seeds are rich in storage proteins collectively known as conglutins, which are classified into four subfamilies α-, β-, γ-, and δ [56, 57]. These proteins not only serve as nutritional reserves during germination but have also attracted interest due to their potential functional and therapeutic properties. Among them, γ-conglutin have been particularly studied for their ability to modulate insulin metabolism and reduce blood glucose levels, suggesting potential applications in the management of type 2 diabetes [4, 58].

In our analysis, we identified a total of 30 conglutin genes in the *L. luteus* genome, nearly twice the number previously reported *L. angustifolius* [56]. Notably, this expansion is especially pronounced in the δ-conglutin subfamily, and more specifically in δ2-conglutins, which appear to have undergone lineage-specific duplication events in *L. luteus*. This finding aligns with recent transcriptomic studies showing that δ-conglutin are more highly expressed in wild *L. angustifolius* accessions than domesticated varieties, suggesting a potential role in adaptation to natural environments and biotic stress [57].

The δ-conglutin subfamily is associated with storage functions and also plant defence, due to its structural homology with the α-amylase/trypsin inhibitor family [57]. These proteins play roles in herbivore deterrence and pathogen resistance, functioning as protease inhibitors that interfere with the digestive enzymes of pests and microbes [56]. In this context, the expansion of δ2-conglutins in *L. luteus* may reflect an adaptive strategy to enhance seed protection under environmental pressure.

However, while δ-conglutin may confer defensive benefits, several studies have raised concerns regarding their low digestibility and allergenic potential in humans [59, 60, 61]. These properties pose challenges for food safety and nutritional applications, especially given the growing interest in lupins as alternative protein sources. Interestingly, proteomic analyses in *L. luteus* cv. Tapper has revealed that δ-conglutin are the predominant storage proteins in this species, contrasting with the dominance of β-conglutins in *L. albus* and *L. angustifolius* [62]. These findings lead us to hypothesise that this species-specific variation in storage protein composition and the expansion of the δ-conglutin gene family in *L. luteus* results from both evolutionary and functional divergence.

Terpene metabolism

Our comparative genomic analysis revealed a significant expansion of gene families involved in terpene metabolism in *L. luteus*, particularly those associated with the biosynthesis of monoterpenes, diterpenes, and sesquiterpenes.

Terpenes constitute one of the largest and most functionally diverse classes of secondary metabolites in plants. They play a crucial role in plant defence mechanisms, acting as antimicrobial, antifungal, and insect-repellent compounds [63, 64]. In addition to their antimicrobial and insect-repellent activities, terpenes can act as signalling molecules that mediate complex interactions with the environment and other organisms [64].

The enrichment of terpene biosynthesis genes in *L. luteus* suggests that this species has evolved enhanced chemical defence strategies, which may contribute to its adaptive response to environmental pressures. These findings are consistent with recent reports in legumes and other crop species, where terpene-based defences have been linked to pathogen resistance and herbivore deterrence [65]; suggesting that *L. luteus* has developed a robust multi-layered defence system, integrating secondary metabolite production and hormonal signalling pathways to enhance survival in diverse environmental conditions.

The expansion of terpene biosynthesis genes in *L. luteus* not only highlights its evolutionary adaptation to environmental challenges but also suggests potential agronomic applications. Some terpenes, such as phytoalexins and volatile organic compounds (VOCs), contribute to allelopathic interactions that influence plant-plant competition and soil microbiome composition [66]. These compounds can modulate the composition and activity of the soil microbiota, thereby affecting nutrient cycling and plant health. In *L. luteus*, the expansion of terpene synthase genes may have implications not only for pathogen resistance but also for interactions with beneficial microbes, such as rhizobia or mycorrhizal fungi, which are known to be influenced by root-emitted volatiles.

Moreover, certain terpenes, such as triterpenes and diterpenoids, are known for their antifungal properties. Recent studies have proposed their use as natural biopesticides in integrated pest management strategies [67, 68], which may contribute to enhancing disease resistance in lupin crops. The identification of these biosynthetic pathways in *L. luteus* provides an opportunity to explore their exploitation for sustainable agriculture.

While our genomic findings provide evidence of terpene gene family expansion, functional validation through transcriptomic analysis under biotic and abiotic stress conditions, as well as metabolomic profiling, will be essential to elucidate the specific contributions of these compounds to stress resilience and crop improvement. Transcriptomic analyses under biotic (e.g., *Colletotrichum lupini* infection) and abiotic (e.g., drought, salinity) stress conditions will help identify stress-inducible terpene synthases, and metabolomic profiling will allow correlation of gene expression with terpene accumulation. Additionally, spatial expression studies could reveal tissue-specific deployment of these defences, informing breeding programs aiming to combine resistance with food safety and yield stability.

Disease resistance genes

In plants, nucleotide-binding site leucine-rich repeat (NBS-LRR) proteins constitute a significant class of intracellular immune receptors, known as resistance (R) proteins, that play a critical role in the recognition of pathogen-derived effectors and the activation of downstream defence responses [32]. In our analysis, we

identified 80 NBS-type R genes in the *L. luteus* genome, exceeding the 67 reported in *L. angustifolius* [47], suggesting a moderate expansion of this gene family in *L. luteus*. This observation is particularly relevant given that previous comparative genomic studies in *L. angustifolius* and *L. albus* reported an overall underrepresentation of R genes in lupins relative to other legume species [47].

Interestingly, a large proportion of the R genes in *L. luteus* were found in tandem arrays, consistent with patterns observed in other legumes such as *Glycine max* and *Medicago truncatula*, where tandem duplication events are a known driver of R gene diversification [32, 69]. Many R genes are clustered together and share a closer phylogenetic relationship. Gene clustering facilitates the evolution of novel resistance specificities and allows for coordinated transcriptional regulation under stress conditions. This suggests that tandem duplication plays a significant role in the expansion of the R gene family in *L. luteus*, potentially reflecting species-specific immune adaptations. In other legumes, gene clustering has been associated with increased pathogen recognition capabilities [69], indicating that a similar process may have contributed to pathogen-driven selection in *L. luteus*. Supporting this, recent studies have provided functional evidence supporting the role of NBS-LRR resistance genes in *L. luteus* against *Colletotrichum lupini*, one of the most devastating fungal pathogens affecting lupin crops [27]. The *LluR1* gene, a TIR-NBS-LRR orthologue of *LanR1* from *L. angustifolius*, was significantly upregulated in resistant *L. luteus* genotypes upon pathogen infection. Functional enrichment analysis also revealed the induction of genes involved in phenylpropanoid biosynthesis, a pathway associated with strengthening cell walls and producing antimicrobial compounds, suggesting a coordinated defence response [27].

Mechanistically, NBS-LRR proteins function as intracellular immune receptors, detecting pathogen-derived effectors and activating downstream defence pathways. Specifically, TIR-NBS-LRR proteins typically trigger Effector-Triggered Immunity (ETI), which induces localized hypersensitive response (HR) and activates salicylic acid (SA)-dependent defence pathways [32]. In *L. luteus*, resistance against *C. lupini* appears to be mediated by *LluR1*, which induces genes involved in phenylpropanoid biosynthesis, contributing to reinforcing cell walls and producing antimicrobial compounds [27]. The activation of *LluR1* in *L. luteus* reflects this canonical immune response but with species-specific regulatory dynamics. Indeed, expression analyses of *LanR1* in *L. angustifolius* revealed divergent regulatory patterns in response to *C. lupini*, suggesting that even conserved R gene orthologues may be differentially deployed among lupin species [70].

These findings further emphasize the uniqueness of the R protein repertoire in lupins and suggest that evolutionary divergence among species has shaped distinct immune response strategies. The observed differences in R gene copy number, clustering, and expression regulation across *Lupinus* species highlight the evolutionary flexibility of their immune systems. In particular, the expansion and tandem duplication of R gene clusters in *L. luteus* may reflect species-specific adaptation to pathogen pressure and environmental challenges. Future studies could explore gene expression analysis under pathogen infection conditions to determine whether these expanded R gene clusters are actively involved in resistance responses, shedding light on their potential functional diversification to determine whether these R gene clusters are actively involved in defence responses and uncover the regulatory mechanisms driving their diversification.

Alkaloid biosynthesis

Alkaloids contribute to plant defence but also influence the bitterness of lupin seeds, affecting their nutritional and agronomic value [20, 24, 25, 71]. Using domain-based protein search and sequence similarity analysis, we identified 16 genes in *L. luteus* previously associated with quinolizidine alkaloid (QA) biosynthesis and regulation in other lupins. These include key enzimes such as Lysine Decarboxylase (*LDC*) and Copper Amine Oxidase (*CAO*) involved in the early biosynthetic steps of QA biosynthesis [20, 22, 45, 72]. We also identified genes encoding acetyltransferase-like enzymes (*AT*) and N-methyltransferase (*NMT*), which participate in the structural modification of QAs, affecting their bioactivity, toxicity, and accumulation in seeds [20, 22, 45, 72]. Specifically, NMT catalyses the methylation of alkaloid precursors and plays a crucial role in generating bioactive QA derivatives [73]. Similarly, amide synthases (e.g., acetyltransferases) contribute to chemical diversification and solubility of alkaloids, influencing their tissue localization and transport efficiency, affecting alkaloid content in seeds [74].

Interestingly, the two key enzymes crucial for QA biosynthesis, lysine/ornithine decarboxylase (*LDC*) and copper amine oxidase (*CAO*), were predominantly expressed in tissues such as leaves and pedicels [20, 75, 76]. In this way, similar to *L. angustifolius*, QA biosynthesis in *L. luteus* primarily occurs in aerial tissues, with subsequent transportation to the seeds [77]. This spatial separation underscores the importance of regulatory mechanisms controlling temporal and tissue-specific gene expression, which may offer targets for metabolic engineering or selective breeding.

The translocation of QAs in lupins has been previously investigated in species such as *Lupinus polyphyllus*, where lupanine accumulates in stems and petioles after

biosynthesis in leaves [75]. A recent study proposed that a purine permease transporter (*PUP1*) may be involved in this process, as its expression correlates with QA biosynthetic gene clusters in *L. angustifolius* [20]. Based on this, we searched for orthologues of this transporter in *L. luteus* and identified Llu07983, an orthologue of *LanPUP* (TanjilG_28431). Although its expression was highest in leaves, pedicels, and cotyledons, no strong correlation was observed with core QA biosynthetic genes. Nevertheless, the co-expression patterns suggest a possible role in QA transport, highlighting the complexity of secondary metabolite trafficking in lupins.

Our findings also indicate that QA biosynthesis genes are active in aerial tissues, suggesting that transport mechanisms are crucial in alkaloid accumulation in seeds. These findings prompt further investigation into the functional role of this ortholog in QA translocation within *L. luteus*, highlighting the complexity of alkaloid transport mechanisms in lupins.

The relationship between alkaloid biosynthesis and seed quality is a key factor to improving *L. luteus* as a food and feed crop. While alkaloids contribute to plant defence, their presence in high concentrations negatively impacts seed palatability and safety for human and animal consumption [20, 24, 25, 71]. Our study revealed that *L. luteus* possesses a diverse and functionally redundant set of QA-related genes, which may underlie the variation in alkaloid content observed among different cultivars. Recent studies have also shown that QA accumulation among different *L. luteus* cultivars is strongly influenced by environmental conditions such as temperature, soil type, and water availability, reinforcing the role of genotype-environment interactions [17, 78]. The interaction between genotype and environment is particularly relevant for breeding low-alkaloid cultivars adapted to climate change [17, 78].

This variability is particularly relevant for the development of low-alkaloid varieties adapted to changing climatic scenarios. Among quinolizidine alkaloids, lupinine, sparteine, lupanine, and multiflorine are the most frequently detected in *L. luteus* [17]. These alkaloids contribute to plant defence against herbivores and pathogens but also pose challenges for food and feed safety due to their bitter taste and potential toxicity [17]. Interestingly, indole alkaloids such as gramine, are rare in lupins, but is present in *L. luteus*, supporting the idea of species-specific diversification in alkaloid biosynthetic pathways.

The presence of indole alkaloids in lupins are not common. In a reset work by Valente et al. (2024) demonstrate that *L. luteus* was the unique species that presented this type of alkaloid suggests a determinant role in the defence and ecological interactions [78]. These alkaloids, particularly gramine, are known for their toxicity and deterrent effects against a range of herbivores and insect pests, contributing to the resistance against predation [79]. In addition, gramine exhibits antimicrobial properties, which may help *L. luteus* defend itself against bacterial and fungal pathogens.

The indole alkaloid pathway begins with the amino acid tryptophan as a precursor, and the biosynthesis of various indole alkaloids occurs through distinct steps. The enzyme 3-aminomethylindole N-methyltransferase (NMT) plays a central role in the biosynthesis of indole alkaloids [80]. It catalyzes the N-methylation of 3-aminomethylindole (AMI), producing N-methyl-3-aminomethylindole (MAMI), which leads to the formation of gramine. This pathway originates from the amino acid tryptophan and is involved in the production of compounds essential for plant defence [81, 82]. The methylation reaction uses S-adenosylmethionine (SAM) as the methyl donor, a mechanism commonly found in plant alkaloid biosynthesis [83].

Recent studies have characterized NMT genes in species such as barley. These genes are essential for the final step in gramine synthesis [81]. Similar pathways and enzymatic activities have been reported in other plants that produce indole alkaloids. This suggests that NMTs are broadly conserved and crucial in specialized metabolism. Their substrate specificity and enzymatic function have likely evolved in response to ecological pressures, allowing different species to fine-tune their alkaloid profiles [84, 85].

In *L. luteus*, direct studies on NMTs are limited. However, based on findings from related species, it is likely that similar enzymes are present and functional. *L. luteus* produces indole alkaloids, epecially gramine and gramine derivative in seeds and leaves [78, 86], and it may use NMTs in comparable biosynthetic routes. These enzymes could be involved in defence mechanisms and possibly in producing compounds with cytotoxic activity [87, 88]. Evidence from transcriptomic comparisons with other legumes supports this hypothesis [89]. Our results demonstrate, for the first time, the presence of a key enzyme involved in the biosynthetic pathway of indole alkaloids, thereby opening new avenues for investigating the role of 3-aminomethylindole N-methyltransferase (NMT) in the synthesis of gramine in Lupinus luteus. Furthermore, we detected the expression of this enzyme across multiple tissues, with notably high levels in roots and pod walls. These findings, combined with emerging evidence, suggest that indole alkaloids may contribute to allelopathic interactions, potentially suppressing the growth of neighbouring plant species and conferring a competitive ecological advantage to *L. luteus*.

NMT activity also appears to be influenced by environmental conditions such as abiotic stress [81]. This highlights the dual role of NMTs in both metabolism and stress response. Their contribution goes beyond

biochemistry, as they help plants adapt to changing environments.

Conclusions and future prospects

This study provides a comprehensive genomic framework for *L. luteus*, delivering this species' first high-quality chromosome-scale genome assembly. The assembly reveals key features of genome organization, including large-scale structural rearrangements, recent gene family expansions, and a high level of completeness and continuity, positioning it as a valuable reference for future comparative, functional, and evolutionary studies in the *Lupinus* genus and other legumes.

Our findings clarify several important aspects of *L. luteus* biology. First, comparative genomic analyses underscore the evolutionary proximity of *L. luteus* to *L. angustifolius* while revealing extensive chromosomal translocations that may underlie species-specific adaptations. Second, the expansion and clustering of NBS-type resistance (R) genes suggest the evolution of a unique immune repertoire in *L. luteus*, likely shaped by selective pressures from pathogens such as *C. lupini*. Third, we identified a significant increase in genes involved in terpene biosynthesis, which may contribute to enhanced chemical defence and ecological interactions. Fourth, the conglutin gene family, particularly δ-conglutin, was found to be expanded in *L. luteus*, potentially influencing both nutritional traits and allergenic potential. Finally, the diversity and tissue-specific expression of alkaloid biosynthetic genes and identifying potential transporters highlight the complex regulation and ecological importance of quinolizidine alkaloids.

Collectively, these insights advance our understanding of *L. luteus* genome biology and open new avenues for targeted crop improvement. Future research should prioritize the functional validation of candidate genes through transcriptomic profiling, stress-induced expression analysis, and reverse genetics approaches such as CRISPR/Cas9. In addition, integrating metabolomic data will be essential to link gene function with phenotypic traits such as disease resistance, alkaloid content, and nutritional quality.

This genomic resource lays the foundation for breeding *L. luteus* cultivars with enhanced resistance to pathogens, reduced antinutritional compounds, improved protein profiles, and greater environmental adaptability. In the context of climate change and increasing demand for plant-based proteins, *L. luteus* represents a promising candidate for sustainable agriculture, functional foods, and resilient cropping systems.

Materials and methods

Plant materials and sequencing

Lupinus luteus cultivar AluProt-CGNA® (National Official Register Number 47/13, SAG, Chile), developed by the Centro de Genómica Nutricional Agroacuícola (CGNA) and denoted as C195 in the text, was used as plant material in this study. The formal taxonomic identification of the cultivar was performed by Haroldo Edinson Salvo Garrido, breeder and creator of the cultivar. A voucher specimen has not been deposited in a public herbarium, as the material is proprietary to CGNA. Researchers interested in accessing the plant material or further information may contact CGNA directly.

Seeds were scarified and hydrated for 20 min in water, washed with 70% ethanol for one minute, then in 2.5% sodium hypochlorite, and finally washed three times with sterile distilled water. Then, seeds were put in Petri dishes with moist filter paper (Whatman) and kept in darkness at 23 °C for 24 h to germinate. The germinated seeds were grown in controlled conditions with a 16-h day and 8-h night photoperiod at 23 °C. The nucleic acid extractions were performed from young leaves (~60 mg). According to the manufacturer's instructions, DNA extraction was performed with Quick-DNA™ HMW MagBead and Quick-DNA™ Plant/Seed Kits (Zymo Research). The quantity and quality ratios (260/280 & 260/230) of nucleic acid were measured by absorbance in the Synergy HTX multi-plate reader with a Take3 Trio plate (Biotek), and the integrity was assessed through 1% agarose gel electrophoresis run at 100 V for 1 h. After extraction, ~500 ng of gDNA was sequenced on the Illumina NovaSeq 6000 and PacBio Sequel II platforms.

RNA data were extracted from three tissues (leaf, hypocotyl, and cotyledon). According to the manufacturer's instructions, RNA extraction was performed separately with a Quick-RNA™ plant/seed kit (Zymo Research) from ~150 mg of fresh hypocotyl and cotyledon, and ~60 mg of young leaves. Then, the tissues were frozen with liquid nitrogen and powdered by grinding in a mortar. In addition, a 30 min DNAse digestion was performed. RNA concentration and A260/A280 were measured by absorbance in Synergy HTX multi-plate reader (Biotek). The integrity of RNA assessment was performed through 2% agarose gel electrophoresis run at 100 V for 1 h. PE150 sequencing was performed in the Illumina Novaseq 6000 platform.

A Hi-C library was generated as follows: (i) Fresh, young leaves were treated with formaldehyde to fix the conformation of DNA. (ii) Cells were lysed, and cross-linked DNA was digested by the restriction enzyme *Dpn*II. (iii) Digested fragments were ligated and biotinylated. (iv) Purify and fragment the DNA in random fragments of ~300–500 bp. Finally, the Hi-C library was also

sequenced based on the Illumina NovaSeq 6000 platform in PE150 mode.

Genome assembly and assessment

The estimation of genomic characteristics was determined before assembly, based on Illumina DNA short reads data. Low-quality reads were filtered using SOAPnuke v2.1.8 [90]. Then, the cleaned data were used for *K-mer-based* frequency distribution in Jellyfish v2.3.0 [91]. To estimate the genome size, heterozygosity, and repeat content of the *L. luteus* genome, GenomeScope v2.0 was used [92].

For the novo assembly, PaBio HiFi long reads were cleaned and assembled using the HifiASM v0.19.6 [36, 93] assembler with an aggressive purge-dups option. Then, the contigs were polished using two rounds of nextpolish v1.4.1 [37], employing Illumina clean reads.

Hi-C data were used to build a chromosome-level genome assembly. First, Hi-C reads were trimmed and cleaned for adapters and low-quality reads. The cleaned data were aligned and filtered using Juicer v1.6 [94] with default parameters. The ordering and orientation of contigs were carried out using 3D-DNA v180922 [95].

The Burrows-Wheeler Aligner (BWA) v0.7.17-r1188 [96] with default parameters was used to align Illumina reads to the final assembly for estimating the coverage ratio. Completeness and quality of the genome were evaluated using QUAST v5.2.0 [97] and BUSCO v5.4.7 [98] by using the Embryophyta and Fabales linages database.

Genome annotation

Repetitive DNA regions were predicted from the assembled genome sequences using a combination of homologous and ab initio prediction. First RepeatMasker v.4.1.5 [99] was used for a homology search against known repeat sequences in Repbase database v28.08 [100]. For ab initio prediction, a custom-built *de novo* repeat database with RepeatModeler v2.0.4 [101] was produced and then annotated with RepeatMasker.

Non-coding RNA (ncRNA) was predicted using StructRNAFinder [102] with ViennaRNA v2.6.3 [103], Infernal v1.1.4 [104], and Rfam 14.9 (November 2022, 4108 families) [105] as reference for rRNA, microRNA and other ncRNA classes prediction. tRNA annotations were carried out with tRNAscanSE v2.0.12 [106].

Protein-coding genes were predicted by integrating ab initio, homology-based, and transcriptome-based annotation. *De novo* gene prediction was conducted using Augustus v3.5.0 [107] employing the Arabidopsis training gene model, and, v4.71 [108]. For RNA-seq-based prediction, RNA-seq data of *L. luteus* from different tissues (see Supplementary Table S6) were downloaded from NCBI SRA (accessed on August 4th, 2023). Next, low-quality reads were filtered using fastp v0.23.2 [109], and the remaining high-quality RNA-seq reads were mapped to the assembly of the C195 genome using hisat2 v2.2.1 [110]. Mapped reads were then employed for genome-based transcript assembly using Cufflinks v2.2.1 [111] and Stringtie v2.2.1 [112]. Afterwards, all transcripts were used to generate a transcriptome assembly using PASA v2.5.3 [113, 114].

For the protein-based homology search, MMseqs2 v14-7e284 [115] and GeMoMa v1.9 [116] were used to annotate the gene models by comparing with the predicted protein sequences of *Arabidopsis thaliana* (TAIR10), *Glycine max* (Williams), *Lupinus albus* (Amiga), *Lupinus angustifolius* (Tanjil), and *Medicago truncatula* (Jemalong A17) downloaded from NCBI (downloaded on August 17th, 2023). The results of these three strategies were combined by EVidence Modeler v2.1.0 [113].

The predicted protein-coding genes were functionally annotated through several methods, such as InterProscan v5.62-94.0 [117] for protein domain annotation, EggNOG-mapper v2.1.11-1 [118] with EggNOG v5.0.2 [119] as a reference database and BLAST v2.14.1 [120], for queried protein sequence against all UniProt Release 2023_05 (SwissProt + Trembl) database [121] (downloaded on September 25th, 2023), using a threshold of E-value < 1e-10. KEGG Orthologs (KO) were assigned using KOfamScan v1.3.0 with the KOfam HMM database as a reference [122]. Finally, an in-house developed R script was used to merge all annotation files and create a GO file annotation for enrichment analysis.

To identify Resistance genes (R-genes) in the *L. luteus* genome. First, protein-coding genes were scanned using RGAugury v2.2 [123], and DRAGO2 (DRAGO2-API) deposited in PRGdb 3.0 [124] (accessed on November 10th, 2023) with default parameters. Then, the consensus annotated R-genes were selected to detect the nucleotide-binding site leucine-rich repeat (NBS-LRR) proteins. All sequences of RefPlantNLR v20210712_481 [125] were downloaded and the NBS (NB-ARC) domain (PF00931) HMM profile from Pfam v36.0 [126]. Next, following the methods of Yang et al. [127], BLASTp using DIAMOND v2.0.14.152 [128] (E-value = 1e-10, query coverage = 80, subject coverage = 50) and hmmsearch (E-value = 1e-5) from HMMER v3.4 [104] were used to determine the NBS-LRR candidates genes. Finally, all four outcomes were merged to distinguish if an R-gene belonged to the NBS class. The NBS subclass was annotated from RGAugury and DRAGO2 outcomes.

The alkaloid biosynthesis pathway and related genes in the *L. luteus* genome were identified using information derived from different publications associated with different lupin species [22, 45, 72, 77]. First, all sequences of the genes related to alkaloid production in lupins were downloaded and made a non-redundant database for all

genes. For gramine indole alkaloid, all biosynthesis genes associated were downloaded from the KEGG pathways database [129, 130, 131]. Then, a search was performed using the same BLASTp parameters employed for the R-genes search.

Secondary metabolite transporter families were characterised according to those described by Nogia and Pati, where they noted that the ABC, MATE, PUP, and NTR/PTR (NPF) families were mainly responsible for the transport of secondary metabolites in plants [43]. First, Transporter Classification Database (TCDB) [132] data was downloaded (on November 20th, 2023). The TCDB sequence was used to search the potential secondary metabolite transporters in the *L. luteus* genome, with the same BLASTp parameters search described above. In addition, the HMM profiles of MATE (MatE) domain (PF01554), PUP (PUNUT) domain (PF16913), and NPF (PTR2) domain (PF00854), were downloaded from Pfam and used to search in protein-coding genes with hmmsearch using the same parameters described above. Likewise, a dataset of manually selected ABC transporters, obtained from Hou et al. [133] was queried with the same search parameters used with TCDB. Transmembrane regions of transporter proteins were predicted by DeepTMHMM v1.0.24 [134].

Comparative genomic analysis and evolutionary analysis

The genomes of *A. thaliana*, *G. max*, *L. albus*, *L. angustifolius*, and *M. truncatula* were collected and used for comparative genomics and phylogenetic analysis of the *L. luteus* genome. First, OrthoFinder v2.5.5 [135] was employed to identify ortholog groups with *A. thaliana* as an outgroup. Then a set of single-copy gene orthogroups, detected in the previous step, were used for phylogeny reconstruction. The protein sequences were aligned using MAFFT v7.490 [136] with default parameters. These data were employed for phylogenetic tree inference using PhyML v3.3.3 [137]. The divergence time was calculated and calibrated as reported by Jiu et al. [41], using MCMCtree from PAML v4.9j [138] and TimeTree v5 [139]. To measure the expansion and contraction of gene families in the *L. luteus* genome, CAFE v5.1.0 [140] was used. GO enrichment analysis was performed on unique genes, and expanded and contracted families using the Cytoscape app BiNGO v3.0.5 [141] in Cytoscape v3.10.1 [142].

Collinearity was performed by comparing the *L. luteus* genome with their relatives *L. albus* and *L. angustifolius* using the methods proposed by Jiu et al. [41] using MUMmer v4.0.0 [143]. Then, to evaluate synteny blocks between the three genomes MCScan [144] was employed through the JCVI package v1.3.8 [145] with default parameters.

Expression analysis

All RNA-seq raw data files (Supplementary Table S6) corresponding to nine tissues of *L. luteus* were filtered to eliminate low-quality reads and mapped to the reference genome using the same protocol explained above. Gene count matrices for each experiment were calculated using feature Counts v2.0.6 [146]. Next, normalised counts TMM were obtained with EdgeR v4.0.3 [147]. The batch effect was removed by applying parametric empirical Bayes frameworks adjustment with SVA v3.50.0 [148]. Figure S15 summarizes the entry pipeline employed in this work.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11678-8.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

Not Applicable.

Author contributions

H.S.G. Scientific design of the research, planned scientific analysis, experiment and decided plant material. Oriented and guided the writing, results and discussions of the paper. Responsible for the research and overall project. J.E.M. Designed, performed bioinformatics analysis and writing. D.L. Responsible for laboratory methods, guide DNA samples preparation and discussion with external companies. Writing methods and paper discussions. P.D.S.C. Provided extensive revisión of the manuscript and scientific discussions. G.A. Contributed with paper preparation and scientific discussions. A.R., M.C., S.H. and F.W. Established and maintained plant material in a controlled environment and composts. Performed genomic DNA extraction measurement and qualifications, samples preparations. T.M. and G.L. Contributed with bioinformatics analysis. All authors read the final version of the manuscript and approved its publication.

Funding

This research was funded by the Agencia Nacional de Investigación y Desarrollo de Chile under grant number R20F0003. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability

The raw genome sequencing data of ***L. luteus*** are available at National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/) under the following BioProject accession numbers: PRJNA1087712, PRJNA1087709, PRJNA1087391, PRJNA1087246, and GCA_964019355.1. All data are available from the corresponding author upon request.

Declarations

Ethics, consent to participate, and consent to publish declarations

All plant materials used in the current study complied with the relevant institutional, national, and international guidelines and legislation. This manuscript does not contain any studies with human participants or animals performed by the authors. These methods were carried out in accordance with relevant guidelines and regulations.

Competing interests

The authors declare no competing interests.

Author details
[1]CGNA (Agriaquaculture Nutritional Genomic Center), Las Heras 350, Temuco 4781158, Chile
[2]Present address: Núcleo de Investigación en Data Science, Facultad de Ingeniería y Negocios, Universidad de Las Américas, Santiago 7500975, Chile
[3]Departamento de Genética Molecular y Microbiología, Pontificia Universidad Católica de Chile, Santiago 8331150, Chile
[4]Biocontrol Research Laboratory, Universidad de La Frontera, Temuco 4811230, Chile

Received: 17 July 2024 / Accepted: 6 May 2025
Published online: 14 May 2025

References

1. Kroc M, Tomaszewska M, Czepiel K, Bitocchi E, Oppermann M, Neumann K, et al. Towards development, maintenance, and standardized phenotypic characterization of single-seed-descent genetic resources for lupins. Curr Protoc. 2021;1:e191.
2. Nevado B, Atchison GW, Hughes CE, Filatov DA. Widespread adaptive evolution during repeated evolutionary radiations in new world lupins. Nat Commun. 2016;7:12384.
3. Musco N, Cutrignelli MI, Calabrò S, Tudisco R, Infascelli F, Grazioli R, et al. Comparison of nutritional and antinutritional traits among different species (Lupinus albus L., Lupinus luteus L., Lupinus angustifolius L.) and varieties of lupin seeds. J Anim Physiol Anim Nutr (Berl). 2017;101:1227–41.
4. Ishaq AR, El-Nashar HAS, Younis T, Mangat MA, Shahzadi M, Ul Haq AS, et al. Genus Lupinus (Fabaceae): a review of ethnobotanical, phytochemical and biological studies. J Pharm Pharmacol. 2022;74:1700–17.
5. Carvajal-Larenas FE. Nutritional, rheological and sensory evaluation of Lupinus mutabilis food products - a review. Czech J Food Sci. 2019;37:301–11.
6. Hane JK, Ming Y, Kamphuis LG, Nelson MN, Garg G, Atkins CA, et al. A comprehensive draft genome sequence for lupin (Lupinus angustifolius), an emerging health food: insights into plant-microbe interactions and legume evolution. Plant Biotechnol J. 2017;15:318–30.
7. Pueyo JJ, Quiñones MA, de la Coba T, Fedorova EE, Lucas MM. Nitrogen and phosphorus interplay in lupin root nodules and cluster roots. Front Plant Sci. 2021;12:644218.
8. Burgos-Diaz C, Wandersleben T, Olivos M, Lichtin N, Bustamante M, Solans C. Food-grade Pickering stabilizers obtained from a protein-rich lupin cultivar (AluProt-CGNA®): chemical characterization and emulsifying properties. Food Hydrocolloids. 2019;87:847–57.
9. Parra-González LB, Aravena-Abarzúa GA, Navarro-Navarro CS, Udall J, Maughan J, Peterson LM, et al. Yellow lupin (Lupinus luteus L.) transcriptome sequencing: molecular marker development and comparative studies. BMC Genomics. 2012;13:425.
10. Martínez-Villaluenga C, Frías J, Vidal-Valverde C. Functional lupin seeds (Lupinus albus L. and Lupinus luteus L.) after extraction of α-galactosides. Food Chem. 2006;98:291–9.
11. Glazinska P, Kulasek M, Glinkowski W, Wysocka M, Kosiński JG. LuluDB-The database created based on small RNA, transcriptome, and degradome sequencing shows the wide landscape of Non-coding and coding RNA in yellow lupine (L.) flowers and pods. Front Genet. 2020;11:455.
12. Abraham EM, Ganopoulos I, Madesis P, Mavromatis A, Mylona P, Nianiou-Obeidat I et al. The use of lupin as a source of protein in animal feeding: genomic tools and breeding approaches. Int J Mol Sci. 2019;20.
13. Pexas G, Doherty B, Kyriazakis I. The future of protein sources in livestock feeds: implications for sustainability and food safety. Front Sustain Food Syst. 2023;7.
14. van de Noort M, Lupin. An important protein and nutrient source. Sustainable protein sources. Elsevier; 2024. pp. 219–39.
15. Lichtin N, Salvo-Garrido H, Till B, Caligari PDS, Rupayan A, Westermeyer F, et al. Genetic and comparative mapping of Lupinus luteus L. highlight syntenic regions with major orthologous genes controlling anthracnose resistance and flowering time. Sci Rep. 2020;10:19174.
16. Iqbal MM, Huynh M, Udall JA, Kilian A, Adhikari KN, Berger JD, et al. The first genetic map for yellow lupin enables genetic dissection of adaptation traits in an orphan grain legume crop. BMC Genet. 2019;20:68.
17. Valente IM, Sousa C, Almeida M, Miranda C, Pinheiro V, Garcia-Santos S, et al. Insights from the yield, protein production, and detailed alkaloid composition of white (Lupinus albus), narrow-leafed (Lupinus angustifolius), and yellow (Lupinus luteus) lupin cultivars in the Mediterranean region. Front Plant Sci. 2023;14:1231777.
18. Ruiz-López MA, Barrientos-Ramírez L, García-López PM, Valdés-Miramontes EH, Zamora-Natera JF, Rodríguez-Macias R, et al. Nutritional and bioactive compounds in Mexican lupin beans species: A mini-review. Nutrients. 2019;11:1785.
19. Singh M, Jyoti, Kumar N, Singh H. Plant functional traits assisted crop adaptation to abiotic and biotic stress. Plant functional traits for improving productivity. Singapore: Springer Nature Singapore; 2024. pp. 239–55.
20. Frick KM, Kamphuis LG, Siddique KHM, Singh KB, Foley RC. Quinolizidine alkaloid biosynthesis in lupins and prospects for grain quality improvement. Front Plant Sci. 2017;8:87.
21. Osorio CE, Till BJ. A Bitter-Sweet story: unraveling the genes involved in Quinolizidine alkaloid synthesis in Lupinus albus. Front Plant Sci. 2022;12:795091.
22. Czepiel K, Krajewski P, Wilczura P, Bielecka P, Święcicki W, Kroc M. Expression profiles of alkaloid-related genes across the organs of narrow-leafed lupin (Lupinus angustifolius L.) and in response to anthracnose infection. Int J Mol Sci. 2021;22:2676.
23. Frick KM, Foley RC, Kamphuis LG, Siddique KHM, Garg G, Singh KB. Characterization of the genetic factors affecting Quinolizidine alkaloid biosynthesis and its response to abiotic stress in narrow-leafed lupin (Lupinus angustifolius L). Plant Cell Environ. 2018;41:2155–68.
24. Rodés-Bachs C, Van der Fels-Klerx HJ. Impact of environmental factors on the presence of Quinolizidine alkaloids in lupins: a review. Food Addit Contam Part Chem Anal Control Expo Risk Assess. 2023;40:757–69.
25. Mancinotti D, Frick KM, Geu-Flores F. Biosynthesis of Quinolizidine alkaloids in lupins: mechanistic considerations and prospects for pathway Elucidation. Nat Prod Rep. 2022;39:1423–37.
26. Alkemade JA, Nazzicari N, Messmer MM, Annicchiarico P, Ferrari B, Voegele RT, et al. Genome-wide association study reveals white lupin candidate gene involved in anthracnose resistance. Züchter Genet Breed Res. 2022;135:1011–24.
27. Godoy GA, Levicoy D, Martinez-Hernandez JE, Rupayan A, Hernandez S, Carrasco M et al. Defence mechanisms in*Lupinus luteus*against*Colletotrichum lupini*, involving TIR-NBS-LRR protein, hypersensitive response and phenylpropanoid pathways. BioRxiv. 2024.
28. Fischer K, Dieterich R, Nelson MN, Kamphuis LG, Singh KB, Rotter B, et al. Characterization and mapping of LanrBo: a locus conferring anthracnose resistance in narrow-leafed lupin (Lupinus angustifolius L). Züchter Genet Breed Res. 2015;128:2121–30.
29. Zhou G, Yang H, Renshaw D, Zou M, Thomas G, Li C. Identification of Gray leaf spot disease candidate gene in narrow-leafed lupin (Lupinus angustifolius L). Front Genet. 2021;12:695791.
30. Mithöfer A, Boland W. Plant defence against herbivores: chemical aspects. Annu Rev Plant Biol. 2012;63:431–50.
31. Li P, Day B. Battlefield cytoskeleton: turning the tide on plant immunity. Mol Plant Microbe Interact. 2019;32:25–34.
32. Ngou BPM, Ding P, Jones JDG. Thirty years of resistance: Zig-zag through the plant immune system. Plant Cell. 2022;34:1447–78.
33. Wolko B, Clements JC, Naganowska B, Nelson MN, Yang H 'an, editors. Lupinus. Wild Crop Relatives: Genomic and Breeding Resources. 2011;:153–206.
34. Aïnouche A, Bayer RJ, Misset M-T. Molecular phylogeny, diversification and character evolution in Lupinus (Fabaceae) with special attention to mediterranean and African lupines. Osterr Bot Z. 2004;246.
35. Kazimierski T, Kazimierska M. Morphological and cytological differences within the species Lupinus luteus L. Acta Soc Bot Pol Pol Tow Bot. 2015;44:265–75.
36. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de Novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18:170–5.
37. Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome Polishing tool for long-read assembly. Bioinformatics. 2020;36:2253–5.
38. Ou S, Chen J, Jiang N. Assessing genome assembly quality using the LTR assembly index (LAI). Nucleic Acids Res. 2018;46:e126.
39. Xu L, Wang Y, Dong J, Zhang W, Tang M, Zhang W, et al. A chromosome-level genome assembly of radish (Raphanus sativus L.) reveals insights into genome adaptation and differential bolting regulation. Plant Biotechnol J. 2023;21:990–1004.

40. Bayer PE, Golicz AA, Tirnaz S, Chan C-KK, Edwards D, Batley J. Variation in abundance of predicted resistance genes in the Brassica oleracea pangenome. Plant Biotechnol J. 2019;17:789–800.
41. Jiu S, Chen B, Dong X, Lv Z, Wang Y, Yin C, et al. Chromosome-scale genome assembly of Prunus pusilliflora provides novel insights into genome evolution, disease resistance, and dormancy release in cerasus L. Hortic Res. 2023;10:uhad062.
42. Rychel S, Książkiewicz M. Development of gene-based molecular markers tagging low alkaloid pauper locus in white lupin (Lupinus albus L). J Appl Genet. 2019;60:269–81.
43. Nogia P, Pati PK. Plant secondary metabolite transporters: diversity, functionality, and their modulation. Front Plant Sci. 2021;12:758202.
44. Hildreth SB, Gehman EA, Yang H, Lu R-H, Ritesh KC, Harich KC, et al. Tobacco nicotine uptake permease (NUP1) affects alkaloid metabolism. Proc Natl Acad Sci U S A. 2011;108:18179–84.
45. Plewiński P, Książkiewicz M, Rychel-Bielska S, Rudy E, Wolko B. Candidate domestication-related genes revealed by expression quantitative trait loci mapping of narrow-leafed lupin (Lupinus angustifolius L). Int J Mol Sci. 2019;20:5670.
46. Lo B, Kasapis S, Farahnaky A. Lupin protein: isolation and techno-functional properties, a review. Food Hydrocoll. 2021;112:106318.
47. Garg G, Kamphuis LG, Bayer PE, Kaur P, Dudchenko O, Taylor CM, et al. A pan-genome and chromosome-length reference genome of narrow-leafed lupin (Lupinus angustifolius) reveals genomic diversity and insights into key industry and biological traits. Plant J. 2022;111:1252–66.
48. Wang P, Zhou G, Jian J, Yang H, Renshaw D, Aubert MK, et al. Whole-genome assembly and resequencing reveal genomic imprint and key genes of rapid domestication in narrow-leafed lupin. Plant J. 2021;105:1192–210.
49. Hufnagel B, Marques A, Soriano A, Marquès L, Divol F, Doumas P, et al. High-quality genome sequence of white lupin provides insight into soil exploration and seed quality. Nat Commun. 2020;11:492.
50. Zhang H, Mascher M, Abbo S, Jayakodi M. Advancing grain legumes domestication and evolution studies with genomics. Plant Cell Physiol. 2022;63:1540–53.
51. Garg V, Dudchenko O, Wang J, Khan AW, Gupta S, Kaur P, et al. Chromosome-length genome assemblies of six legume species provide insights into genome organization, evolution, and agronomic traits for crop improvement. J Adv Res. 2022;42:315–29.
52. Drummond CS, Eastwood RJ, Miotto STS, Hughes CE. Multiple continental radiations and correlates of diversification in Lupinus (Leguminosae): testing for key innovation with incomplete taxon sampling. Syst Biol. 2012;61:443–60.
53. Książkiewicz M, Nazzicari N, Yang H, 'an, Nelson MN, Renshaw D, Rychel S, et al. A high-density consensus linkage map of white lupin highlights synteny with narrow-leafed lupin and provides markers tagging key agronomic traits. Sci Rep. 2017;7:15335.
54. Susek K, Bielski W, Czyż KB, Hasterok R, Jackson SA, Wolko B, et al. Impact of chromosomal rearrangements on the interpretation of lupin karyotype evolution. Genes (Basel). 2019;10:259.
55. Bielski W, Książkiewicz M, Šimoníková D, Hřibová E, Susek K, Naganowska B. The puzzling fate of a lupin chromosome revealed by reciprocal oligo-FISH and BAC-FISH mapping. Genes (Basel). 2020;11:1489.
56. Foley RC, Jimenez-Lopez JC, Kamphuis LG, Hane JK, Melser S, Singh KB. Analysis of conglutin seed storage proteins across lupin species using transcriptomic, protein and comparative genomic approaches. BMC Plant Biol. 2015;15:106.
57. Tahmasian A, Juhász A, Broadbent JA, Nye-Wood MG, Le TT, Colgrave ML. Evaluation of the major seed storage proteins, the conglutins, across genetically diverse narrow-leafed lupin varieties. Front Nutr. 2022;9:842168.
58. Vargas-Guerrero B, García-López PM, Martínez-Ayala AL, Domínguez-Rosales JA, Gurrola-Díaz CM. Administration of Lupinus albus gamma conglutin (Cγ) to n5 STZ rats augmented Ins-1 gene expression and pancreatic insulin content. Plant Foods Hum Nutr. 2014;69:241–7.
59. Ogura T, Hernández A, Aizawa T, Ogihara J, Sunairi M, Alcaino J, et al. Identification of a low digestibility δ-Conglutin in yellow lupin (Lupinus luteus L.) seed meal for Atlantic salmon (Salmo Salar L.) by coupling 2D-PAGE and mass spectrometry. PLoS ONE. 2013;8:e80369.
60. Dooper MMBW, Plassen C, Holden L, Lindvik H, Faeste CK. Immunoglobulin E cross-reactivity between lupine conglutins and peanut allergens in serum of lupine-allergic individuals. J Investig Allergol Clin Immunol. 2009;19:283–91.
61. Holden L, Sletten GBG, Lindvik H, Faeste CK, Dooper MMBW. Characterization of IgE binding to Lupin, peanut and almond with Sera from lupin-allergic patients. Int Arch Allergy Immunol. 2008;146:267–76.
62. Klajn N, Kapczyńska K, Pasikowski P, Glazińska P, Kugiel H, Kęsy J et al. Regulatory effects of ABA and GA on the expression of conglutin genes and LAFL network genes in yellow lupine (Lupinus luteus L.) seeds. Int J Mol Sci. 2023;24.
63. Toffolatti SL, Maddalena G, Passera A, Casati P, Bianco PA, Quaglino F. Role of terpenes in plant defence to biotic stress. Biocontrol agents and secondary metabolites. Elsevier; 2021. pp. 401–17.
64. Ninkuu V, Zhang L, Yan J, Fu Z, Yang T, Zeng H. Biochemistry of terpenes and recent advances in plant protection. Int J Mol Sci. 2021;22:5710.
65. Makhlouf L, El Fakhouri K, Kemal SA, Maafa I, Meftah Kadmiri I, El Bouhssini M. Potential of volatile organic compounds in the management of insect pests and diseases of food legumes: a comprehensive review. Front Plant Sci. 2024;15:1430863.
66. Pichersky E, Raguso RA. Why do plants produce so many terpenoid compounds? New Phytol. 2018;220:692–702.
67. Adamczyk S, Latvala S, Poimala A, Adamczyk B, Hytönen T, Pennanen T. Diterpenes and triterpenes show potential as biocides against pathogenic fungi and oomycetes: a screening study. Biotechnol Lett. 2023;45:1555–63.
68. Wiart C, Kathirvalu G, Raju CS, Nissapatorn V, Rahmatullah M, Paul AK et al. Antibacterial and antifungal terpenes from the medicinal angiosperms of Asia and the Pacific: haystacks and gold needles. Molecules. 2023;28.
69. Zheng F, Wu H, Zhang R, Li S, He W, Wong F-L, et al. Molecular phylogeny and dynamic evolution of disease resistance genes in the legume family. BMC Genomics. 2016;17:402.
70. Książkiewicz M, Rychel-Bielska S, Plewiński P, Bielski W, Nuc M, Kozak B, et al. A successful defence of the narrow-leafed lupin against anthracnose involves quick and orchestrated reprogramming of oxidation-reduction, photosynthesis and pathogenesis-related genes. Sci Rep. 2022;12:8164.
71. Schryvers S, Arinzechukwu C, Miserez B, Eeckhout M, Jacxsens L. The fate of Quinolizidine alkaloids during the processing of lupins (Lupinus spp.) for human consumption. Food Chem. 2023;429:136847.
72. Kroc M, Koczyk G, Kamel KA, Czepiel K, Fedorowicz-Strońska O, Krajewski P, et al. Transcriptome-derived investigation of biosynthesis of Quinolizidine alkaloids in narrow-leafed lupin (Lupinus angustifolius L.) highlights candidate genes linked to Iucundus locus. Sci Rep. 2019;9:2231.
73. Wink M. N-Methylation of Quinolizidine alkaloids: an S-adenosyl-L-methionine: cytisine N-methyltransferase from Laburnum anagyroides plants and cell cultures of L. alpinum and Cytisus canariensis. Planta. 1984;161:339–44.
74. Bunsupa S, Okada T, Saito K, Yamazaki M. An acyltransferase-like gene obtained by differential gene expression profiles of Quinolizidine alkaloid-producing and nonproducing cultivars of Lupinus angustifolius. Plant Biotechnol (Tsukuba). 2011;28:89–94.
75. Wink M, Hartmann T. Localization of the enzymes of Quinolizidine alkaloid biosynthesis in leaf chloroplasts of Lupinus polyphyllus. Plant Physiol. 1982;70:74–7.
76. Yang T, Nagy I, Mancinotti D, Otterbach SL, Andersen TB, Motawia MS, et al. Transcript profiling of a bitter variety of narrow-leafed lupin to discover alkaloid biosynthetic genes. J Exp Bot. 2017;68:5527–37.
77. Otterbach SL, Yang T, Kato L, Janfelt C, Geu-Flores F. Quinolizidine alkaloids are transported to seeds of bitter narrow-leafed lupin. J Exp Bot. 2019;70:5799–808.
78. Valente IM, Monteiro A, Sousa C, Miranda C, Maia MRG, Castro C, et al. Agronomic, nutritional traits, and alkaloids of Lupinus Albus, Lupinus angustifolius and Lupinus luteus genotypes: effect of sowing dates and locations. ACS Agric Sci Technol. 2024;4:450–62.
79. Matsuura HN, Fett-Neto AG. Plant alkaloids: main features, toxicity, and mechanisms of action. Plant toxins. Dordrecht: Springer Netherlands; 2015. pp. 1–15.
80. Lashley A, Miller R, Provenzano S, Jarecki S-A, Erba P, Salim V. Functional diversification and structural origins of plant natural product methyltransferases. Molecules. 2022;28:43.
81. Ishikawa E, Kanai S, Shinozawa A, Hyakutake M, Sue M. Hordeum vulgare CYP76M57 catalyzes C2 shortening of Tryptophan side chain by C-N bond rearrangement in Gramine biosynthesis. Plant J. 2024;118:892–904.
82. Levac D, Cázares P, De Yu F. A picrinine N-methyltransferase belongs to a new family of γ-tocopherol-like methyltransferases found in medicinal plants that make biologically active monoterpenoid Indole alkaloids. Plant Physiol. 2016;170:1935–44.
83. Pathak S, Lakhwani D, Gupta P, Mishra BK, Shukla S, Asif MH, et al. Comparative transcriptome analysis using high Papaverine mutant of Papaver somniferum reveals pathway and uncharacterized steps of Papaverine biosynthesis. PLoS ONE. 2013;8:e65622.

84. Zhao D, Yao Z, Zhang J, Zhang R, Mou Z, Zhang X, et al. Melatonin synthesis genes N-acetylserotonin methyltransferases evolved into caffeic acid O-methyltransferases and both assisted in plant terrestrialization. J Pineal Res. 2021;71:e12737.
85. Antosiak A, Milowska K, Maczynska K, Rozalska S, Gabryelak T. Cytotoxic activity of genistein-8-C-glucoside form Lupinus luteus L. and genistein against human SK-OV-3 ovarian carcinoma cell line. Med Chem Res. 2017;26:64–73.
86. Adhikari KN, Edwards OR, Wang S, Ridsdill-Smith TJ, Buirchell B. The role of alkaloids in conferring aphid resistance in yellow lupin (Lupinus luteus L). Crop Pasture Sci. 2012;63:444.
87. Génard T, Etienne P, Laîné P, Yvin J-C, Diquélou S. Nitrogen transfer from Lupinus albus L., Trifolium incarnatum L. and Vicia sativa L. contribute differently to rapeseed (Brassica Napus L.) nitrogen nutrition. Heliyon. 2016;2:e00150.
88. Macaulay M, Ramsay L, Åhman I. Quantitative trait locus for resistance to the aphid rhopalosiphum Padi L. in barley (Hordeum vulgare L.) is not linked with a genomic region for Gramine concentration. Arthropod Plant Interact. 2020;14:57–65.
89. Yang M, Zhu L, Li L, Li J, Xu L, Feng J, et al. Digital gene expression analysis provides insight into the transcript profile of the genes involved in aporphine alkaloid biosynthesis in lotus (Nelumbo nucifera). Front Plant Sci. 2017;8:80.
90. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, et al. SOAPnuke: a mapreduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience. 2018;7:1–6.
91. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27:764–70.
92. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 2017;33:2202–4.
93. Cheng H, Jarvis ED, Fedrigo O, Koepfli K-P, Urban L, Gemmell NJ, et al. Haplotype-resolved assembly of diploid genomes without parental data. Nat Biotechnol. 2022;40:1332–5.
94. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution hi-C experiments. Cell Syst. 2016;3:95–8.
95. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De Novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science. 2017;356:92–5.
96. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26:589–95.
97. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics. 2018;34:i142–50.
98. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021;38:4647–54.
99. Tempel S. Using and Understanding repeatmasker. Methods Mol Biol. 2012;859:29–51.
100. Bao W, Kojima KK, Kohany O. Repbase update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6:11.
101. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117:9451–7.
102. Arias-Carrasco R, Vásquez-Morán Y, Nakaya HI, Maracaja-Coutinho V. StructRNAfinder: an automated pipeline and web server for RNA families prediction. BMC Bioinformatics. 2018;19.
103. Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2 0 Algorithms Mol Biol. 2011;6:26.
104. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29:2933–5.
105. Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, et al. Rfam 14: expanded coverage of metagenomic, viral and MicroRNA families. Nucleic Acids Res. 2021;49:D192–200.
106. Chan PP, Lowe TM. TRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol. 2019;1962:1–14.
107. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de Novo gene finding. Bioinformatics. 2008;24:637–44.
108. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 2005;33:6494–506.
109. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90.
110. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15.
111. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28:511–5.
112. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–5.
113. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J et al. Automated eukaryotic gene structure annotation using evidencemodeler and the program to assemble spliced alignments. Genome Biol. 2008;9.
114. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, et al. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 2003;31:5654–66.
115. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017;35:1026–8.
116. Keilwagen J, Hartung F, Grau J, GeMoMa. Homology-based gene prediction utilizing intron position conservation and RNA-seq data. Methods Mol Biol. 2019;1962:161–77.
117. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.
118. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. EggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. BioRxiv. 2021.
119. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47:D309–14.
120. Ghahramanlu A, Rezaei M, Heidari P, Balandari A. Species survey of Iranian barberry genotypes using ITS2 sequences and basic local alignment search tools. Erwerbs-Obstbau. 2023;65:2491–9.
121. UniProt Consortium. UniProt: the universal protein knowledgebase in 2023. Nucleic Acids Res. 2023;51:D523–31.
122. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 2020;36:2251–2.
123. Li P, Quan X, Jia G, Xiao J, Cloutier S, You FM. RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. BMC Genomics. 2016;17:852.
124. Osuna-Cruz CM, Paytuvi-Gallart A, Di Donato A, Sundesha V, Andolfo G, Aiese Cigliano R, et al. PRGdb 3.0: a comprehensive platform for prediction and analysis of plant disease resistance genes. Nucleic Acids Res. 2018;46:D1197–201.
125. Kourelis J, Sakai T, Adachi H, Kamoun S. RefPlantNLR is a comprehensive collection of experimentally validated plant disease resistance proteins from the NLR family. PLoS Biol. 2021;19:e3001124.
126. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: the protein families database in 2021. Nucleic Acids Res. 2021;49:D412–9.
127. Yang J-S, Qian Z-H, Shi T, Li Z-Z, Chen J-M. Chromosome-level genome assembly of the aquatic plant Nymphoides indica reveals transposable element bursts and NBS-LRR gene family expansion shedding light on its invasiveness. DNA Res. 2022;29.
128. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021;18:366–8.
129. Kanehisa M. Toward Understanding the origin and evolution of cellular organisms. Protein Sci. 2019;28:1947–51.
130. Kanehisa M, Furumichi M, Sato Y, Matsuura Y, Ishiguro-Watanabe M. KEGG: biological systems database as a model of the real world. Nucleic Acids Res. 2025;53:D672–7.
131. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.
132. Saier MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, Iddamsetty V, et al. The transporter classification database (TCDB): 2021 update. Nucleic Acids Res. 2021;49:D461–7.
133. Hou R, Wang L, Wu Y-J. Predicting ATP-binding cassette transporters using the random forest method. Front Genet. 2020;11:156.

134. Hallgren J, Tsirigos KD, Pedersen MD, Almagro Armenteros JJ, Marcatili P, Nielsen H et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. BioRxiv. 2022.
135. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238.
136. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
137. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59:307–21.
138. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 1997;13:555–6.
139. Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M et al. TimeTree 5: an expanded resource for species divergence times. Mol Biol Evol. 2022;39.
140. Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics. 2021;36:5516–8.
141. Maere S, Heymans K, Kuiper M. BiNGO: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics. 2005;21:3448–9.
142. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.
143. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5:R12.
144. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and collinearity in plant genomes. Science. 2008;320:486–8.
145. Tang H, Krishnakumar V, Zeng X, Xu Z, Taranto A, Lomas JS, et al. JCVI: A versatile toolkit for comparative genomics analysis. Imeta. 2024;3:e211.
146. Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.
147. Robinson MD, McCarthy DJ, Smyth GK. EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
148. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The Sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28:882–3.

Publisher's note