Chromosome-level genome assembly of yellow lupin (*Lupinus luteus*) provides novel insights into genome evolution, crop adaptation and seed protein in the three most cultivated lupins

J. Eduardo Martinez-Hernandez
CGNA (Agriaquaculture Nutritional Genomic Center), Las Heras 350, Temuco, 4781158, Chile.
https://orcid.org/0000-0003-0739-0999

Haroldo Salvo-Garrido
haroldo.salvo@cgna.cl

CGNA (Agriaquaculture Nutritional Genomic Center), Las Heras 350, Temuco, 4781158, Chile.

Daniela Levicoy
CGNA (Agriaquaculture Nutritional Genomic Center), Las Heras 350, Temuco, 4781158, Chile.

Peter D. S. Caligari
CGNA (Agriaquaculture Nutritional Genomic Center), Las Heras 350, Temuco, 4781158, Chile.

Annally Rupayán
CGNA (Agriaquaculture Nutritional Genomic Center), Las Heras 350, Temuco, 4781158, Chile.

Tomas Moyano
Departamento de Genética Molecular y Microbiología, Pontificia Universidad Católica de Chile, Santiago, 8331150, Chile

Makarena Carrasco
CGNA (Agriaquaculture Nutritional Genomic Center), Las Heras 350, Temuco, 4781158, Chile.

Sebastián Hernandez
CGNA (Agriaquaculture Nutritional Genomic Center), Las Heras 350, Temuco, 4781158, Chile.

Grace Armijo-Godoy
CGNA (Agriaquaculture Nutritional Genomic Center), Las Heras 350, Temuco, 4781158, Chile.

Fernando Westermeyer
CGNA (Agriaquaculture Nutritional Genomic Center), Las Heras 350, Temuco, 4781158, Chile.

Giovanni Larama
Biocontrol Research Laboratory, Universidad de La Frontera, Temuco 4811230, Chile

Research Article

Keywords: Lupinus luteus, genome assembly, Alkaloid biosynthesis, Resistance genes, Chromosomal rearrangements

Posted Date: March 27th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4171664/v1

Additional Declarations: The authors declare no competing interests.

Abstract

Lupinus luteus is a grain legume crop of agricultural importance due to its high seed protein content. In this study, the first chromosome-scale genome assembly of *L. luteus* (962.97 Mb) is presented, integrating data from Illumina, PacBio, and Hi-C platforms. The assembly exhibits exceptional completeness (98.9% BUSCO score) and a high repetition rate (76.15%). Genomic annotation identifies 36,884 protein-coding genes, including 2,492 transcription factors and 23 microRNA families.

Synteny analysis with lupin species reveals important chromosomal rearrangements, indicating complex interactions between conserved regions and structural variations. Our analyses suggest that chromosome 8 may have originated from a translocation event involving two chromosomes during the speciation of *L. luteus*. Orthologous group characterization between *L. luteus* and related species indicates an enrichment in gene families associated with biotic and abiotic stress responses, secondary metabolism, and nutrient reservoir activity. Moreover, 911 resistance (R) genes are identified, highlighting their importance in pathogen defence. Exploration of alkaloid biosynthesis and regulation reveals 16 genes associated with quinolizidine alkaloids (QAs) with expression analysis revealing tissue-specific expression patterns for key enzymes in QA biosynthesis. Furthermore, secondary metabolite transporters are explored, including a *Lupinus angustifolius* PUP1 ortholog, providing insights into QA translocation mechanisms. This comprehensive genome analysis provides valuable resources for further understanding the genetic basis of important traits in *Lupinus luteus*, facilitating advancements in crop adaptation, improvement, and sustainability.

Introduction

Lupins are valuable in agriculture because of their substantial seed protein content (Lucas et al. 2015) but generally underexploited. They are found in various regions of the world and are grouped into Old World lupin (Mediterranean) and New World lupin (American) (Ainouche and Bayer 1999) species. The four principal species grown, with the highest protein content, are *Lupinus albus*, *Lupinus luteus*, *Lupinus angustifolius* and *Lupinus mutabilis* (Musco et al. 2017; Carvajal-Larenas 2019). They have different nutritional profiles but the highest average protein, making them suitable for producing high-quality food and feed in both human and animal diets (Ishaq et al. 2022). In addition, they contribute to the sustainability of cropping systems owing to their low fertilizer requirements and positive impact on soil fertility, as well as their ability to fix atmospheric nitrogen in symbiosis with beneficial bacteria and efficiently take up phosphorus from soils (Hane et al. 2017).

L. luteus has high dehulled seed protein content, 60% dry matter (DM). (Burgos-Díaz et al. 2019; Gresta et al. 2010) and twice the cysteine and methionine content of most lupins (Parra-González et al. 2012; Martínez-Villaluenga, Frías, and Vidal-Valverde 2006). 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 (Glazinska et al. 2020). *L. luteus* has been studied as an alternative protein source for countries that import soybeans (Abraham et al. 2019; Pexas, Doherty, and Kyriazakis 2023; Duranti et al.

2008). However, despite their importance, there are few studies related to genomics and molecular tools in this species (Lichtin et al. 2020; Iqbal et al. 2019; Parra-González et al. 2012).

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 (Lambers, Clements, and Nelson 2013). The intricate balance between biotic and abiotic stress factors significantly shapes the adaptive responses of plants, influencing their genetic landscape and biochemical composition (Musco et al. 2017; Frick et al. 2017; Ruiz-López et al. 2019). 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 (Musco et al. 2017; Frick et al. 2017; Gresta et al. 2010). However, the regulation of alkaloids becomes critical due to challenges associated with bitterness and potential toxicity, highlighting a nuanced aspect of lupin biology (Ruiz and Sotelo 2001). In tandem to the dynamics of alkaloid synthesis, the presence of resistance (R) proteins assumes a pivotal role in the plant's ability to withstand biotic stresses (Mithöfer and Boland 2012). Serving as sentinels of the cellular defence system, R proteins play a crucial role in recognizing pathogenic invaders and triggering cascades of molecular responses that fortify the plant's resistance against diseases (Zhu et al. 2010; Pai Li and Day 2019). This comprehensive exploration of the intricate crosstalk between environmental stress, alkaloid production, and the deployment of R proteins is fundamental for unravelling the genetic mechanisms that underpin the adaptability and defence strategies of *Lupinus luteus*.

In the current work, a chromosome-level genome assembly for yellow lupin is presented, offering insights into its structural organization, 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 quality validation

Lupinus luteus C195 (cultivar Alu*Prot*-CGNA; 2n = 2x = 52) was chosen for genome sequencing and assembly (Fig. 1A). To 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 C195 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 (Cheng et al. 2021) 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 S2). The assembly was polished for error correction using Illumina short reads from the same cultivar with Nextpolish (Hu et al. 2020) 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 S3) 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, Fig. 1D). 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 (Ou, Chen, and Jiang 2018) (Supplementary Figure S2). 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 lab. The ratio of mapped reads to the genome varies between 76–94%, depending on the tissue of origin (Supplementary Table S6).

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

Assembly features	
Estimated genome size	1,024.49 Mb
Total assembly size	962.97 Mb
GC content (%)	35.4
Repeat content (%)	76.15
Contig number	1,053
Contig N50	15.38 Mb
Longest contig	26.27 Mb
Scaffold number	111
Scaffold N50	37.36 Mb
Longest scaffold	46.32 Mb
Complete BUSCO (%):	98.9
Embryophyta	96.2
Fabales	
Annotation features	
No. of protein-coding genes	36,895
Average gene length (bp)	5,370
Average CDS length (bp)	1,191
Complete BUSCO (%):	93.9
Embryophyta	91.1
Fabales	
Protein-coding genes with annotation	35,940

Annotation of L. luteus genome

Repetitive elements were identified and masked in the *L. luteus* 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% of all LTR elements. The genome of *L. luteus* 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.

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 *L. luteus* genome (Supplementary Table S8, Supplementary Table S9, and Supplementary Table S10). Among these genes, 2,492 Transcription Factors belonging to 58 families were identified (Supplementary Table S11). BUSCO annotation of predicted genes showed 93.9% and 91.7% completeness using Embryophyta and Fabales lineages (Table 1, Fig. 1D).

Non-coding RNAs (ncRNAs) were predicted in the genome as well. 23 microRNA (miRNA) families with putative functions based on homology with other miRNA families in other plants (Supplementary Table S13). 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 in L.luteus

The genome of *L. luteus* was compared with other phylogenetically related Fabales (*G. max*, *L. albus*, *L. angustifolius*, and *M. truncatula*) (Fig. 2A). In addition, the *A. thaliana* genome was used as an outgroup in the phylogenetic analysis (Fig. 2B). Among the Fabales species, 26,956 orthologous groups were identified, of which 13,076 are shared in common (Fig. 2A) and 7,561 (19.5%) yellow lupin genes in 483 species-specific orthologous groups (Supplementary Table 14). 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 S3 and Supplementary Table S15). 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. 2B). 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 as response to stress, response to flooding, response to herbicide, defence response to bacterium and defence response to insect, response to jasmonic acid stimulus, and response to salicylic acid stimulus (Fig. 2C, Supplementary Table S16). In addition, 30 members of MYB-related and 10 Nin-like transcription factor families were significantly expanded in the *L. luteus* genome (Supplementary Table S17). 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 S17). On the other hand, contracted gene families were involved in e.g. regulation of catalytic activity, protein modification, and cold acclimation (Supplementary Table S18, Supplementary Fig. 4).

Orthologous groups across the *Lupinus* species (*L. albus*, *L. angustifolius*, and *L. luteus*) were analysed. Remarkably, 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. 3A). Notably, 20,198 gene families were found to be shared among all three *Lupinus* species, underscoring a core genomic repertoire conserved across the most cultivated lupins (Fig. 3A). Intriguingly, our analysis unveiled 442 unique gene families 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, underscoring the adaptive genomic repertoire of *Lupinus luteus* (Supplementary Figure S5, Supplementary Table S19).

Furthermore, within the shared gene families, we detected significant expansions and contractions of 362 and 306 families, respectively, in *L. luteus* (Fig. 3B). 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. 3C and Supplementary Table S20), illuminating its evolutionary adaptations to environmental challenges. Notably, we elucidated the genetic basis underlying the superior protein content of *L. luteus* seeds, a trait of considerable agronomic importance. Our comparative genomic analysis identified 30 genes encoding conglutin proteins in the genome of *L. luteus* (Supplementary Table S21), surpassing the counts of 19 and 20 observed in *L. albus* and *L. angustifolius*, respectively. To note was the significant expansion of the conglutin delta 2 family in *L. luteus*, shedding light on the genetic underpinnings of its enhanced nutritional profile.

Syntenic analysis between L. luteus and other lupin species

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-B and Supplementary Figure S6). However, more significant chromosomal rearrangements were evident between *L. luteus* and *L. angustifolius* compared to *L. luteus* and *L. albus* (Fig. 4A-B). Moreover, multiple chromosome inversions were identified between *L. luteus* and the chromosomes of its two related genomes. Synteny analysis indicated that *L. luteus* chromosome 8 partially aligns with *L. albus* chromosomes 1 and 15 (see Fig. 4D and Supplementary Figure S7A). Similarly, it showed partial alignment of *L. luteus* chromosome 8 with *L. angustifolius* chromosomes 4 and 9 (Fig. 4C and Supplementary Figure S7B). To note, the inversion appears in the homologous region between chromosome 8 of *L. luteus* and the two chromosomes of *L. albus* (Fig. 4D-F and Supplementary Figure S7A). Further examination mapped 328 synteny blocks within *L. luteus* chromosome 8 and *L. albus* chromosomes 1 and 15, sharing an identity of 87–96% (Fig. 4F, Supplementary Figure S7A and

Supplementary Table S22). Similarly, 500 blocks were identified between chromosome 8 of *L. luteus* and chromosomes 4 and 9 of *L. angustifolius*, with sequence identity ranging from 87–97% (Fig. 4E, Supplementary Figure S7B and Supplementary Table S23). Here 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 24), thus demonstrating high genomic conservation among the genomes of *Lupinus* genus.

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 S8, Supplementary Table S25). 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 S8B). 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 in L. luteus

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 of *L. luteus.* 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) (Pingchuan Li et al. 2016). 80 NBS-type, 580 RLK-type, 61 RLP, and 190 TM-CC-type proteins in *L. luteus* (Supplementary Figure S9A) were identified. In addition, all R gene types were unevenly located in all the twenty-six chromosomes (Supplementary Figure S9B). Then, NBS-type R genes were analysed by their key role in host resistance to diseases (Xu et al. 2023; Bayer et al. 2019; Jiu et al. 2023). From 80 NBS-type genes encoded in the *L. luteus* genome, were identified as: 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) (Supplementary Table S26), distributed over all chromosomes, except for chromosomes 02, 15, 18 and 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 (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 S27).

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 cluster were more phylogenetically related (Supplementary Figure S10).

Alkaloid biosynthesis-related genes and secondary metabolite transporters

In lupins, the content of alkaloids is an important factor that affects their nutritional impact (Rychel and Książkiewicz 2019). 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 (Frick et al. 2017). In addition to QAs, in *L. luteus* the indole alkaloid, gramine, is accumulated in high amounts in seeds. 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 S28 and Supplementary Figure S11). While, 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 (Nogia and Pati 2021). 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 S29); they are unevenly distributed along the 26 chromosomes (Supplementary Figure S12). 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* (Hildreth et al. 2011)d *angustifolius* (Plewiński et al. 2019). An ortholog gene of *Lan*PUP1 was identified in the *L. luteus* genome. The expression patterns of *Llu*PUP1 (Llu07983) present ones similar to *Llu*LDC and *Llu*CAO. However, in a correlation expression analysis, a significant correlation was not found between these genes (Supplementary Figure S13 and Supplementary Table S30).

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

Alkaloid	Gene	Name	Chr	Description
Quinolizidine alkaloids	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
Gramine	Llu28618	NMT	YL-04	3-aminomethylindole N-methyltransferase

Discussion

The *Lupinus* genus includes species economically important for protein production, and valuable functional food crops such as white lupin (*L. albus*), narrow-leaf lupin (*L. angustifolius*) and yellow lupin (*L. luteus*) (Lo, Kasapis, and Farahnaky 2021). Despite the economic and nutritional potential of lupins, genomic information is relatively scarce, and genome sequences are available only for two species *L. angustifolius* (Garg et al. 2022; Hane et al. 2017; Wang et al. 2021)d *albus* (Hufnagel et al. 2020). Here, the first chromosome-scale genome assembly of *Lupinus luteus* is presented, as obtained by combining data from the Illumina, PacBio, and Hi-C platforms. This genome assembly shows a remarkable level of completeness in protein-coding genes and repetitive sequences, as well as good 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) (Garg et al. 2022)d *albus* (451 Mb) (Hufnagel et al. 2020). 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%). Further, *L. luteus* has a higher scaffold N50 (33.87 Mb) than the latest version of *L. angustifolius* assembly (scaffold N50 = 30.7 Mb) (Garg et al. 2022)d *albus* (scaffold N50 = 17.4 Mb) (Hufnagel et al. 2020). 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. This implies that *L. luteus* assembly will serve as a high-quality reference genome according to the LAI classification score proposed by Ou *et al* (Ou, Chen, and Jiang 2018).

The analysis of the genome of *L. luteus*, in comparison to other Fabales species, revealed a recent divergence within the lupin branch. This finding is consistent with previous research that indicated a divergence between *L. albus, L. angustifolius*, and *L. luteus* approximately 8 million years ago (Drummond et al. 2012). Further study demonstrated that *L. angustifolius* is more closely related to *L. luteus*. Our research group recently employed comparative mapping of genetic linkage groups to establish that the genomes of these species retain extensive syntenic blocks linked to flowering time and resistance to anthracnose (Lichtin et al. 2020). However, synteny and collinearity analyses also revealed a higher incidence of chromosomal rearrangements between the genomes of *L. angustifolius* and *L. luteus*. This suggests a complex interplay between conserved genomic regions and structural variation in lupin genomes, potentially influencing key agronomic traits.

Other research employing techniques such as comparative mapping of linkage markers and BAC-FISH (Bacterial Artificial Chromosomes Fluorescence In Situ Hybridization) have consistently revealed remarkable conservation within the genomes of Old World lupins (Książkiewicz et al. 2017; Lichtin et al. 2020; Susek et al. 2019; Bielski et al. 2020). Our comprehensive synteny analysis and orthologous group characterization robustly corroborate these established findings. In a previous work Książkiewicz et al. found that linkage groups Lalb01 and Lalb15 of *L. albus* exhibited partial synteny with the NLL-04 and NLL-09 linkage groups of *L. angustifolius* (Książkiewicz et al. 2017). Notably, we uncovered the same partial synteny between chromosome 8 of *L. luteus* and corresponding sequence of chromosomes 1 and 15 of *L. albus* and 4 and 9 of *L. angustifolius*. This evidence suggests a potential translocation event that could have led to the formation of chromosome 8 in *L. luteus*. These findings underscore the dynamic nature of lupine genomes and provide valuable insights into their evolutionary history.

Conglutins, categorized into α-, β-, γ-, and δ conglutin subfamilies, serve as the primary storage proteins in lupin seeds (Foley et al. 2015; Tahmasian et al. 2022). Recent work has explored the potential therapeutic effects of these proteins on various metabolic disorders such as diabetes, with a particular focus on γ conglutins for their insulin regulatory capacity in animal models, highlighting their nutritional and therapeutic impact on health (Vargas-Guerrero et al. 2014; Ishaq et al. 2022). In this study, we identified 30 genes belonging to the conglutin family in the genome of *L. luteus*, nearly doubling that previously reported (Foley et al. 2015). Intriguingly, our comparative analysis across *Lupinus* species revealed a significant expansion in the conglutin δ 2 subfamily in *L. luteus*. The conglutin δ subfamily has been studied not only for its role as a storage protein but also as a defence protein due to its homology with the α-amylase/trypsin inhibitor family (Tahmasian et al. 2022). Additionally, several publications have highlighted the low digestibility and allergenic potential (Ogura et al. 2013; Dooper et al. 2009; Holden et al. 2008). Furthermore, a recent study showcased higher expression of conglutin δ 2 in wild accessions of *L. angustifolius* compared to domesticated ones (Tahmasian et al. 2022). Interestingly, a study conducted in *L. luteus* cv. Tapper revealed that conglutin δ is the predominant storage protein, contrasting with other lupin species where conglutin β is predominant (Klajn et al. 2023). These findings lead us to hypothesize that this phenomenon may be attributed to the expansion of the delta conglutin family in *L. luteus*.

Plant resistance to pathogens relies predominantly on disease resistance proteins, specifically those belonging to the NBS-type class, which are associated with defense against infections caused by fungi, bacteria, or viruses (McHale et al. 2006; Dangl and Jones 2001). In this work, we identified 80 NBS-type proteins encoded in the genome of *L. luteus*. This finding aligns with a previous report indicating that the genome of *L. angustifolius* contains 67 NBS-type R genes (Garg et al. 2022). Notably, a comparative analysis between *L. angustifolius* and *L. albus* revealed an underrepresentation of R proteins in these lupin species when compared to other leguminous species (Garg et al. 2022). Our results are consistent with previous reports for both lupin species, emphasizing the uniqueness of their R protein repertoire.

Additionally, our study revealed that genes located in similar clusters not only shared proximity but also exhibited a closer phylogenetic relationship. This observation suggests that tandem duplication may serve as a mechanism driving the expansion of this gene family. Importantly, this mechanism has been previously documented in species such as *N. indica* (J.-S. Yang et al. 2022). These insights into the genetic basis of pathogen resistance in lupins, emphasises the significance of tandem duplication in shaping the diversity and evolution of R proteins in these leguminous plants.

In lupins, alkaloids serve protective functions and are also linked to the bitter taste of the seeds, reducing their attractiveness for exploitation and consumption (Frick et al. 2017; Schryvers et al. 2023; Rodés-Bachs and Van der Fels-Klerx 2023; Mancinotti, Frick, and Geu-Flores 2022). Utilizing a domain-based protein search and sequence similarity analysis, 16 genes were identified as previously associated with the biosynthesis and regulation of quinolizidine alkaloids (QA) in other lupins (Frick et al. 2017; Plewiński et al. 2019; Czepiel et al. 2021; Kroc et al. 2019). Thus the two key enzymes crucial for QA biosynthesis, lysine/ornithine decarboxylase (LDC) and copper amine oxidase (CAO), are predominantly expressed in

tissues such as leaves and pedicels (Frick et al. 2017; Wink and Hartmann 1982). This finding is consistent with previous reports in *L. angustifolius*, where higher expression of these key enzymes was observed in leaves and pedicels (T. Yang et al. 2017). In this way, similar to *L. angustifolius*, QA biosynthesis in *L. luteus* primarily occurs in aerial tissues, with subsequent transportation to the seeds (Otterbach et al. 2019). Thus the importance of the spatial and temporal regulation of QA biosynthesis in lupins, hinting at the potential for future research and applications in crop improvement.

The translocation of quinolizidine alkaloids (QA) in lupins has been previously investigated, revealing that Lupanine is transported from aerial tissues and accumulates in stems and petioles in *L. polyphylus*. In a recent study, it was determined that a purine protease transporter 1 (PUP1) might actively participate in the biosynthesis and transport of QA in *L. angustifolius*, as its expression is positively associated with genes in the QA biosynthetic pathway. Based on this assumption, we explored various secondary metabolite transporters, focusing on the PUP transporter family (Frick et al. 2017). This revealed the presence of a *L. angustifolius Lan*PUP ortholog (TanjilG_28431) in the genome of *L. luteus* (Llu07983). Expression analysis of this gene showed predominant expression in leaves, pedicels, and cotyledons, although no statistically significant correlations were found between its expression and key genes associated with QA biosynthesis. 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.

In summary, our work gives insights into the genomic landscape of *Lupinus luteus* that might be exploited in research and breeding. The high-quality genome assembly, coupled with comprehensive gene annotations, gives information about the genetic basis of important agronomic traits. The identified disease resistance genes, alkaloid biosynthesis pathways, and transporters shed light on lupin's defence mechanisms and alkaloid translocation.

Material and methods

Plant materials and sequencing

L. luteus seeds (Alu-Prot CGNA; C195) were scarified and hydrated for 20 minutes in water, washed with 70% ethanol for one minute, in 2.5% sodium hypochlorite, and finally washed three times with sterile distilled water. Then, seeds were put in Petri dishes with moistened filter paper (Whatman) and kept in darkness at 23°C for 24h to germination. 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 quantification 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 and run at 100V for 1 hour. After extraction, ~ 500 ng of gDNA was sequenced on the Illumina NovaSeq 6000 and PacBio Sequel II platforms.

RNA data were extracted separately, from three different 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 them in 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 100V for 1 hour. 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 into ~ 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 (Y. Chen et al. 2018). Then, the cleaned data were used for *K-mer-based* frequency distribution in Jellyfish v2.3.0 (Marçais and Kingsford 2011). To estimate the genome size, heterozygosity, and repeat content of the *L. luteus* genome, GenomeScope v2.0 was used (Vurture et al. 2017).

For the novo assembly, PaBio HiFi long reads were cleaned and assembled using the HifiASM v0.19.6 (Cheng et al. 2021, 2022) assembler with an aggressive purge-dups option. Then, the contigs were polished using two rounds of nextpolish v1.4.1 (Hu et al. 2020), 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 (Durand et al. 2016) with default parameters. The ordering and orientation of contigs was carried out using 3D-DNA v180922 (Dudchenko et al. 2017).

The Burrows-Wheeler Aligner (BWA) v0.7.17-r1188 (H. Li and Durbin 2010) 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 (Mikheenko et al. 2018) and BUSCO v5.4.7 (Manni et al. 2021) by using 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 (Tempel 2012) (http://www.repeatmasker.org) was used for a homology search against known repeat sequences in Repbase database v28.08 (Bao, Kojima, and Kohany 2015). For *ab initio* prediction, a custom-built *de*

novo repeat database with RepeatModeler v2.0.4 (Flynn et al. 2020) was produced and then annotated with RepeatMasker.

Non-coding RNA (ncRNA) was predicted using StructRNAFinder (Arias-Carrasco et al. 2018) with ViennaRNA v2.6.3 (Lorenz et al. 2011), Infernal v1.1.4 (Nawrocki and Eddy 2013), and Rfam 14.9 (November 2022, 4108 families) (Kalvari et al. 2021) as reference for rRNA, microRNA and other ncRNA classes prediction. tRNA annotations were carried out with tRNAscanSE v2.0.12 (Chan and Lowe 2019).

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 (Stanke et al. 2008) employing the Arabidopsis training gene model, and GeneMark-ES v4.71 (Lomsadze et al. 2005). For RNA-seq-based prediction, RNA-seq data of *L. luteus* from different tissues (see Supplementary Table S6) were downloaded from NCBI SRA (https://www.ncbi.nlm.nih.gov/sra/, accessed on August 4th, 2023). Next, low-quality reads were filtered using fastp v0.23.2 (S. Chen et al. 2018), and the remaining high-quality RNA-seq reads were mapped to the assembly of the C195 genome using hisat2 v2.2.1 (Kim et al. 2019). Mapped reads were then employed for genome-based transcript assembly using Cufflinks v2.2.1 (Trapnell et al. 2010) and Stringtie v2.2.1 (Pertea et al. 2015). Afterwards, all transcripts were used to generate a transcriptome assembly using PASA v2.5.3.

For the protein-based homology search, MMseqs2 v14-7e284 (Steinegger and Söding 2017) and GeMoMa v1.9 (Keilwagen, Hartung, and Grau 2019) 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 (https://www.ncbi.nlm.nih.gov; downloaded on August 17th, 2023). The results of these three strategies were combined by EVidence Modeler v2.1.0 (Haas et al. 2008).

The predicted protein-coding genes were functionally annotated through several methods, such as InterProscan v5.62-94.0 (Jones et al. 2014) for protein domain annotation, EggNOG-mapper v2.1.11-1 (Cantalapiedra et al. 2021) with EggNOG v5.0.2 (Huerta-Cepas et al. 2019)(REF) as a reference database and BLAST v2.14.1 (Altschul et al. 1990), for queried protein sequence against all UniProt Release 2023_05 (SwissProt + Trembl) database (UniProt Consortium 2023) (https://www.uniprot.org/, 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 (Aramaki et al. 2020). 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 (Pingchuan Li et al. 2016), and DRAGO2 (DRAGO2-API) deposited in PRGdb 3.0 (Osuna-Cruz et al. 2018) (http://prgdb.org/prgdb/, 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 (Kourelis et al. 2021) were downloaded and the NBS (NB-ARC) domain (PF00931) HMM profile from Pfam v36.0 (Mistry

et al. 2021). Next, following the methods of Yang et al. (J.-S. Yang et al. 2022), BLASTp using DIAMOND v2.0.14.152 (E-value = 1e-10, query coverage = 80, subject coverage = 50) and hmmsearch (E-value = 1e-5) from HMMER v3.4 (Eddy 2011) (http://hmmer.org) 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 (Kroc et al. 2019; Czepiel et al. 2021; Plewiński et al. 2019; Otterbach et al. 2019). 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 (https://www.genome.jp/kegg/). Then, a search was performed using the same BLASTp parameters employed for the R-genes search.

Secondary metabolite transporter families were characterized according to those described by Nogia and Pati, where they mention that the ABC, MATE, PUP, and NTR/PTR (NPF) families are mainly responsible for the transport of secondary metabolites in plants (Nogia and Pati 2021). First, Transporter Classification Database (TCDB) (Saier et al. 2021) data were downloaded (https://tcdb.org/, 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 and used to search in protein-coding genes with the same parameters described above. Likewise, a dataset of manually selected ABC transporters, obtained from Hou et al. (Hou, Wang, and Wu 2020) was queried with the same search parameters used with TCDB. Transmembrane regions of transporter proteins were predicted by DeepTMHMM v1.0.24 (https://dtu.biolib.com/DeepTMHMM).

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 (Emms and Kelly 2019) 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 with default parameters. These data were employed for phylogenetic tree inference using PhyML v3.3.3 (Guindon et al. 2010). The divergence time was calculated and calibrated as reported by Jiu et al. (Jiu et al. 2023), using MCMCtree from PAML v4.9j (Z. Yang 1997) and TimeTree v5 (http://timetree.org/) (Kumar et al. 2022). To measure the expansion and contraction of gene families in the *L. luteus* genome, CAFE v5.1.0 (Mendes et al. 2021) was used. GO enrichment analysis was performed on unique genes, expanded and contracted families using the Cytoscape app BiNGO v3.0.5 (Maere, Heymans, and Kuiper 2005) in Cytoscape v3.10.1 (https://cytoscape.org/).

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. (Jiu et al. 2023) using MUMmer v4.0.0 (Kurtz et al. 2004). Then, to evaluate synteny blocks between the three genomes MCScan(Tang et al. 2008) was employed through the JCVI package v1.3.8 (https://github.com/tanghaibao/jcvi) 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 featureCounts v2.0.6 (Liao, Smyth, and Shi 2014). Next, normalized counts TMM were obtained with EdgeR v4.0.3 (Robinson, McCarthy, and Smyth 2010). The batch effect was removed by applying parametric empirical Bayes frameworks adjustment with SVA v3.50.0 (Leek et al. 2012).

Declarations

Author Contributions

H.S.G. Scientific design of the research, planned scientific analysis, experiment and decided plant material. Oriented and guide 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 revision 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.

Acknowledgments

This research was funded by the Agri-aquaculture Nutritional Genomic Center (CGNA), ANID project R20F0003 and GORE La Araucanía, Chile.

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. All data are available from the corresponding author upon request.

References

1. Abraham, Eleni M., Ioannis Ganopoulos, Panagiotis Madesis, Athanasios Mavromatis, Photini Mylona, Irini Nianiou-Obeidat, Zoi Parissi, Alexios Polidoros, Eleni Tani, and Dimitrios Vlachostergios. 2019. "The Use of Lupin as a Source of Protein in Animal Feeding: Genomic Tools and Breeding Approaches." *International Journal of Molecular Sciences* 20 (4). https://doi.org/10.3390/ijms20040851.
2. Ainouche, A. K., and R. J. Bayer. 1999. "Phylogenetic Relationships in Lupinus (Fabaceae: Papilionoideae) Based on Internal Transcribed Spacer Sequences (ITS) of Nuclear Ribosomal DNA." *American Journal of Botany* 86 (4): 590–607.
3. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10.
4. Aramaki, Takuya, Romain Blanc-Mathieu, Hisashi Endo, Koichi Ohkubo, Minoru Kanehisa, Susumu Goto, and Hiroyuki Ogata. 2020. "KofamKOALA: KEGG Ortholog Assignment Based on Profile HMM and Adaptive Score Threshold." *Bioinformatics* 36 (7): 2251–52.
5. Arias-Carrasco, Raúl, Yessenia Vásquez-Morán, Helder I. Nakaya, and Vinicius Maracaja-Coutinho. 2018. "StructRNAfinder: An Automated Pipeline and Web Server for RNA Families Prediction." *BMC Bioinformatics* 19 (1): 55.
6. Bao, Weidong, Kenji K. Kojima, and Oleksiy Kohany. 2015. "Repbase Update, a Database of Repetitive Elements in Eukaryotic Genomes." *Mobile DNA* 6 (June): 11.
7. Bayer, Philipp E., Agnieszka A. Golicz, Soodeh Tirnaz, Chon-Kit Kenneth Chan, David Edwards, and Jacqueline Batley. 2019. "Variation in Abundance of Predicted Resistance Genes in the Brassica Oleracea Pangenome." *Plant Biotechnology Journal* 17 (4): 789–800.
8. Bielski, Wojciech, Michał Książkiewicz, Denisa Šimoníková, Eva Hřibová, Karolina Susek, and Barbara Naganowska. 2020. "The Puzzling Fate of a Lupin Chromosome Revealed by Reciprocal Oligo-FISH and BAC-FISH Mapping." *Genes* 11 (12). https://doi.org/10.3390/genes11121489.
9. Burgos-Díaz, César, Traudy Wandersleben, Marcos Olivos, Nicole Lichtin, Mariela Bustamante, and Conxita Solans. 2019. "Food-Grade Pickering Stabilizers Obtained from a Protein-Rich Lupin Cultivar (AluProt-CGNA®): Chemical Characterization and Emulsifying Properties." *Food Hydrocolloids* 87 (February): 847–57.
10. Cantalapiedra, Carlos P., Ana Hernández-Plaza, Ivica Letunic, Peer Bork, and Jaime Huerta-Cepas. 2021. "eggNOG-Mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale." *Molecular Biology and Evolution* 38 (12): 5825–29.

11. Carvajal-Larenas, Francisco E. 2019. “Nutritional, Rheological and Sensory Evaluation of Lupinus Mutabilis Food Products - a Review.” *Czech Journal of Food Science* 37 (5): 301–11.
12. Chan, Patricia P., and Todd M. Lowe. 2019. “tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences.” *Methods in Molecular Biology* 1962: 1–14.
13. Cheng, Haoyu, Gregory T. Concepcion, Xiaowen Feng, Haowen Zhang, and Heng Li. 2021. “Haplotype-Resolved de Novo Assembly Using Phased Assembly Graphs with Hifiasm.” *Nature Methods* 18 (2): 170–75.
14. Cheng, Haoyu, Erich D. Jarvis, Olivier Fedrigo, Klaus-Peter Koepfli, Lara Urban, Neil J. Gemmell, and Heng Li. 2022. “Haplotype-Resolved Assembly of Diploid Genomes without Parental Data.” *Nature Biotechnology* 40 (9): 1332–35.
15. Chen, Shifu, Yanqing Zhou, Yaru Chen, and Jia Gu. 2018. “Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor.” *Bioinformatics* 34 (17): i884–90.
16. Chen, Yuxin, Yongsheng Chen, Chunmei Shi, Zhibo Huang, Yong Zhang, Shengkang Li, Yan Li, et al. 2018. “SOAPnuke: A MapReduce Acceleration-Supported Software for Integrated Quality Control and Preprocessing of High-Throughput Sequencing Data.” *GigaScience* 7 (1): 1–6.
17. Czepiel, Katarzyna, Paweł Krajewski, Paulina Wilczura, Patrycja Bielecka, Wojciech Święcicki, and Magdalena Kroc. 2021. “Expression Profiles of Alkaloid-Related Genes across the Organs of Narrow-Leafed Lupin (L.) and in Response to Anthracnose Infection.” *International Journal of Molecular Sciences* 22 (5). https://doi.org/10.3390/ijms22052676.
18. Dangl, J. L., and J. D. Jones. 2001. “Plant Pathogens and Integrated Defence Responses to Infection.” *Nature* 411 (6839): 826–33.
19. Dooper, M. M. B. W., C. Plassen, L. Holden, H. Lindvik, and C. K. Faeste. 2009. “Immunoglobulin E Cross-Reactivity between Lupine Conglutins and Peanut Allergens in Serum of Lupine-Allergic Individuals.” *Journal of Investigational Allergology & Clinical Immunology: Official Organ of the International Association of Asthmology and Sociedad Latinoamericana de Alergia E Inmunologia* 19 (4): 283–91.
20. Drummond, Christopher S., Ruth J. Eastwood, Silvia T. S. Miotto, and Colin E. Hughes. 2012. “Multiple Continental Radiations and Correlates of Diversification in Lupinus (Leguminosae): Testing for Key Innovation with Incomplete Taxon Sampling.” *Systematic Biology* 61 (3): 443–60.
21. Dudchenko, Olga, Sanjit S. Batra, Arina D. Omer, Sarah K. Nyquist, Marie Hoeger, Neva C. Durand, Muhammad S. Shamim, et al. 2017. “De Novo Assembly of the Genome Using Hi-C Yields Chromosome-Length Scaffolds.” *Science* 356 (6333): 92–95.
22. Durand, Neva C., Muhammad S. Shamim, Ido Machol, Suhas S. P. Rao, Miriam H. Huntley, Eric S. Lander, and Erez Lieberman Aiden. 2016. “Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments.” *Cell Systems* 3 (1): 95–98.
23. Duranti, Marcello, Alessandro Consonni, Chiara Magni, Fabio Sessa, and Alessio Scarafoni. 2008. “The Major Proteins of Lupin Seed: Characterisation and Molecular Properties for Use as Functional and Nutraceutical Ingredients.” *Trends in Food Science & Technology* 19 (12): 624–33.

24. Eddy, Sean R. 2011. “Accelerated Profile HMM Searches.” *PLoS Computational Biology* 7 (10): e1002195.
25. Emms, David M., and Steven Kelly. 2019. “OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics.” *Genome Biology* 20 (1): 238.
26. Flynn, Jullien M., Robert Hubley, Clément Goubert, Jeb Rosen, Andrew G. Clark, Cédric Feschotte, and Arian F. Smit. 2020. “RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families.” *Proceedings of the National Academy of Sciences of the United States of America* 117 (17): 9451–57.
27. Foley, Rhonda C., Jose C. Jimenez-Lopez, Lars G. Kamphuis, James K. Hane, Su Melser, and Karam B. Singh. 2015. “Analysis of Conglutin Seed Storage Proteins across Lupin Species Using Transcriptomic, Protein and Comparative Genomic Approaches.” *BMC Plant Biology* 15 (April): 106.
28. Frick, Karen M., Lars G. Kamphuis, Kadambot H. M. Siddique, Karam B. Singh, and Rhonda C. Foley. 2017. “Quinolizidine Alkaloid Biosynthesis in Lupins and Prospects for Grain Quality Improvement.” *Frontiers in Plant Science* 8 (January): 87.
29. Garg, Gagan, Lars G. Kamphuis, Philipp E. Bayer, Parwinder Kaur, Olga Dudchenko, Candy M. Taylor, Karen M. Frick, et al. 2022. “A Pan-Genome and Chromosome-Length Reference Genome of Narrow-Leafed Lupin (Lupinus Angustifolius) Reveals Genomic Diversity and Insights into Key Industry and Biological Traits.” *The Plant Journal: For Cell and Molecular Biology* 111 (5): 1252–66.
30. Glazinska, Paulina, Milena Kulasek, Wojciech Glinkowski, Marta Wysocka, and Jan Grzegorz Kosiński. 2020. “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.” *Frontiers in Genetics* 11 (May): 455.
31. Gresta, Fabio, Valerio Abbate, Giovanni Avola, Giuseppe Magazzù, and Biagina Chiofalo. 2010. “Lupin Seed for the Crop-Livestock Food Chain.” *Italian Journal of Agronomy* 5 (4): 333.
32. Guindon, Stéphane, Jean-François Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, and Olivier Gascuel. 2010. “New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0.” *Systematic Biology* 59 (3): 307–21.
33. Haas, Brian J., Steven L. Salzberg, Wei Zhu, Mihaela Pertea, Jonathan E. Allen, Joshua Orvis, Owen White, C. Robin Buell, and Jennifer R. Wortman. 2008. “Automated Eukaryotic Gene Structure Annotation Using EVidenceModeler and the Program to Assemble Spliced Alignments.” *Genome Biology* 9 (1): R7.
34. Hane, James K., Yao Ming, Lars G. Kamphuis, Matthew N. Nelson, Gagan Garg, Craig A. Atkins, Philipp E. Bayer, et al. 2017. “A Comprehensive Draft Genome Sequence for Lupin (Lupinus Angustifolius), an Emerging Health Food: Insights into Plant-Microbe Interactions and Legume Evolution.” *Plant Biotechnology Journal* 15 (3): 318–30.
35. Hildreth, Sherry B., Elizabeth A. Gehman, Haibing Yang, Rong-He Lu, K. C. Ritesh, Kim C. Harich, Shi Yu, et al. 2011. “Tobacco Nicotine Uptake Permease (NUP1) Affects Alkaloid Metabolism.”

Proceedings of the National Academy of Sciences of the United States of America 108 (44): 18179–84.

36. Holden, Lise, Gaynour B. G. Sletten, Helene Lindvik, Christiane K. Faeste, and Maaike M. B. W. Dooper. 2008. “Characterization of IgE Binding to Lupin, Peanut and Almond with Sera from Lupin-Allergic Patients.” *International Archives of Allergy and Immunology* 146 (4): 267–76.
37. Hou, Ruiyan, Lida Wang, and Yi-Jun Wu. 2020. “Predicting ATP-Binding Cassette Transporters Using the Random Forest Method.” *Frontiers in Genetics* 11 (March): 156.
38. Huerta-Cepas, Jaime, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K. Forslund, Helen Cook, Daniel R. Mende, et al. 2019. “eggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses.” *Nucleic Acids Research* 47 (D1): D309–14.
39. Hufnagel, Bárbara, André Marques, Alexandre Soriano, Laurence Marquès, Fanchon Divol, Patrick Doumas, Erika Sallet, et al. 2020. “High-Quality Genome Sequence of White Lupin Provides Insight into Soil Exploration and Seed Quality.” *Nature Communications* 11 (1): 492.
40. Hu, Jiang, Junpeng Fan, Zongyi Sun, and Shanlin Liu. 2020. “NextPolish: A Fast and Efficient Genome Polishing Tool for Long-Read Assembly.” *Bioinformatics* 36 (7): 2253–55.
41. Iqbal, Muhammad Munir, Mark Huynh, Joshua A. Udall, Andrzej Kilian, Kedar N. Adhikari, Jens D. Berger, William Erskine, and Matthew N. Nelson. 2019. “The First Genetic Map for Yellow Lupin Enables Genetic Dissection of Adaptation Traits in an Orphan Grain Legume Crop.” *BMC Genetics* 20 (1): 68.
42. Ishaq, Ali Raza, Heba A. S. El-Nashar, Tahira Younis, Muhammad Asad Mangat, Mashal Shahzadi, Amina Shamsheer Ul Haq, and Mohamed El-Shazly. 2022. “Genus Lupinus (Fabaceae): A Review of Ethnobotanical, Phytochemical and Biological Studies.” *The Journal of Pharmacy and Pharmacology* 74 (12): 1700–1717.
43. Jiu, Songtao, Baozheng Chen, Xiao Dong, Zhengxin Lv, Yuxuan Wang, Chunjin Yin, Yan Xu, et al. 2023. “Chromosome-Scale Genome Assembly of Provides Novel Insights into Genome Evolution, Disease Resistance, and Dormancy Release in L.” *Horticulture Research* 10 (5): uhad062.
44. Jones, Philip, David Binns, Hsin-Yu Chang, Matthew Fraser, Weizhong Li, Craig McAnulla, Hamish McWilliam, et al. 2014. “InterProScan 5: Genome-Scale Protein Function Classification.” *Bioinformatics* 30 (9): 1236–40.
45. Kalvari, Ioanna, Eric P. Nawrocki, Nancy Ontiveros-Palacios, Joanna Argasinska, Kevin Lamkiewicz, Manja Marz, Sam Griffiths-Jones, et al. 2021. “Rfam 14: Expanded Coverage of Metagenomic, Viral and microRNA Families.” *Nucleic Acids Research* 49 (D1): D192–200.
46. Keilwagen, Jens, Frank Hartung, and Jan Grau. 2019. “GeMoMa: Homology-Based Gene Prediction Utilizing Intron Position Conservation and RNA-Seq Data.” *Methods in Molecular Biology* 1962: 161–77.
47. Kim, Daehwan, Joseph M. Paggi, Chanhee Park, Christopher Bennett, and Steven L. Salzberg. 2019. “Graph-Based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype.” *Nature*

Biotechnology 37 (8): 907–15.

48. Klajn, Natalia, Katarzyna Kapczyńska, Paweł Pasikowski, Paulina Glazińska, Hubert Kugiel, Jacek Kęsy, and Waldemar Wojciechowski. 2023. “Regulatory Effects of ABA and GA on the Expression of Conglutin Genes and Network Genes in Yellow Lupine (L.) Seeds.” *International Journal of Molecular Sciences* 24 (15). https://doi.org/10.3390/ijms241512380.
49. Kourelis, Jiorgos, Toshiyuki Sakai, Hiroaki Adachi, and Sophien Kamoun. 2021. “RefPlantNLR Is a Comprehensive Collection of Experimentally Validated Plant Disease Resistance Proteins from the NLR Family.” *PLoS Biology* 19 (10): e3001124.
50. Kroc, Magdalena, Grzegorz Koczyk, Katarzyna A. Kamel, Katarzyna Czepiel, Olga Fedorowicz-Strońska, Paweł Krajewski, Joanna Kosińska, Jan Podkowiński, Paulina Wilczura, and Wojciech Święcicki. 2019. “Transcriptome-Derived Investigation of Biosynthesis of Quinolizidine Alkaloids in Narrow-Leafed Lupin (Lupinus Angustifolius L.) Highlights Candidate Genes Linked to Iucundus Locus.” *Scientific Reports* 9 (1): 2231.
51. Książkiewicz, Michał, Nelson Nazzicari, Hua ’an Yang, Matthew N. Nelson, Daniel Renshaw, Sandra Rychel, Barbara Ferrari, et al. 2017. “A High-Density Consensus Linkage Map of White Lupin Highlights Synteny with Narrow-Leafed Lupin and Provides Markers Tagging Key Agronomic Traits.” *Scientific Reports* 7 (1): 15335.
52. Kumar, Sudhir, Michael Suleski, Jack M. Craig, Adrienne E. Kasprowicz, Maxwell Sanderford, Michael Li, Glen Stecher, and S. Blair Hedges. 2022. “TimeTree 5: An Expanded Resource for Species Divergence Times.” *Molecular Biology and Evolution* 39 (8). https://doi.org/10.1093/molbev/msac174.
53. Kurtz, Stefan, Adam Phillippy, Arthur L. Delcher, Michael Smoot, Martin Shumway, Corina Antonescu, and Steven L. Salzberg. 2004. “Versatile and Open Software for Comparing Large Genomes.” *Genome Biology* 5 (2): R12.
54. Lambers, Hans, Jon C. Clements, and Matthew N. Nelson. 2013. “How a Phosphorus-Acquisition Strategy Based on Carboxylate Exudation Powers the Success and Agronomic Potential of Lupines (Lupinus, Fabaceae).” *American Journal of Botany* 100 (2): 263–88.
55. Leek, Jeffrey T., W. Evan Johnson, Hilary S. Parker, Andrew E. Jaffe, and John D. Storey. 2012. “The Sva Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments.” *Bioinformatics* 28 (6): 882–83.
56. Liao, Yang, Gordon K. Smyth, and Wei Shi. 2014. “featureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features.” *Bioinformatics* 30 (7): 923–30.
57. Lichtin, Nicole, Haroldo Salvo-Garrido, Bradley Till, Peter D. S. Caligari, Annally Rupayan, Fernando Westermeyer, and Marcos Olivos. 2020. “Genetic and Comparative Mapping of Lupinus Luteus L. Highlight Syntenic Regions with Major Orthologous Genes Controlling Anthracnose Resistance and Flowering Time.” *Scientific Reports* 10 (1): 19174.
58. Li, Heng, and Richard Durbin. 2010. “Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform.” *Bioinformatics* 26 (5): 589–95.

59. Li, Pai, and Brad Day. 2019. “Battlefield Cytoskeleton: Turning the Tide on Plant Immunity.” *Molecular Plant-Microbe Interactions: MPMI* 32 (1): 25–34.
60. Li, Pingchuan, Xiande Quan, Gaofeng Jia, Jin Xiao, Sylvie Cloutier, and Frank M. You. 2016. “RGAugury: A Pipeline for Genome-Wide Prediction of Resistance Gene Analogs (RGAs) in Plants.” *BMC Genomics* 17 (1): 852.
61. Lomsadze, Alexandre, Vardges Ter-Hovhannisyan, Yury O. Chernoff, and Mark Borodovsky. 2005. “Gene Identification in Novel Eukaryotic Genomes by Self-Training Algorithm.” *Nucleic Acids Research* 33 (20): 6494–6506.
62. Lorenz, Ronny, Stephan H. Bernhart, Christian Höner Zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F. Stadler, and Ivo L. Hofacker. 2011. “ViennaRNA Package 2.0.” *Algorithms for Molecular Biology: AMB* 6 (November): 26.
63. Lucas, M. Mercedes, Frederick L. Stoddard, Paolo Annicchiarico, Juana Frías, Cristina Martínez-Villaluenga, Daniela Sussmann, Marcello Duranti, Alice Seger, Peter M. Zander, and José J. Pueyo. 2015. “The Future of Lupin as a Protein Crop in Europe.” *Frontiers in Plant Science* 6 (September): 705.
64. Maere, Steven, Karel Heymans, and Martin Kuiper. 2005. “BiNGO: A Cytoscape Plugin to Assess Overrepresentation of Gene Ontology Categories in Biological Networks.” *Bioinformatics* 21 (16): 3448–49.
65. Mancinotti, Davide, Karen Michiko Frick, and Fernando Geu-Flores. 2022. “Biosynthesis of Quinolizidine Alkaloids in Lupins: Mechanistic Considerations and Prospects for Pathway Elucidation.” *Natural Product Reports* 39 (7): 1423–37.
66. Manni, Mosè, Matthew R. Berkeley, Mathieu Seppey, Felipe A. Simão, and Evgeny M. Zdobnov. 2021. “BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes.” *Molecular Biology and Evolution* 38 (10): 4647–54.
67. Marçais, Guillaume, and Carl Kingsford. 2011. “A Fast, Lock-Free Approach for Efficient Parallel Counting of Occurrences of K-Mers.” *Bioinformatics* 27 (6): 764–70.
68. Martínez-Villaluenga, C., J. Frías, and C. Vidal-Valverde. 2006. “Functional Lupin Seeds (Lupinus Albus L. and Lupinus Luteus L.) after Extraction of α-Galactosides.” *Food Chemistry* 98 (2): 291–99.
69. McHale, Leah, Xiaoping Tan, Patrice Koehl, and Richard W. Michelmore. 2006. “Plant NBS-LRR Proteins: Adaptable Guards.” *Genome Biology* 7 (4): 212.
70. Mendes, Fábio K., Dan Vanderpool, Ben Fulton, and Matthew W. Hahn. 2021. “CAFE 5 Models Variation in Evolutionary Rates among Gene Families.” *Bioinformatics* 36 (22-23): 5516–18.
71. Mikheenko, Alla, Andrey Prjibelski, Vladislav Saveliev, Dmitry Antipov, and Alexey Gurevich. 2018. “Versatile Genome Assembly Evaluation with QUAST-LG.” *Bioinformatics* 34 (13): i142–50.
72. Mistry, Jaina, Sara Chuguransky, Lowri Williams, Matloob Qureshi, Gustavo A. Salazar, Erik L. L. Sonnhammer, Silvio C. E. Tosatto, et al. 2021. “Pfam: The Protein Families Database in 2021.” *Nucleic Acids Research* 49 (D1): D412–19.

73. Mithöfer, Axel, and Wilhelm Boland. 2012. “Plant Defense against Herbivores: Chemical Aspects.” *Annual Review of Plant Biology* 63 (February): 431–50.
74. Musco, N., M. I. Cutrignelli, S. Calabrò, R. Tudisco, F. Infascelli, R. Grazioli, V. Lo Presti, F. Gresta, and B. Chiofalo. 2017. “Comparison of Nutritional and Antinutritional Traits among Different Species (Lupinus Albus L., Lupinus Luteus L., Lupinus Angustifolius L.) and Varieties of Lupin Seeds.” *Journal of Animal Physiology and Animal Nutrition* 101 (6): 1227–41.
75. Nawrocki, Eric P., and Sean R. Eddy. 2013. “Infernal 1.1: 100-Fold Faster RNA Homology Searches.” *Bioinformatics* 29 (22): 2933–35.
76. Nogia, Panchsheela, and Pratap Kumar Pati. 2021. “Plant Secondary Metabolite Transporters: Diversity, Functionality, and Their Modulation.” *Frontiers in Plant Science* 12 (October): 758202.
77. Ogura, Takahiro, Adrián Hernández, Tomoko Aizawa, Jun Ogihara, Michio Sunairi, Javier Alcaino, Haroldo Salvo-Garrido, and Iván J. Maureira-Butler. 2013. “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* 8 (11): e80369.
78. Osuna-Cruz, Cristina M., Andreu Paytuvi-Gallart, Antimo Di Donato, Vicky Sundesha, Giuseppe Andolfo, Riccardo Aiese Cigliano, Walter Sanseverino, and Maria R. Ercolano. 2018. “PRGdb 3.0: A Comprehensive Platform for Prediction and Analysis of Plant Disease Resistance Genes.” *Nucleic Acids Research* 46 (D1): D1197–1201.
79. Otterbach, Sophie Lisa, Ting Yang, Lucilia Kato, Christian Janfelt, and Fernando Geu-Flores. 2019. “Quinolizidine Alkaloids Are Transported to Seeds of Bitter Narrow-Leafed Lupin.” *Journal of Experimental Botany* 70 (20): 5799–5808.
80. Ou, Shujun, Jinfeng Chen, and Ning Jiang. 2018. “Assessing Genome Assembly Quality Using the LTR Assembly Index (LAI).” *Nucleic Acids Research* 46 (21): e126.
81. Parra-González, Lorena B., Gabriela A. Aravena-Abarzúa, Cristell S. Navarro-Navarro, Joshua Udall, Jeff Maughan, Louis M. Peterson, Haroldo E. Salvo-Garrido, and Iván J. Maureira-Butler. 2012. “Yellow Lupin (Lupinus Luteus L.) Transcriptome Sequencing: Molecular Marker Development and Comparative Studies.” *BMC Genomics* 13 (August): 425.
82. Pertea, Mihaela, Geo M. Pertea, Corina M. Antonescu, Tsung-Cheng Chang, Joshua T. Mendell, and Steven L. Salzberg. 2015. “StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads.” *Nature Biotechnology* 33 (3): 290–95.
83. Pexas, Georgios, Bob Doherty, and Ilias Kyriazakis. 2023. “The Future of Protein Sources in Livestock Feeds: Implications for Sustainability and Food Safety.” *Frontiers in Sustainable Food Systems* 7 (September). https://doi.org/10.3389/fsufs.2023.1188467.
84. Plewiński, Piotr, Michał Książkiewicz, Sandra Rychel-Bielska, Elżbieta Rudy, and Bogdan Wolko. 2019. “Candidate Domestication-Related Genes Revealed by Expression Quantitative Trait Loci Mapping of Narrow-Leafed Lupin (L.).” *International Journal of Molecular Sciences* 20 (22). https://doi.org/10.3390/ijms20225670.

85. Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. 2010. “edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data.” *Bioinformatics* 26 (1): 139–40.
86. Rodés-Bachs, Clàudia, and H. J. Van der Fels-Klerx. 2023. “Impact of Environmental Factors on the Presence of Quinolizidine Alkaloids in Lupins: A Review.” *Food Additives & Contaminants. Part A, Chemistry, Analysis, Control, Exposure & Risk Assessment* 40 (6): 757–69.
87. Ruiz-López, Mario Alberto, Lucia Barrientos-Ramírez, Pedro Macedonio García-López, Elia Herminia Valdés-Miramontes, Juan Francisco Zamora-Natera, Ramón Rodríguez-Macias, Eduardo Salcedo-Pérez, Jacinto Bañuelos-Pineda, and J. Jesús Vargas-Radillo. 2019. “Nutritional and Bioactive Compounds in Mexican Lupin Beans Species: A Mini-Review.” *Nutrients* 11 (8). https://doi.org/10.3390/nu11081785.
88. Ruiz, M. A., and A. Sotelo. 2001. “Chemical Composition, Nutritive Value, and Toxicology Evaluation of Mexican Wild Lupins.” *Journal of Agricultural and Food Chemistry* 49 (11): 5336–39.
89. Rychel, Sandra, and Michał Książkiewicz. 2019. “Development of Gene-Based Molecular Markers Tagging Low Alkaloid Pauper Locus in White Lupin (Lupinus Albus L.).” *Journal of Applied Genetics* 60 (3-4): 269–81.
90. Saier, Milton H., Vamsee S. Reddy, Gabriel Moreno-Hagelsieb, Kevin J. Hendargo, Yichi Zhang, Vasu Iddamsetty, Katie Jing Kay Lam, et al. 2021. “The Transporter Classification Database (TCDB): 2021 Update.” *Nucleic Acids Research* 49 (D1): D461–67.
91. Schryvers, Sofie, Chinaza Arinzechukwu, Bram Miserez, Mia Eeckhout, and Liesbeth Jacxsens. 2023. “The Fate of Quinolizidine Alkaloids during the Processing of Lupins (Lupinus Spp.) for Human Consumption.” *Food Chemistry* 429 (December): 136847.
92. Stanke, Mario, Mark Diekhans, Robert Baertsch, and David Haussler. 2008. “Using Native and Syntenically Mapped cDNA Alignments to Improve de Novo Gene Finding.” *Bioinformatics* 24 (5): 637–44.
93. Steinegger, Martin, and Johannes Söding. 2017. “MMseqs2 Enables Sensitive Protein Sequence Searching for the Analysis of Massive Data Sets.” *Nature Biotechnology* 35 (11): 1026–28.
94. Susek, Karolina, Wojciech Bielski, Katarzyna B. Czyż, Robert Hasterok, Scott A. Jackson, Bogdan Wolko, and Barbara Naganowska. 2019. “Impact of Chromosomal Rearrangements on the Interpretation of Lupin Karyotype Evolution.” *Genes* 10 (4). https://doi.org/10.3390/genes10040259.
95. Tahmasian, Arineh, Angéla Juhász, James A. Broadbent, Mitchell G. Nye-Wood, Thao T. Le, and Michelle L. Colgrave. 2022. “Evaluation of the Major Seed Storage Proteins, the Conglutins, Across Genetically Diverse Narrow-Leafed Lupin Varieties.” *Frontiers in Nutrition* 9 (May): 842168.
96. Tang, Haibao, John E. Bowers, Xiyin Wang, Ray Ming, Maqsudul Alam, and Andrew H. Paterson. 2008. “Synteny and Collinearity in Plant Genomes.” *Science* 320 (5875): 486–88.
97. Tempel, Sébastien. 2012. “Using and Understanding RepeatMasker.” *Methods in Molecular Biology* 859: 29–51.
98. Trapnell, Cole, Brian A. Williams, Geo Pertea, Ali Mortazavi, Gordon Kwan, Marijke J. van Baren, Steven L. Salzberg, Barbara J. Wold, and Lior Pachter. 2010. “Transcript Assembly and Quantification

by RNA-Seq Reveals Unannotated Transcripts and Isoform Switching during Cell Differentiation." *Nature Biotechnology* 28 (5): 511–15.

99. UniProt Consortium. 2023. "UniProt: The Universal Protein Knowledgebase in 2023." *Nucleic Acids Research* 51 (D1): D523–31.
100. Vargas-Guerrero, Belinda, Pedro M. García-López, Alma L. Martínez-Ayala, José A. Domínguez-Rosales, and Carmen M. Gurrola-Díaz. 2014. "Administration of Lupinus Albus Gamma Conglutin (Cγ) to n5 STZ Rats Augmented Ins-1 Gene Expression and Pancreatic Insulin Content." *Plant Foods for Human Nutrition* 69 (3): 241–47.
101. Vurture, Gregory W., Fritz J. Sedlazeck, Maria Nattestad, Charles J. Underwood, Han Fang, James Gurtowski, and Michael C. Schatz. 2017. "GenomeScope: Fast Reference-Free Genome Profiling from Short Reads." *Bioinformatics* 33 (14): 2202–4.
102. Wang, Penghao, Gaofeng Zhou, Jianbo Jian, Huaan Yang, Daniel Renshaw, Matthew K. Aubert, Jonathan Clements, Tianhua He, Mark Sweetingham, and Chengdao Li. 2021. "Whole-Genome Assembly and Resequencing Reveal Genomic Imprint and Key Genes of Rapid Domestication in Narrow-Leafed Lupin." *The Plant Journal: For Cell and Molecular Biology* 105 (5): 1192–1210.
103. Wink, M., and T. Hartmann. 1982. "Localization of the Enzymes of Quinolizidine Alkaloid Biosynthesis in Leaf Chloroplasts of Lupinus Polyphyllus." *Plant Physiology* 70 (1): 74–77.
104. Xu, Liang, Yan Wang, Junhui Dong, Wei Zhang, Mingjia Tang, Weilan Zhang, Kai Wang, et al. 2023. "A Chromosome-Level Genome Assembly of Radish (Raphanus Sativus L.) Reveals Insights into Genome Adaptation and Differential Bolting Regulation." *Plant Biotechnology Journal* 21 (5): 990–1004.
105. Yang, Jing-Shan, Zhi-Hao Qian, Tao Shi, Zhi-Zhong Li, and Jin-Ming Chen. 2022. "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 Research: An International Journal for Rapid Publication of Reports on Genes and Genomes* 29 (4). https://doi.org/10.1093/dnares/dsac022.
106. Yang, Ting, Istvan Nagy, Davide Mancinotti, Sophie Lisa Otterbach, Trine Bundgaard Andersen, Mohammed Saddik Motawia, Torben Asp, and Fernando Geu-Flores. 2017. "Transcript Profiling of a Bitter Variety of Narrow-Leafed Lupin to Discover Alkaloid Biosynthetic Genes." *Journal of Experimental Botany* 68 (20): 5527–37.
107. Yang, Z. 1997. "PAML: A Program Package for Phylogenetic Analysis by Maximum Likelihood." *Computer Applications in the Biosciences: CABIOS* 13 (5): 555–56.
108. Zhu, Zhaohai, Fang Xu, Yaxi Zhang, Yu Ti Cheng, Marcel Wiermer, Xin Li, and Yuelin Zhang. 2010. "Arabidopsis Resistance Protein SNC1 Activates Immune Responses through Association with a Transcriptional Corepressor." *Proceedings of the National Academy of Sciences of the United States of America* 107 (31): 13960–65.

Figures

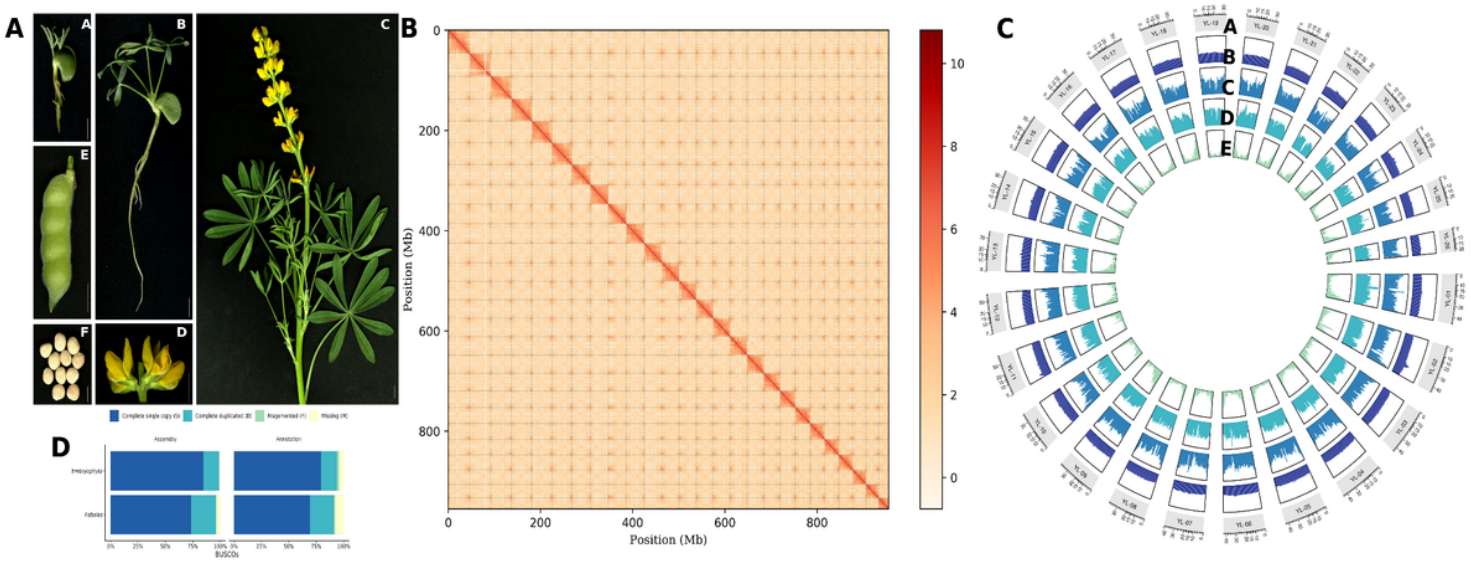

Figure 1

Genome assembly of yellow lupin. A) Morphological characteristics of a *Lupinus 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. C) Circos plot depicting 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 are shown with a window of 1 Mb. D) BUSCO scores of the yellow lupin genome and gene annotation using the Embryophyta and Fabales reference lineage.

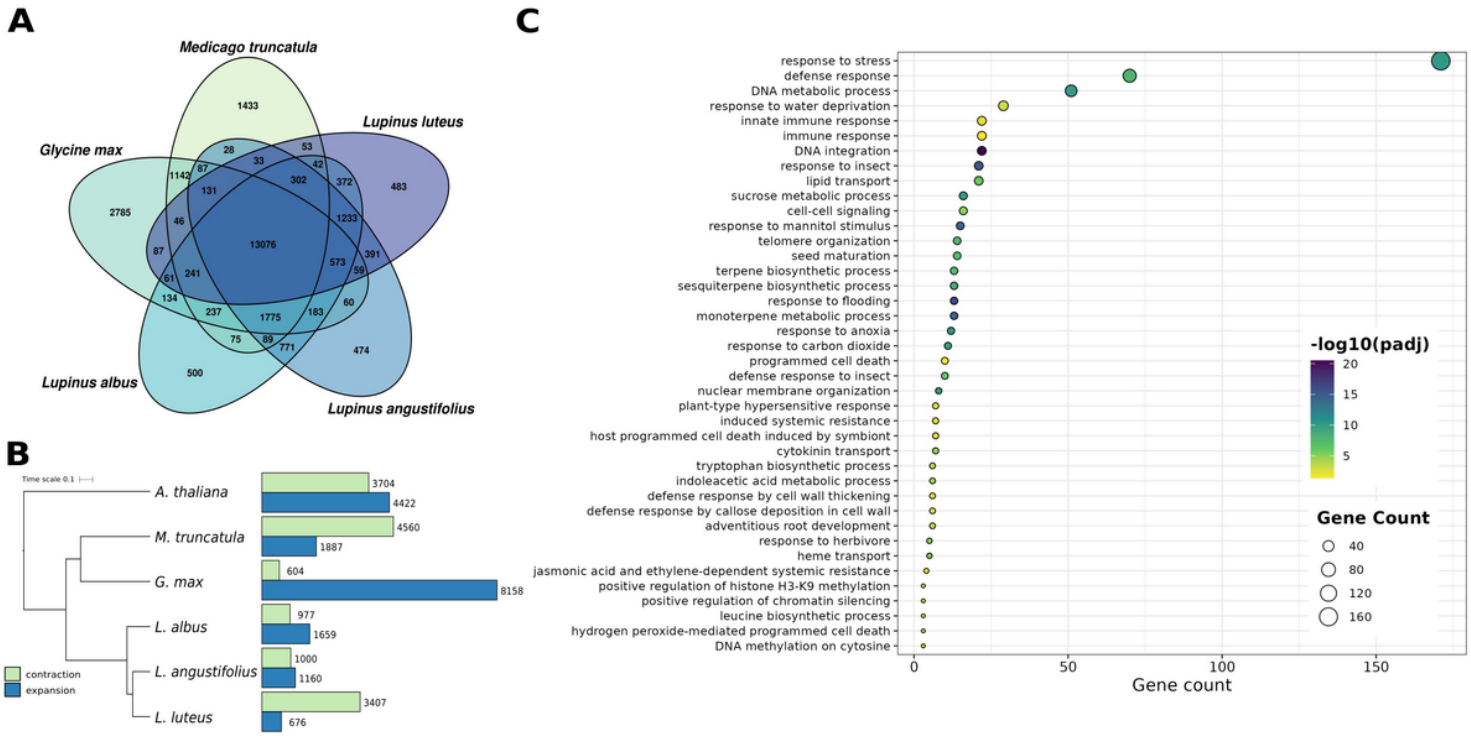

Figure 2

Comparative analysis of gene families between *L. luteus* and related species. A) Venn diagram of the number of gene families unique and shared between *L. luteus* and *Glycine max*, *Lupinus albus*, *Lupinus angustifolius*, and *Medicago truncatula*. B) Phylogenetic analysis of 6 species using *A. thaliana* as an outgroup, and the profiles of expanded and contracted gene families in each species. C) Gene ontology terms enriched biological processes in the set of expanded gene families in the *L. luteus* genome.

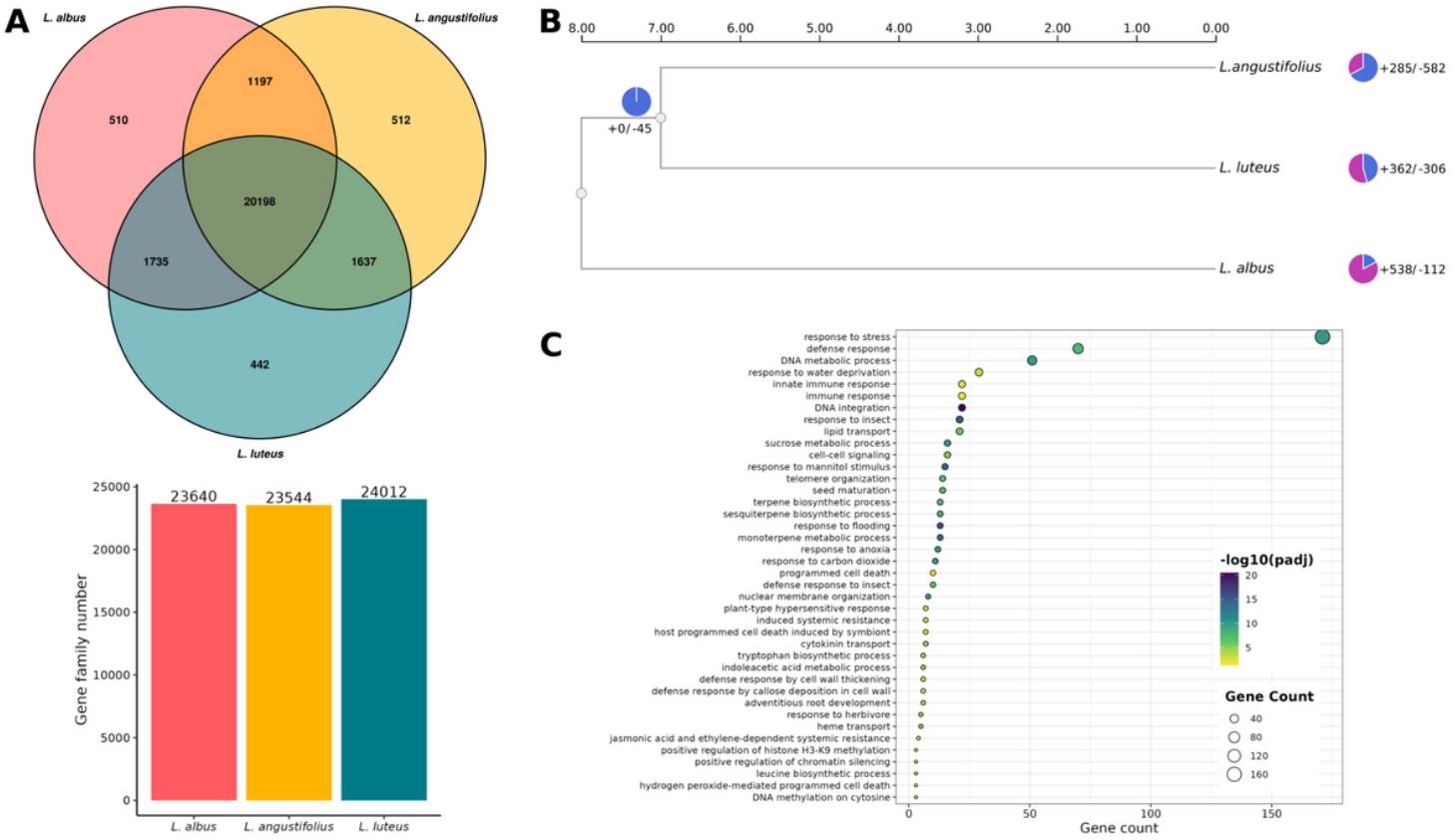

Figure 3

Comparative analysis of gene families between *Lupinus* species. A) Venn diagram of the number of gene families unique and shared between *L. luteus*, *Lupinus albus* and *Lupinus angustifolius*. B) Phylogenetic analysis of lupin species, and the profiles of expanded and contracted gene families in each species. In the Pie-chart significantly expanded families are represent in magenta and in blue contracted gene families. C) Gene ontology terms enriched biological processes in the set of expanded gene families in the *L. luteus* genome.

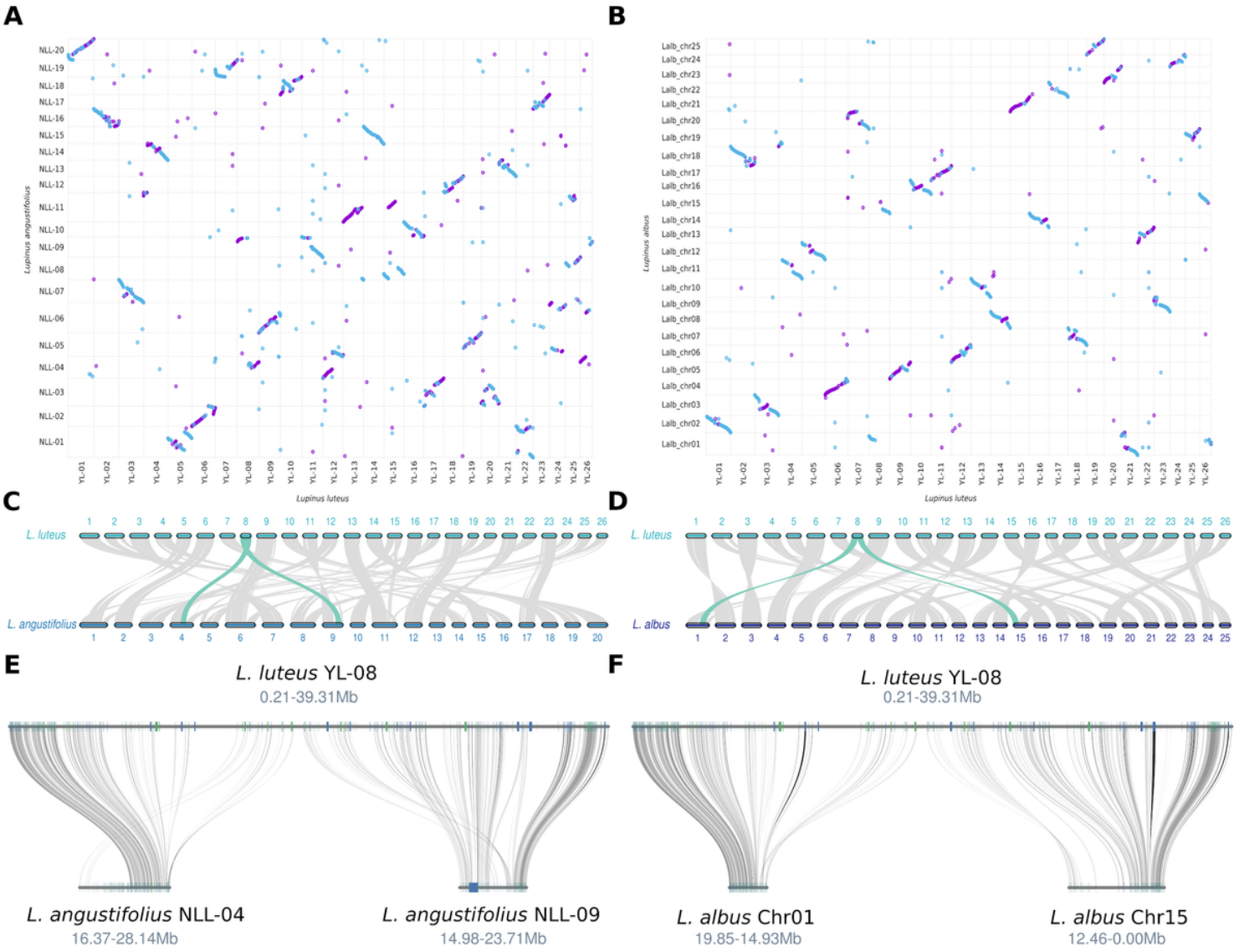

Figure 4

Synteny analysis of *L. luteus*, *L. albus*, and *L. angustifolius*. Synteny map of the *L. luteus* genome and the A) *L. angustifolius*, and B) *L. albus* genomes. Purple and blue dots indicate similar sequences in the same orientation and inversions, respectively. Syntenic blocks among *L. luteus* and C) *L. angustifolius*, and D) *L. albus*. Numbers represent the chromosome order. Each line represents one block. Green lines indicate the event of fusion which originates the chromosome 8 in the *L. luteus* genome. Local synteny between chromosome 8 of *L. luteus* (YL-08) and its homologous chromosome in E) *L. angustifolius* and F) *L. albus*

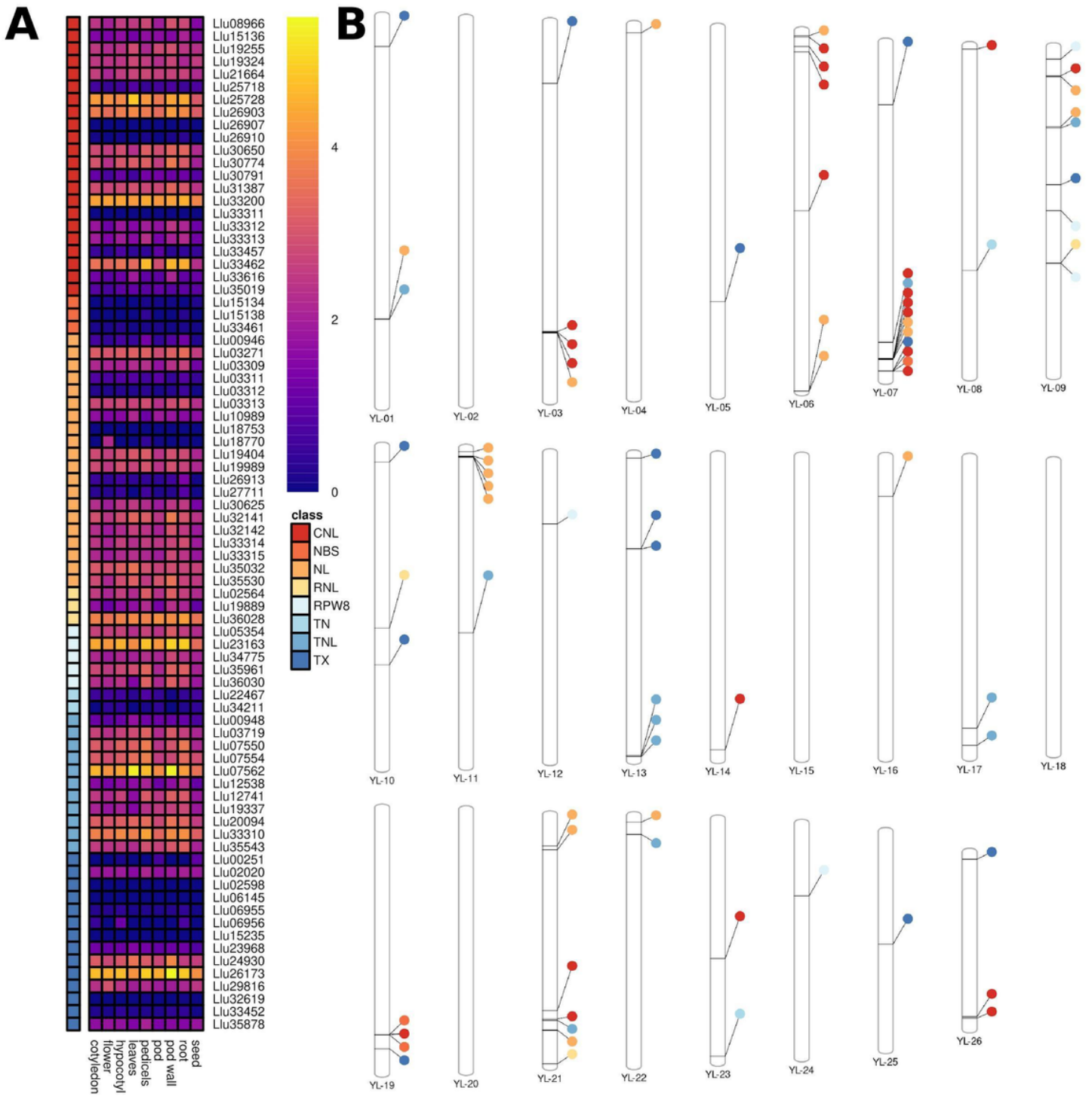

Figure 5

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 genes.

Supplementary Files

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

- supplementaryfiguresLluteusgenomeAssembly.docx
- SupplementaryTablesLluteusGenomeAssembly.xlsx