

Multiomic analysis of genes related to oil traits in legumes provide insights into lipid metabolism and oil richness in soybean

Dayana K. Turquetti-Moraes¹, Cláudio Benício Cardoso-Silva^{1,2}, Fabricio Almeida-Silva^{1,3,4}, Thiago M. Venancio^{1*}

¹ Laboratório de Química e Função de Proteínas e Peptídeos, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacazes, RJ, Brazil.

² Laboratório Nacional de Ciência e Tecnologia do Bioetanol, Centro Nacional de Pesquisa em Energia e Materiais, Universidade de Campinas, São Paulo, SP, Brazil.

³ Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium.

⁴ VIB Center for Plant Systems Biology, VIB, 9052 Ghent, Belgium.

* TMV: Laboratório de Química e Função de Proteínas e Peptídeos, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro. Av. Alberto Lamego 2000, P5, sala 217, Campos dos Goytacazes, RJ, Brazil. Email: thiago.venancio@gmail.com.

Abstract

Soybean (*Glycine max*) and common bean (*Phaseolus vulgaris*) diverged approximately 19 million years ago. While these species share a whole-genome duplication (WGD), the *Glycine* lineage experienced a second, independent WGD. Despite the significance of these WGDs, their impact on gene families related to oil-traits remains poorly understood. Here, we report an in-depth investigation of oil-related gene families in soybean, common bean, and twenty-eight other legume species. We adopted a systematic approach that included transcriptome and co-expression analysis, identification of orthologous groups, gene duplication modes and evolutionary rates, and family expansions and contractions. We curated a list of oil candidate genes and found that 91.5% of the families containing these genes expanded in soybean in comparison to common bean. Notably, we observed an expansion of triacylglycerol (TAG) biosynthesis (~3:1) and an erosion of TAG degradation (~1.4:1) families in soybean in comparison to common bean. In addition, TAG degradation genes were two-fold more expressed in common bean than in soybean, suggesting that oil degradation is also important for the sharply contrasting seed oil contents in these species. We found 17 transcription factor hub genes that are likely regulators of lipid metabolism. Finally, we inferred expanded and contracted families and correlated these patterns with oil content found in different legume species. In summary, our results do not only shed light on the evolution of oil metabolism genes in soybean, but also present multifactorial evidence supporting the prioritization of candidates for crop improvement.

Key words: polyploidization, transcriptome, coexpression networks, Fabaceae.

1. Introduction

The Fabaceae family (legumes) is a large group of flowering plants, with around 19,500 species. This family is remarkably diverse and stands out as the second most economically important family of crop plants after grasses (Azani et al. 2017; Thorne 2002). The notable grain legumes — such as chickpea (*Cicer arietinum*), pea (*Pisum sativum*), peanut (*Arachis hypogaea*), common bean (*Phaseolus vulgaris*), and soybean (*Glycine max*) — play a critical role in human and animal nutrition, as well as in industrial applications. For example, peanut and soybean are rich in oil and protein, while pea and common bean are rich in starch and protein (Azizaram et al. 2021; Pattee et al. 1983; Shen, Hong, and Li 2022; Yao et al. 2020). Interestingly, oil content among legume species can vary dramatically, from ~2% to ~50% in common bean and peanut, respectively. Despite the differences in seed oil content, soybean and common bean are related crops that diverged approximately ~19 million years ago (mya) (Lavin, Herendeen, and Wojciechowski 2005; Stefanović et al. 2009). These species shared a whole-genome duplication event (WGD, also referred to as polyploidization) ~58 mya, while a second WGD (~13 mya) occurred in the common ancestor of the *Glycine* genus, making soybean a suitable model for investigating the effects of WGD on gene family evolution (Shoemaker, Schlueter, and Doyle 2006).

Gene and genome duplications have been extensively associated with plant adaptation and diversification (Van de Peer, Mizrahi, and Marchal 2017; Zhuang et al. 2022). Gene duplication may occur through mechanisms such as WGDs or small-scale duplications (SSD) (Ren et al. 2018; Flagel and Wendel 2009). WGDs also played a role in the domestication of plants that eventually became

modern crops, such as wheat and maize (Carretero-Paulet and Van de Peer 2020; Hake and Ross-Ibarra 2015; Rahman et al. 2020). WGDs are typically followed by genome fractionation and rearrangement, restoring bivalent chromosome pairing and disomic inheritance — a process also known as diploidization (Z. Li et al. 2021). Intriguingly, the prevalence of WGDs is significantly greater in plants than in other lineages of multicellular organisms (Panchy, Lehti-Shiu, and Shiu 2016). Gene duplicates that survive diploidization form families and often diverge at the sequence, epigenetic, and transcriptional levels (Freeling 2009), resulting in neofunctionalization or subfunctionalization. Neofunctionalization involves the acquisition of new functions, while subfunctionalization results in the division of the original function of the gene ancestor among its copies (Freeling 2009; Freeling, Scanlon, and Fowler 2015). Over time, gene families can gain or lose genes, generating a wealth of genetic material for adaptation (Cheng et al. 2018; Moharana and Venancio 2020; Renny-Byfield and Wendel 2014; Soltis et al. 2009). A comprehensive analysis of gene family evolution and expression data can contribute to the selection of candidates to improve oil-related traits. A remarkable example was the development of a lipoxygenase-free soybean, leading to improvements in the palatability of soybean oil and protein products. This advancement was achieved through the implementation of a pooled CRISPR-Cas9 system specifically targeting three soybean lipoxygenase genes from a set of 36 previously reported candidates (J. Wang et al. 2020; H. Song et al. 2016).

Although substantial efforts have been dedicated to the investigation of pivotal genes involved in oil content and quality (Borisjuk et al. 2005; B. Chen et al. 2020; Kanai et al. 2019; Li-Beisson et al. 2017; L. Lu et al. 2021; X. Lu et al. 2016; Manan et al. 2017; Marchive et al. 2014; Meinke, Chen, and Beachy 1981; Nguyen et al. 2016; Pham, Shannon, and Bilyeu 2012; Sandhu et al. 2007; Turquetti-Moraes et al. 2022), the evolution of these gene families and how WGDs shaped them remain largely unexplored. In the present study, we have undertaken an in-depth exploration of gene families associated with oil traits in legumes, with particular emphasis in soybean. We observed that the increase in soybean oil content was deeply impacted by the expansion of gene families shared with common bean. In addition, we hypothesize that most genes associated with lipid and fatty acid (FA) metabolism reverted to single copy after the ~58 mya WGD and duplicated at the ~13 mya WGD. In contrast, genes responsible for regulatory functions were often retained as duplicates in both species after the ~58 mya WGD and duplicated again in the ~13 mya WGD. Further, TAG degradation genes were two-fold more expressed in common bean than in soybean. Co-expression analysis uncovered 17 transcription factor (TF) hub genes that are strong candidate regulators of lipid metabolism. Finally, we inferred expanded and contracted orthologous groups and correlated these patterns with the oil contents found in different legume species. Our study expands the knowledge of several metabolic pathways, pinpoint key TFs, and show evidence for novel gene candidates involved in oil biosynthesis. Together, the results presented here also bear the potential to have practical applications by presenting the most promising targets to improve soybean oil content and quality according to the current landscape of genomics data.

2. Results and discussion

2.1 Soybean genes involved in oil traits belong to families shared with common bean

We selected 2,176 soybean genes related to oil traits as a reference to find candidate homologous gene families (Supplementary Table 1; see materials and methods). This set of genes, henceforth

called oil genes, are distributed along 567 soybean homologous families, out of which 562 contain at least one common bean homolog (Supplementary Table 2). Only five families do not have a common bean homolog (HOM05D015518, HOM05D006604, HOM05D031525, HOM05D130031, and HOM05D039847) and include some genes (e.g. Glyma.02G006100, Glyma.02G281500, Glyma.08G064400, Glyma.16G133700, and Glyma.20G068000) reported as candidates to improve oil quality in soybean (Niu et al. 2020) (Supplementary Table 1; Supplementary Figure 1). These results indicate that the basic genetic machinery responsible for oil accumulation in soybean was already present in its last common ancestor shared with common bean, strongly suggesting that oil richness was acquired via expansions or contractions of extant gene families, mutations, and changes in transcriptional and epigenetic regulation, which were at least partially influenced by anthropogenic processes such as domestication and breeding (Turquetti-Moraes et al. 2022; J. Wang et al. 2020; M. Zhang et al. 2022).

About 14.57% (7,706 of 52,872) of the soybean genes and 15.44% (4,236 of 27,433) of the common bean genes belong to oil homologous families (Supplementary Table 3). These genes are enriched in functional terms related to metabolism and regulation of gene expression in both species (Supplementary Table 4), particularly TFs (e.g. GmDof4: Glyma.17G081800, GmDof11: Glyma.13G329000, GmMYB73: Glyma.06G303100, GmDREBL: Glyma.12G103100); response to oxidative stress (e.g. peroxidases: Glyma.19G066200, Glyma.07G263000); and metabolism of lipid and FAs (e.g. phospholipase: Glyma.03G159000, desaturase: Glyma.13G038600). Interestingly, the overexpression of GmMYB73 promotes lipid accumulation in soybean and its ectopic expression with other TFs (GmDof4, GmDof11, and GmDREBL) increased seed size/weight in transgenic *Arabidopsis* (Duan et al. 2023; Y.-F. Liu et al. 2014). In addition, 259 and 42 GO or Interpro terms were enriched only in soybean and common bean, respectively. For example, in soybean, we observed enrichment of terms including lipid glycosylation, FA and lipid metabolic process, and TAG biosynthetic process. On the other hand, in common bean, we found enrichment for lipid transport, microtubule nucleation and polymerization, and protein domains of lipoxygenases (Supplementary Table 4).

We investigated changes in sizes of the 562 gene families containing soybean and common bean genes (Figure 1). Using soybean as a reference, 2.1% (12), 91.5% (514), and 6.4% (36) of the families lost, gained, and had their sizes unchanged (i.e. neutral), respectively (Supplementary Table 2). Loss families were enriched in lipid metabolic process, especially to protein domains related to phospholipase/lysophospholipase. Neutral families were enriched in primary metabolic process, proteolysis, phospholipid biosynthetic processes, lipoate metabolic processes, lipid transport, and some enzyme domains or subunits (e.g. biotinyl protein ligase, FA desaturase, and seed storage helical domain) (Supplementary table 5). In gain families, by far the major group, we identified two scenarios: 25.5% (131 families) exhibited one common bean gene to two or more in soybean (1:2+), while 74.5% (383 families) had at least two common bean genes to three or more soybean homologs (2:3+). In summary, 1:2+ families were enriched in lipid and FA metabolism genes, while those families with duplications in common bean and new duplications in soybean (2:3+) were enriched in regulatory processes (e.g. gene expression and RNA metabolism) and response to stress (Supplementary Table 2; Supplementary Table 5). The 1:2+ families showed one peak in lower K_s values, while the 2:3+ families had two peaks. The peaks in both distributions correspond to the expected WGD K_s ages (Figure 1). Hence, we propose that most genes associated with lipid and FA metabolism (enriched in 1:2+ families) reverted to single copy after the ~58 mya WGD and duplicated

at the ~13 mya WGD. In contrast, genes responsible for regulatory functions (enriched in 2:3+ families) were often retained as duplicates in both species after the ~58 mya WGD and duplicated again in the ~13 mya WGD. Although one cannot rule out the possibility of extreme sequence divergence and neofunctionalization, these results corroborate our hypothesis that the increase in oil content in soybean was deeply impacted by gene families shared with common bean that independently expanded in the *Glycine* WGD event.

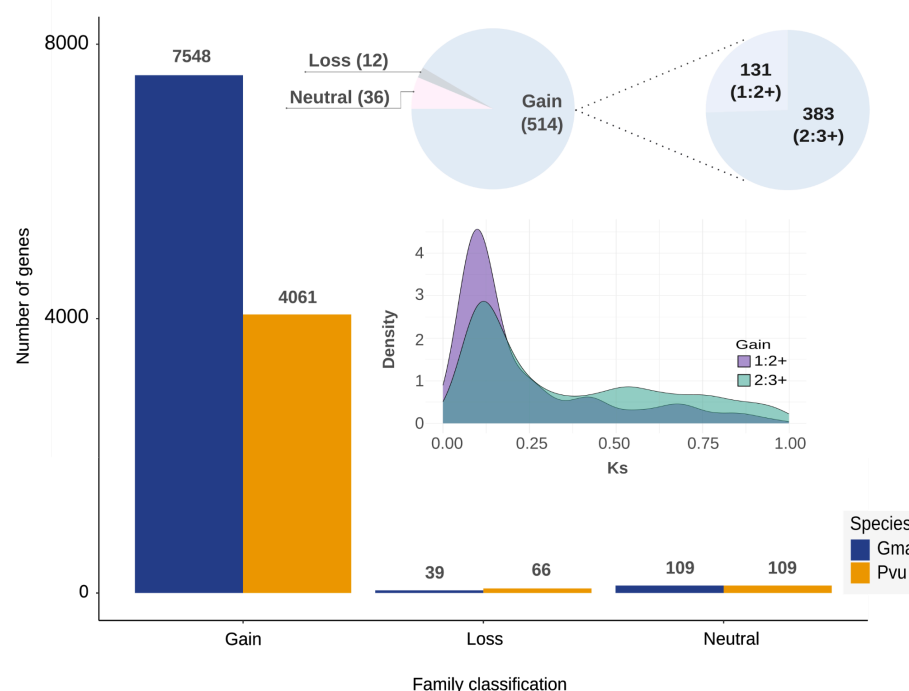


Figure 1. Number of soybean and common bean oil-related genes in different relative family size categories. Inset: number of families in each category. The two categories of gain families (1:2+ and 2:3+) are highlighted, alongside their respective K_s density plots. **Gain:** families with more genes in soybean (Gma) than in common bean (Pvu). **Loss:** families with less genes in Gma than in Pvu. **Neutral:** families with the same number of genes in both species.

2.2 Candidate oil families and expression patterns of genes from triacylglycerol (TAG) pathways

In plants, TAGs can be synthesized by two distinct routes. The classical Kennedy pathway involves the sequential acylation of glycerol-3-phosphate (G3P) (Figure 2.A). G3P is activated through acylation by acyl-CoA:glycerol-3-phosphate acyltransferase (GPAT), leading to the formation of lysophosphatidic acid (LPA); another acyl group is added to LPA by acyl-CoA:lysophosphatidic acid acyltransferase (LPAAT), forming phosphatidic acid (PA). PA is then dephosphorylated by PA phosphatase (PAP), leading to the formation of diacylglycerol (DAG). Finally, acyl-CoA:diacylglycerol acyltransferase (DGAT) adds the third acyl group to DAG, forming TAG. Alternatively, TAGs can be synthesized through the complex pathway, in which DAG originates from preexisting membrane lipids such as phosphatidylcholine (PC) through different pathways (Figure 2.A) (Bates 2016; Bates et al. 2009; Bates, Stymne, and Ohlrogge 2013). Over 90% of the acyl chains esterified to the glycerol backbone in developing soybean embryos originate from the complex pathway (Bates et al. 2009; Bates and Browse 2012). Hence, manipulating FA composition and seed oil content in soybean requires a deep

understanding of these pathways. Furthermore, given its paleopolyploid genome, pinpointing the paralogs that are truly involved in seed lipid metabolism imposes an additional layer of complexity.

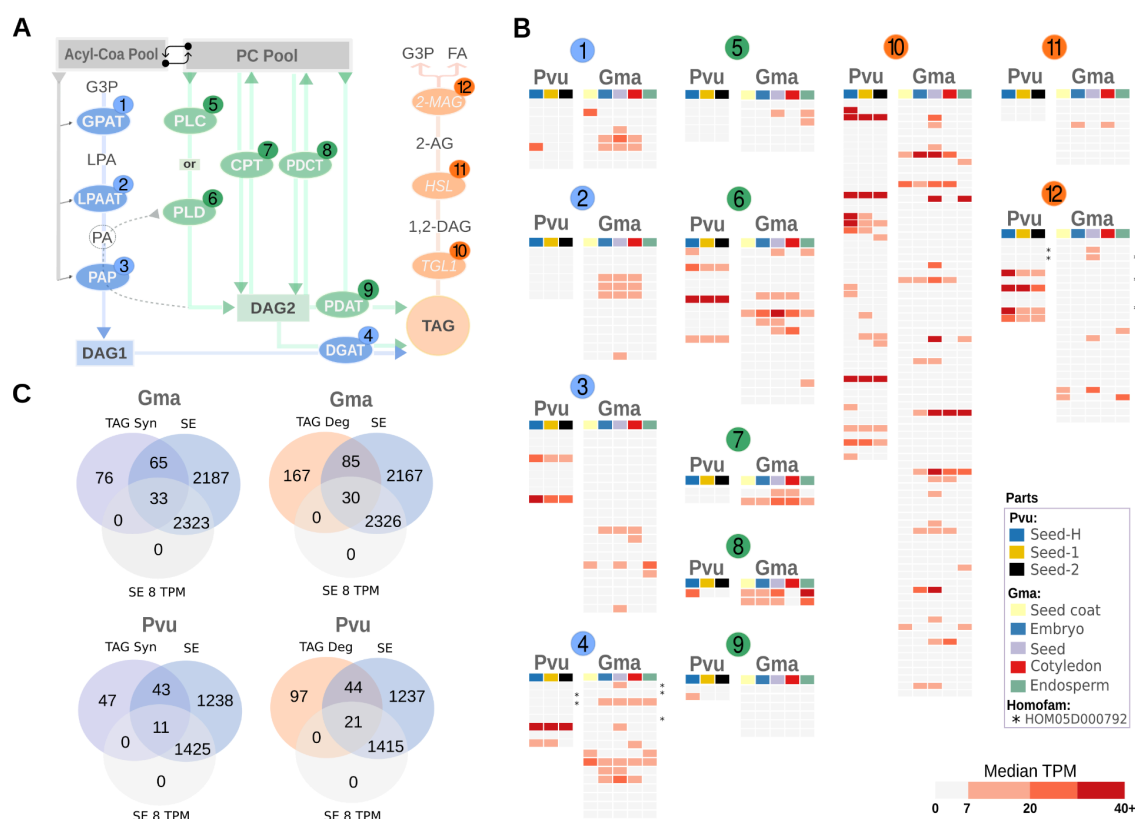


Figure 2. Major enzymes involved in DAG/TAG synthesis and degradation. **A.** Reactions involved in TAG degradation and formation from FAs. Blue: Kennedy pathway (gray arrows indicate the source of FA to feed the pathway). Green: complex pathway; the dashed arrow indicates a reaction PLD-PAP to form DAG from PC; the arrows from CPT and PDCT indicate a reversible reaction for the formation of PC-derived DAG. Orange: TAG degradation pathway. **B.** Expression of soybean and common bean genes encoding enzymes in each step of the pathways from panel A. Asterisks (*) denote unannotated genes from the HOM05D000792 family, which comprises genes associated with TAG synthesis and degradation pathways. **C.** Venn diagram of gene counts in TAG homologous families and their seed expression in Gma and Pvu. Abbreviations: **G3P**: glycerol-3-phosphate; **GPAT**: acyl-CoA glycerol-3-phosphate acyltransferase; **LPA**: lysophosphatidic acid; **LPAAT**: acyl-CoA lysophosphatidic acid acyltransferase; **PA**: phosphatidic acid; **PAP**: phosphatidic acid phosphatase; **DAG**: diacylglycerol; **DAG1**: DAG from the Kennedy pathway; **DAG2**: PC-derived DAG; **DGAT**: acyl-CoA diacylglycerol acyltransferase; **TAG**, triacylglycerol; **PC Pool**: phosphatidylcholine pool; **PLC**: phospholipase C; **PLD**: phospholipase D; **CPT**: cytidine diphosphate-choline diacylglycerol cholinephosphotransferase; **PDCT**: phosphatidylcholine diacylglycerol cholinephosphotransferase; **PDAT**: phospholipid diacylglycerol acyltransferase; **TGL1**: triacylglycerol lipase; **HSL**: hormone-sensitive lipase; **2-MAG**: 2-monoacylglycerol acylhydrolase; **FA**: fatty acid; **Pvu**: common bean; **Gma**: soybean; **TPM**: transcripts per million; **Seed-H**: heart stage seeds (~7 mg); **Seed-1**: state 1 seeds (~50 mg); **Seed-2**: stage 2 seeds (~150 mg). **TAG Syn**: TAG synthesis pathway; **TAG Deg**: TAG degradation pathway; **SE**: seed expression (≥ 1 TPM); **SE 8 TPM**: seed expression (≥ 8 TPM). The genes in this heatmap can be found in Supplementary Table 7 and Supplementary Table 8. Venn diagram was generated using Venn diagram (<https://bioinformatics.psb.ugent.be/webtools/Venn/>).

By meticulously curating homologous families from SoyCyc TAG pathway genes (soybase.org), we found 352 TAG genes in soybean (Supplementary Table 6), which are distributed in 43 families. These families include 204 and 116 genes with seed expression (TPM > 1) in soybean and common bean, respectively (Figure 2.B, Supplementary Table 7, Supplementary Table 8). We found one family, HOM05D000792, comprising genes belonging to TAG synthesis (e.g. DGAT: Glyma.16G051200, Glyma.16G051300) and degradation (e.g. 2-MAG: Glyma.03G243700, Glyma.19G241200). Other five HOM05D000792 members that are expressed in seeds encode WD repeat-containing proteins with unknown functions (e.g. Glyma.19G241800, Glyma.03G244500, Glyma.10G159000, Phvul.006G097800, and Phvul.006G098200). We kept these five genes in both TAG biosynthesis and degradation pathways (asterisk marks in Figure 2.B, Supplementary Table 7 and Supplementary Table 8).

The families encoding enzymes from each step of the TAG pathways presented at least one gene with expression of ~10 TPM in soybean seeds, except for phospholipid diacylglycerol acyltransferase (PDAT) (Figure 2.B). Thus, in order to select the most promising candidate genes, we adopted a threshold of 8 TPM (Supplementary Figure 2), which allowed us to find 33 and 11 (~3:1) TAG synthesis and 30 and 21 (~1.4:1) TAG degradation genes expressed in soybean and common bean seeds, respectively (Figure 2.C, Supplementary Table 7 and Supplementary Table 8). These results suggest both a relative expansion in the TAG biosynthesis and an erosion of TAG degradation components in soybean when compared to common bean, even if we consider the 2:1 ratio expected because of the ~13 mya WGD. We also analyzed the expression of these genes in all plant parts (Supplementary Figure 3, Supplementary Figure 4). Four GmGPATs and only one PvGPAT had at least 8 TPM in seeds (Glyma.07G069700, Glyma.05G131100, Glyma.08G085800, Glyma.09G119200, Phvul.010G099700) (Figure 2.B). Liu et al. showed that only GmGPAT9-2 (Glyma.09G119200) out of sixteen tested GmGPATs with high acyltransferase activity may not play a direct role in TAG formation. However, they found that seed-specific expression of GmGPAT9-2 in *Arabidopsis* increased the proportion of arachidic acid (C20:0) and erucic acid (C22:1) without an increase in the total oil content (H. Liu et al. 2022). Except for Glyma.08G085800, the GmGPATs mentioned above were corroborated by Liu et al. However, we found that this gene is highly expressed in seeds, specially in embryo and cotyledons (Figure 2.B, Supplementary Table 7), supporting its role in TAG synthesis.

In a second group of transferases, LPAATs regulate the synthesis of PA, an intermediate in the formation of membrane, signaling, and storage lipids (Kim and Wang 2020). From the four GmLPAATs reported here (Figure 2.B, Supplementary Table 7), Glyma.10G095500 was highly expressed in cotyledons, corroborating previous studies (X. Wang et al. 2019). Glyma.02G181300 and Glyma.12G163500 are within QTL associated to phosphatidylcholine (qPC-2.1) and phosphatidylinositol (qPI-12.1), respectively (Anshu et al. 2022) (Supplementary Table 7). Finally, DGAT forms an ester linkage between a fatty acyl-CoA and the DAG free hydroxyl group (G. Chen et al. 2022). From a set of 26 recently studied GmDGATs (S. Zhao et al. 2023), we found six showing high expression in seeds (Glyma.16G051200, Glyma.01G156000, Glyma.16G115700, Glyma.13G118300, Glyma.09G065300, and Glyma.13G106100). All these GmDGATs appear to be involved in TAG assembly, although DGAT1s such as GmDGAT1A (Glyma.13G106100) and GmDGAT1C (Glyma.09G065300) influence oil content and quality more prominently (Torabi et al. 2021; J. Zhao et al. 2019).

The complex pathway for DAG/TAG synthesis (Figure 2.A) is influenced by phospholipase C (PLC), phospholipase D (PLD) associated with PAPs (PA phosphatase), PDAT, and CPT or phosphatidylcholine diacylglycerol cholinephosphotransferase (PDCT). PLC hydrolyzes PC to DAG and phosphocholine, while PLD hydrolyses the choline head-group of PC and forms PA, an intermediate to the synthesis of other phospholipids. Subsequently, PAP removes the phosphate head-group of PA, converting it to DAG (Figure 2.A) (Bates, Stymne, and Ohlrogge 2013; Lee et al. 2011; W. Yang et al. 2017; Wakelam 1998). In soybean seeds, PLD is generally more expressed than PLC (Figure 2.B), although the PLCs Glyma.04G196700 and Glyma.06G169100 may be relevant candidates due to their endosperm expression (Figure 2.B). From the six GmPLD genes analyzed here (Glyma.01G162100, Glyma.07G031100, Glyma.08G211700, Glyma.11G081500, Glyma.13G364900, Glyma.15G267200), PLD α 1 (Glyma.13G364900) is a promising target considering its high expression in cotyledons (~20 TPM). Further, suppression of PLD α in soybean results in decreased levels of polyunsaturated FAs in TAG (Lee et al. 2011), suggesting a flow from PC to TAG via PLD-PAP without PDCT. In addition, GmPLD α 3 is also highly expressed in seeds (Glyma.08G211700: cotyledon expression of ~28 TPM) and was reported as associated with malate (J.-Y. Liu, Li, et al. 2020). Considering the expression of GmPAPs (Figure 2, Supplementary Figure 3) we suggest Glyma.10G270000 as a promising candidate to integrate this route.

Other potential routes for DAG2/TAG formation involves the reverse activity (i.e. PC to DAG) of PDCT or amino alcohol phosphotransferase (AAPT) also known as cytidine diphosphate-choline:diacylglycerol cholinephosphotransferase (CPT). These enzymes synthesize unique DAG and PC molecular species. PDCT catalyzes the interconversion between PC and DAG, contributing to the enrichment of polyunsaturated FAs in TAGs. The conversion of PC to DAG varies even between closely related species. PDCT transfers ~40% of oleic acid from PC to DAG in *Arabidopsis* against ~18.2% in canola (S. Bai et al. 2020; C. Lu et al. 2009). We propose the GmPDCTs Glyma.07G029800 and Glyma.08G213100 as interesting candidates because of their high expression in seeds (Supplementary Table 7). CPTs GmAAPT1 (Glyma.02G128300) and GmAAPT2 (Glyma.12G081900) were recently reported as crucial enzymes in TAG metabolism (Y. Bai et al. 2021), which is in line with the high GmAAPT2 expression in seeds. Finally, in addition to DGAT, PDAT is also involved in TAG assembly. GmPDAT (Glyma.13G108100) catalyzes the transfer of a FA moiety from the sn-2 position of a phospholipid to the sn-3-position of sn-1,2-DAG, forming TAG and a lysophospholipid (Pan, Peng, and Weselake 2015). PDATs often have contrasting expression profiles in different plant species (Torabi et al. 2021; Pan, Peng, and Weselake 2015). For example, PDAT was more expressed in plants that accumulate epoxy and hydroxy FAs (e.g. *Vernonia galamensis* and *Erysimum lagascae*) than in soybean and *Arabidopsis* (R. Li, Yu, and Hildebrand 2010). GmPDAT has been reported as associated with acyl-lipid metabolism and likely interacts with GmDGAT1 (Xu et al. 2018), although further studies are warranted to better understand this interaction and its roles in TAG synthesis (J.-Y. Liu, Zhang, et al. 2020).

TAG degradation during seed development is also important for the oil content of mature seeds (Ding et al. 2019). Strikingly, TAG degradation genes exhibited significantly higher expression levels in common bean (mean ~47 TPM) than in soybean (mean ~19 TPM). Furthermore, these genes exhibit high expression levels even in the early stages of seed development. We have identified TAG degradation genes that could be tested in soybean, out of which we highlight GmTGL1s (Glyma.01G067200, Glyma.02G043300, Glyma.06G294900, Glyma.04G255500, and

Glyma.09G233900) (Supplementary Figure 3). In common bean, certain TGL1 genes, such as Phvul.001G168200, Phvul.011G003600, Phvul.011G088700, Phvul.005G114300, and Phvul.010G098300 were highly expressed in seeds (Supplementary Figure 4). Notably, Phvul.011G003600 demonstrated consistently high expression across all seed stages. These findings suggest that the high expression of TAG degradation genes is a key factor to the lower oil accumulation in common bean. In conclusion, achieving high TAG levels in seeds requires an intricate system involving reduced TAG degradation, increased de novo FA synthesis, increased DAG production, and a flow from PC to combine saturated and unsaturated chains. Investigating the concerted action of these genes is key for enhancing oil content and quality in soybean seeds.

2.3 Identification of oil candidate genes with high expression in seeds

In addition to the detailed TAG pathways reported above, we applied the same threshold of 8 TPM in seeds to mine new genes potentially involved in oil metabolism. Approximately 50.7% (17,086/33,684) and 50.2% (8,941/17,805) of the genes expressed in seeds or seed subregions (i.e., endosperm, cotyledon, embryo or seed coat) showed at least 8 TPM in soybean and common bean, respectively (Supplementary Table 9, Supplementary Table 10). Expectedly, these genes are enriched in the metabolism of various molecules such as nitrogen, peptide, amino acid, ribose phosphate, nucleotide, ATP, as well as in translational elongation, protein transport, and gene expression. Out of the 17,086 genes highly expressed in soybean seeds, 562 were seed-specific and are enriched in defense responses and negative regulation of protein metabolic processes (e.g. proteolysis and peptidase activity), essential during seed development and other processes (Santamaría et al. 2014) (Supplementary Table 11). Considering only genes from candidate oil homologous families, we found 1.64 times more soybean (2,356) than common bean (1,436) genes expressed in seeds (Supplementary Table 9, Supplementary Table 10). These genes are distributed in 510 homologous families that can be classified in three categories: I. 402 families with at least one member expressed in soybean and common bean; II. 90 families with members expressed in soybean but not in common bean and; III. 18 families with members expressed in common bean but not in soybean (Supplementary Table 12).

The 402 homologous families from category I comprise 2,198 and 1,413 genes expressed in soybean and common bean, respectively. These families are enriched in a myriad of functions (Supplementary Table 11) and also encompass strong candidates for oil accumulation, such as GmSEIPIN1A (Glyma.09G250400), involved in TAG accumulation and lipid droplet assembly, maintenance, and proliferation (Pyc et al. 2021; Qi et al. 2023; Taurino et al. 2018).

The second category comprises 90 families containing genes expressed in soybean but not in common bean seeds. These genes emerge as promising candidates to account for the contrasting lipid contents found in soybean and common bean. Interestingly, these families are enriched in various lipid metabolism processes (e.g. phospholipid/glycerol acyltransferase and CDP-alcohol phosphatidyltransferase) (Supplementary Table 11). Although most of these genes are expressed in different plant parts, 24 are predominantly expressed in seeds and seed subregions, including lipases (Glyma.03G256900 and Glyma.19G255100), Zinc-finger-CW domain proteins (Glyma.14G204400 and Glyma.18G052100), and a sterol dehydrogenase (Glyma.06G058200) (Supplementary Figure 5). In addition, these families also comprise poorly characterized genes (HOM05D015518: Glyma.20G068000, HOM05D025622: Glyma.05G011200, HOM05D031525: Glyma.08G064400, HOM05D039847: Glyma.02G281500, HOM05D130031: Glyma.02G006100) reported as candidates to

improve oil content and quality (Niu et al. 2020; Fang et al. 2017) (Supplementary Figure 5). Together, these results support the recruitment of multiple genes to soybean seed metabolism after the split between soybean and common bean.

The third category comprises 18 homologous families with genes expressed in common bean but not in soybean seeds (Supplementary Figure 6). For example, the TFs Phvul.007G171333, Phvul.006G179900, and Phvul.009G149200 are highly expressed in the heart seed stage. The phosphatidylinositol PLC (Phvul.009G025300), associated with plant immunity (Tasma et al. 2008), also appears to play roles throughout seed development. In general, genes from this category suggest a deviation from TAG storage (Supplementary Figure 6). For example, Phvul.002G054200 encodes an enzyme with domains associated with wax ester synthesis and acyl group transfer, supporting a role in lipid modification instead of FA storage. These genes may drive FA metabolism towards structural lipid modification or secondary metabolism rather than storage.

About 3.5% (83) of the soybean oil candidate genes showed seed specific expression (Supplementary Table 9). In order to have a more conservative estimate, we retrieved the tissue specificity metrics from a broader and more heterogeneous set of samples available at the Soybean Expression Atlas (Almeida-Silva, Pedrosa-Silva, and Venancio 2023). These 83 genes were enriched in protein domains such as GDSL lipase/esterase and MADS-box TFs (Figure 3, Supplementary Table 11). GDSL-type esterase/lipase proteins belong to the SGNH hydrolase superfamily that can hydrolyze various substrates, including thioesters, acyl esters, phospholipids, and amino acids (Akoh et al. 2004; Su et al. 2020). Out of the 194 previously identified GmGDSL-type genes (Su et al. 2020), eight were found here as seed-specific and are strong candidates to improve oil content (Figure 3). MADS-box TFs are widely known for their roles in flowering, growth, and development (Shu et al. 2013; Zeng et al. 2018). Thirteen GmMADS-box TFs were preferentially expressed in seeds (Figure 3). Studies in *Arabidopsis* and oil palm (*Elaeis guineensis* Jacq.) support MADS-box TFs as important regulators of lipid metabolism and responsible for a decreased accumulation of polyunsaturated fatty acids (S.-Y. Li et al. 2020; Sun et al. 2020). GDSL and MADS-box genes account for 25.3% of the soybean seed-specific genes with at least 8 TPM. Interestingly, the MADS-box Glyma.04G257100 clusters with the sucrose/hexose transporter GmSWEET24/GmSWEET10b (Glyma.08G183500) (Figure 3), known to influence the distribution of sugars from the seed coat to the embryo and playing a crucial role in key soybean seed traits, such as size, oil and protein contents (S. Wang et al. 2020; S. Lu et al. 2022). We also found seed-specific expression of GmSWEET23 (Glyma.08G077200), GmSWEET4 (Glyma.04G198400), and GmSWEET39/GmSWEET10a (Glyma.15G049200), which were also associated with seed oil and protein content (Hooker et al. 2022; C. Liu et al. 2023; H. Yang et al. 2019; Boyang, Wenlong, and Caiying 2023). Finally, we highlight the seed-specific expression of the gibberellin 20-oxidases (GmGA20OX) Glyma.03G019800 and Glyma.07G081700 (Figure 3). GmGA20OX overexpression in *A. thaliana* has been shown to increase seed weight and oil (X. Lu et al. 2016). In this context, it is important to mention that seed size and weight have been related to oil content and several genes associated with these traits have been identified in soybean (Alam et al. 2022; J. Li et al. 2019).

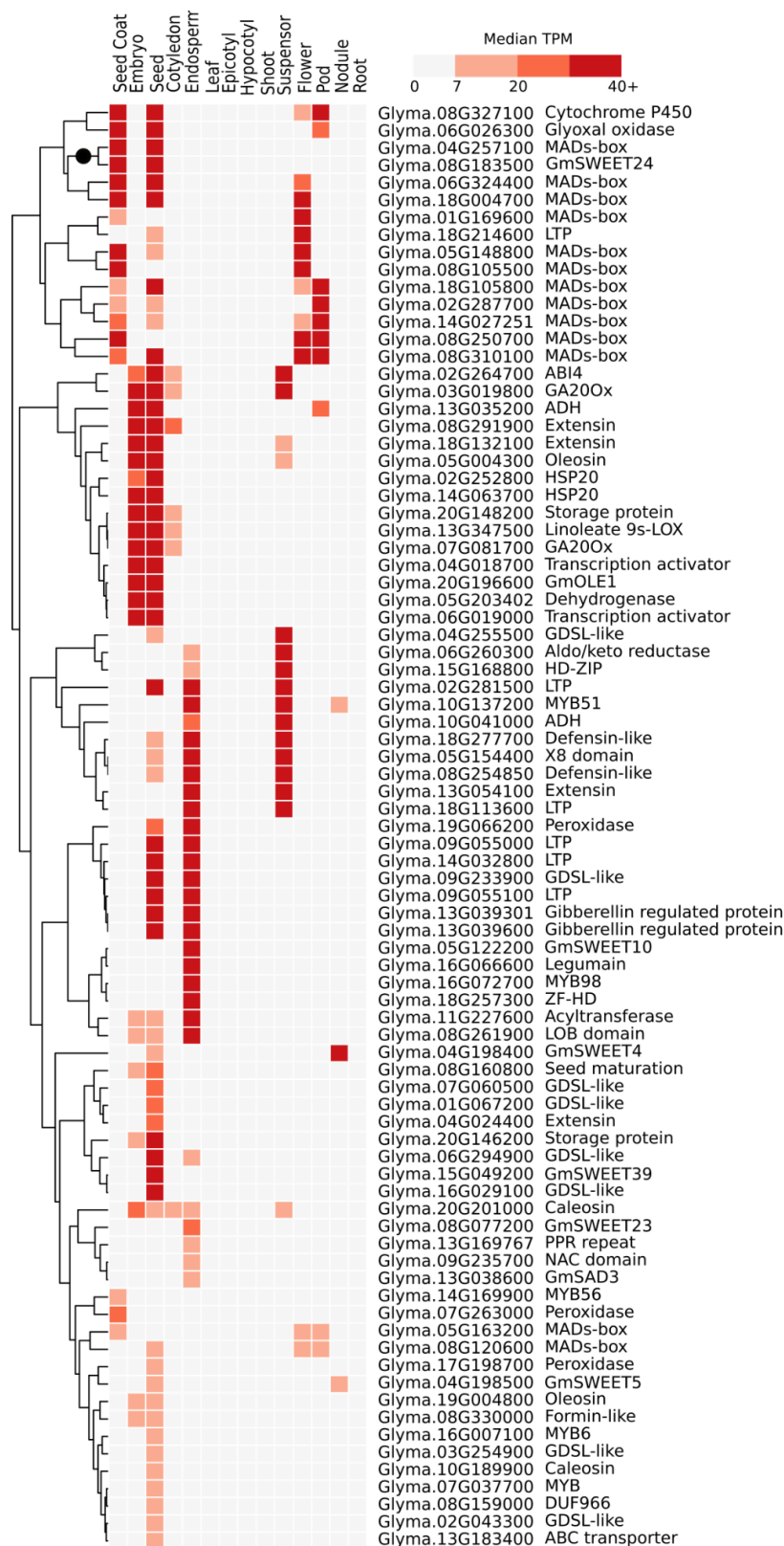


Figure 3. Expression of oil candidate genes with preferential expression in seeds. Tissue specificity metrics were obtained from the Soybean Expression Atlas (Almeida-Silva, Pedrosa-Silva, and Venancio 2023). Annotations were retrieved from Phytozome (V13) and Soybase.org. TPM: transcripts per million. Black circle marks the cluster with the GmSWEET24 and MADS-box genes discussed in the text.

2.4 Co-expression network of soybean genes with high seed expression from gain families

Aiming to better characterize novel seed oil genes, we used BioNERO (Almeida-Silva and Venancio 2022)(see methods for details) to compute a co-expression network (Almeida-Silva et al. 2020; Schaefer, Michno, and Myers 2017) of the gain-family genes with at least 8 TPM in seeds. Out of the 2,269 input genes, 92.9% (2,107) were distributed across 12 co-expression modules (Supplementary Figure 7). Except for the green module, we observed functional enrichment in all other modules: signal transduction and molecular transport, such as lipid transport (blue); cellular process, particularly photosynthesis and oil body stabilization (grey); cellular stress responses (greenyellow, brown); gene expression regulation (purple, magenta); multicellular organism development, cellular signaling and cytoskeletal dynamics. (tan, light green, dark red); protein processes/activity (light yellow); and lipid and FA metabolism (cyan); (Supplementary Table 13).

The cyan module comprises 33 TFs (Supplementary table 14) and important genes involved in seed oil quality and accumulation, such as GmFATA1A (Glyma.18G167300), GmFATA1B (Glyma.08G349200), GmDGAT1C (Glyma.09G065300), biotin carboxyl carrier protein (Glyma.18G243500, Glyma.09G248900), biotin carboxylase (Glyma.08G027600, Glyma.05G221100), ketoacyl-ACP synthase (KASI: Glyma.08G084300; KASII: Glyma.17G047000; KAS III: Glyma.09G277400), and Long-chain Acyl-CoA Synthetase (LACS: Glyma.06G112900, Glyma.13G079900, Glyma.20G060100) (Torabi et al. 2021; X. Wang et al. 2019). Glyma.07G110900 and Glyma.06G122600 encode a cytochrome P450 and an alcohol dehydrogenase, respectively. A previous study suggested that relative expression of these two genes promote the synthesis of linolenic acid in mature soybean seeds (X. Wang et al. 2019). In addition, we encountered four unannotated genes (Glyma.12G105300, Glyma.10G277900, Glyma.04G044200, and Glyma.05G141600) with potential activity in lipid metabolism (Supplementary Table 14).

Aiming to identify genes that may play significant roles in oil regulation and synthesis we analyzed modules that exhibit closest co-expression with the cyan module, i.e., darkred, green, and magenta (Supplementary Figure 8). We identified 15, 6 and 19 TF families containing 63, 8 and 52 genes distributed in darkred, green and magenta modules, respectively (Supplementary Table 14). Approximately 25% (31) of them (Glyma.02G016100, Glyma.02G274600, Glyma.02G303800, Glyma.04G010300, Glyma.04G044900, Glyma.04G050300, Glyma.05G098200, Glyma.05G140400, Glyma.05G175600, Glyma.06G079800, Glyma.08G132800, Glyma.08G360200, Glyma.09G241800, Glyma.10G016500, Glyma.11G242200, Glyma.12G040600, Glyma.13G153200, Glyma.13G202300, Glyma.14G041500, Glyma.14G071400, Glyma.14G205600, Glyma.16G011200, Glyma.16G012600, Glyma.16G152700, Glyma.17G096700, Glyma.17G132600, Glyma.17G157600, Glyma.17G174900, Glyma.18G014900, Glyma.19G022200, Glyma.20G200500) were previously reported as candidates for oil accumulation (Niu et al. 2020). Interestingly, within the magenta module, we found two NF-Y TFs (Glyma.12G236800 and Glyma.13G202300) as potential regulators of seed traits. Although NF-Y TFs are important for oil biosynthesis in *E. guineensis* (Yeap et al. 2017), their roles in soybean lipid metabolism remain unclear. Understanding the effects of gene silencing on TF regulation is also important. For example, we found five TFs (Glyma.01G081100, Glyma.04G010300, Glyma.14G112400, Glyma.17G096700; in darkred and Glyma.16G179900; in magenta) potentially regulated by DNA methylation during seed maturation (An et al. 2017).

We identified 189 hubs (Supplementary Table 14), of which 3.17% (6) are involved in TAG pathways (GmGPAT: Glyma.07G069700, GmPLD: Glyma.07G031100, PLD α 1: Glyma.13G364900, and TGL1: Glyma.14G050600, Glyma.20G127800). The WD-repeat Glyma.03G244500 discussed in section 2.2 was also found as a hub and showed a remarkable expression correlation (~ 0.8) with DGAT2A (Glyma.01G156000) (Supplementary Table 15). A recent study showed that DGAT2A enhances oil and linoleic acid contents in soybean seeds (Jing et al. 2021), highlighting the potential involvement of Glyma.03G244500 in TAG synthesis. In addition, we found three hubs encoding proteins of unknown function in the modules blue (Glyma.11G222000), green (Glyma.10G222300), and greenyellow (Glyma.13G092300), all connected to TFs and lipid metabolism genes (Supplementary Table 15). Interestingly, Glyma.13G092300 was previously reported as a candidate to improve oil quality (Niu et al. 2020). This gene is connected with MOTHER-OF-FT-AND-TFL1 (GmMFT: Glyma.05G244100; Supplementary Figure 9), which was proposed as a major gene of the classical QTL qOil-5-1 that regulates seed oil and protein content (Fang et al. 2017; J. Huang et al. 2020; Zhou et al. 2015; Cai et al. 2023). The connections of these three hubs with relevant genes to lipid metabolism and TFs support their importance in oil-related traits. Finally, we found 17 TF hub genes belonging to ten families, which might constitute bona-fide regulators of transcriptional programs involved in oil accumulation (Supplementary Table 14).

2.5 Conservation of oil candidate genes in legumes

We employed Orthofinder (Emms and Kelly 2019) to investigate the conservation of candidate genes across 30 legume species (Table 2, Supplementary Table 16). Approximately 94.7% (1,086,739/1,147,876) of the legume genes are distributed in 46,003 orthologous groups (OGs). Around 5.5% (63,594) of the genes are distributed in 12,249 species-specific OGs (Supplementary Table 16, Supplementary Table 17). *M. truncatula*, *P. sativum*, *L. japonicus*, *A. hypogaea*, and *S. tora* had the greatest frequencies of species specific OGs, while *A. hypogaea*, *A. ipaensis*, and *L. angustifolius* exhibited the highest frequencies of species-specific duplications (Supplementary Figure 10).

We used the 2,269 soybean gain-family genes with at least 8 TPM in seeds as references to find 1,104 candidate OGs containing 67,577 genes. In general, approximately 47.7% (527) of the OGs are shared by all species (Supplementary Table 18). Expectedly, these core legume genes are enriched in essential metabolic pathways such as phosphatidate metabolism, glycolysis, tricarboxylic citric acid (TCA) cycle, ureide biosynthesis, and other energy metabolism pathways. The remaining 52.3% (577) OGs are enriched in stress response signaling, antioxidant defense mechanisms, and synthesis of bioactive compounds with potential medicinal applications, such as divinyl ether, chlorogenic acid, and justicidin (Grechkin 2002; A. Gupta et al. 2022; Hemmati and Seradj 2016) (Supplementary Table 19).

Three OGs (OG0045944, OG0035121 and OG0018534) contained genes only from two species known for their high oil content, namely soybean and peanut. The genes in these OGs are annotated as amidases (OG0045944: Glyma.08G054300, Glyma.08G197900); and auxin response factor - ARF (OG003512: Glyma.07G134900, AVLU3S; and OG0018534: Glyma.12G153700, AQQ3N0, FQ0W8J, K35AK0, MJ6K6I, NZ22HG, RIMX1M, RZFM32, Y8ANLA, YS88LH, YTF73Z). Interestingly, auxin can alter FA content and composition in soybean and microalgae (W. Liu, Hildebrand, and Collins 1995; Jusoh et al. 2015). Thus, these ARF genes into OG003512 and OG0018534 may be candidates

to influence seed development and oil content. We also investigated the 1,104 OGs mentioned above for the presence of TAG pathway genes. We found 2,397 genes distributed across 41 OGs, out of which 19 are shared by all species (Supplementary Table 18). Among OGs not shared by all species, OG0027103 and OG0018208 contain fewer than 10 species. OG0027103, with HSL genes, is specific to *G. soja* and *G. max*; while OG0018208, with PLD genes, is specific to *A. hypogaea*, *C. fasciculata*, *G. soja*, *L. japonicus*, *M. truncatula*, *N. schottii*, *P. sativum* and *G. max*.

Table 2. Plant species used in this study, genome information and data source.

Species	Assembly	Genome size	Publication	Database
<i>Arachis duranensis</i>	V2.0	1.25 Gb	(Garg et al. 2022)	Legumepedia
<i>Arachis hypogaea</i>	V1.0	2.7 Gb	(Bertioli et al. 2016)	Phytozome 13
<i>Arachis ipaensis</i>	V1.0	1.56 Gb	(Q. Lu et al. 2018)	DRYAD
<i>Cajanus cajan</i>	V2.0	833 Mb	(Garg et al. 2022)	Legumepedia
<i>Cercis canadensis</i>	V1	342 Mb	(Griesmann et al. 2018)	GigaDB
<i>Chamaecrista fasciculata</i>	V1	429 Mb	(Griesmann et al. 2018)	GigaDB
<i>Cicer arietinum</i>	V2.0	738 Mb	(Garg et al. 2022)	Legumepedia
<i>Cicer reticulatum</i>	—	416 Mb	(S. Gupta et al. 2017)	NCBI
<i>Faidherbia albida</i>	—	654 Mb	(Chang et al. 2019)	GigaDB
<i>Glycine max</i>	V4.0	978 Mb	(Schmutz et al. 2010)	Phytozome 13
<i>Glycine soja</i>	V1.1	985 Mb	(Valliyodan et al. 2019)	Phytozome 13
<i>Lablab purpureus</i>	—	615 Mb	(Chang et al. 2019)	GigaDB
<i>Lotus japonicus</i>	V3.0	472 Mb	(Sato et al. 2008)	Kazusa
<i>Lupinus albus</i>	V1	450 Mb	(Hufnagel et al. 2020)	Phytozome 13
<i>Lupinus angustifolius</i>	V1.0	924 Mb	(Hane et al. 2017)	NCBI
<i>Medicago truncatula</i>	V4.0	411 Mb	(Tang et al. 2014)	Phytozome 13
<i>Mimosa pudica</i>	V1	557 Mb	(Griesmann et al. 2018)	GigaDB
<i>Nissolia schottii</i>	V1	466 Mb	(Griesmann et al. 2018)	GigaDB
<i>Phaseolus acutifolius</i>	V1.0	512 Mb	(Moghaddam et al. 2021)	Phytozome 13
<i>Phaseolus lunatus</i>	V1	546 Mb	—	Phytozome 13
<i>Phaseolus vulgaris</i>	V2.0	537 Mb	—	Phytozome 13
<i>Pisum sativum</i>	V1	4 Gb	(Kreplak et al. 2019)	URGI/INRA
<i>Senna tora</i>	V1	547 Mb	(S.-H. Kang et al. 2020)	NCBI
<i>Spatholobus suberectus</i>	V1	793 Mb	(Qin et al. 2019)	NCBI
<i>Trifolium pratense</i>	V2	345 Mb	(De Vega et al. 2015)	Phytozome 13
<i>Trifolium subterraneum</i>	V2.0	540 Mb	(Garg et al. 2022)	Legumepedia
<i>Vigna angularis</i>	V1.1	466 Mb	(K. Yang et al. 2015)	NCBI
<i>Vigna radiata</i>	V1	463 Mb	(Y. J. Kang et al. 2014)	NCBI
<i>Vigna subterranea</i>	—	535 Mb	(Chang et al. 2019)	GigaDB
<i>Vigna unguiculata</i>	V1.2	519 Mb	(Lonardi et al. 2019)	Phytozome 13

We found 46 OGs containing genes with seed-specific expression in soybean, of which 25 are shared by all species (Supplementary Table 20) and associated with regulatory processes (e.g.: metabolic, biosynthetic, and transcriptional), while the remaining 21 are enriched in lipoxygenases, divinyl ether biosynthesis, oleosin, SWEET transporters, among others (Supplementary Table 21). Some studies indicate that lipoxygenases in mature seeds produce conjugated unsaturated FA hydroperoxides, resulting in volatile compounds linked to the undesirable beany flavor (Rackis, Sessa, and Honig 1979; J. Wang et al. 2020). Other studies suggest that SWEET transporters significantly influence seed oil and protein contents (Duan et al. 2023; S. Wang et al. 2020). These results provide some genes that are likely involved in FA oxidation and nutrient uptake, which are important for seed quality and nutritional diversity found in legumes.

Finally, we employed CAFE5 (Mendes et al. 2020), a method based on gene birth (λ) and death (μ) rates, to investigate size changes in OGs. We ran CAFE only with the 1,104 OGs containing at least one of the 2,269 gain-family genes with a minimum expression of 8 TPM in seeds. From these OGs, 163 had significant contractions or expansions (p -value < 0.05, Figure 6). Out of these, 105 are shared by all species (Supplementary Table 18) and contain soybean genes related to lipid metabolism. Notable examples of these OGs include: OG0000430 (NADP-malic enzyme - NADP-ME: Glyma.13G354900; Glyma.15G019300), OG0000544 (GmbZIP123: Glyma.06G010200), OG0000840 (GmSWEET: Glyma.08G183500; Glyma.15G049200), OG0002123 (GmDGAT: Glyma.09G065300; Glyma.13G106100), and OG0000341 (containing the MADS box Glyma.04G257100, indicated here as a promising regulator of FA metabolism) (Torabi et al. 2021; S. Wang et al. 2020; Morley et al. 2023; Q.-X. Song et al. 2013). Conversely, among the remaining 58 OGs, we observed OG0006371 and OG0015023 containing genes from only 17 and 21 species, respectively. In OG0006371 we found Glyma.05G011200, an unannotated gene preferentially expressed in seed, flower, and nodule (Supplementary Figure 5). In OG0015023 we found Glyma.05G140300, a gene that encodes a small subunit of serine palmitoyltransferase-like (SPT-like) that is highly expressed in seeds. The SPT complex catalyzes the first and rate-limiting step in sphingolipid biosynthesis (M. Chen et al. 2006; P. Liu et al. 2023). In general, lipid metabolism genes related to adaptive responses including signaling and response biotic and abiotic stresses are found in OGs not shared by all species.

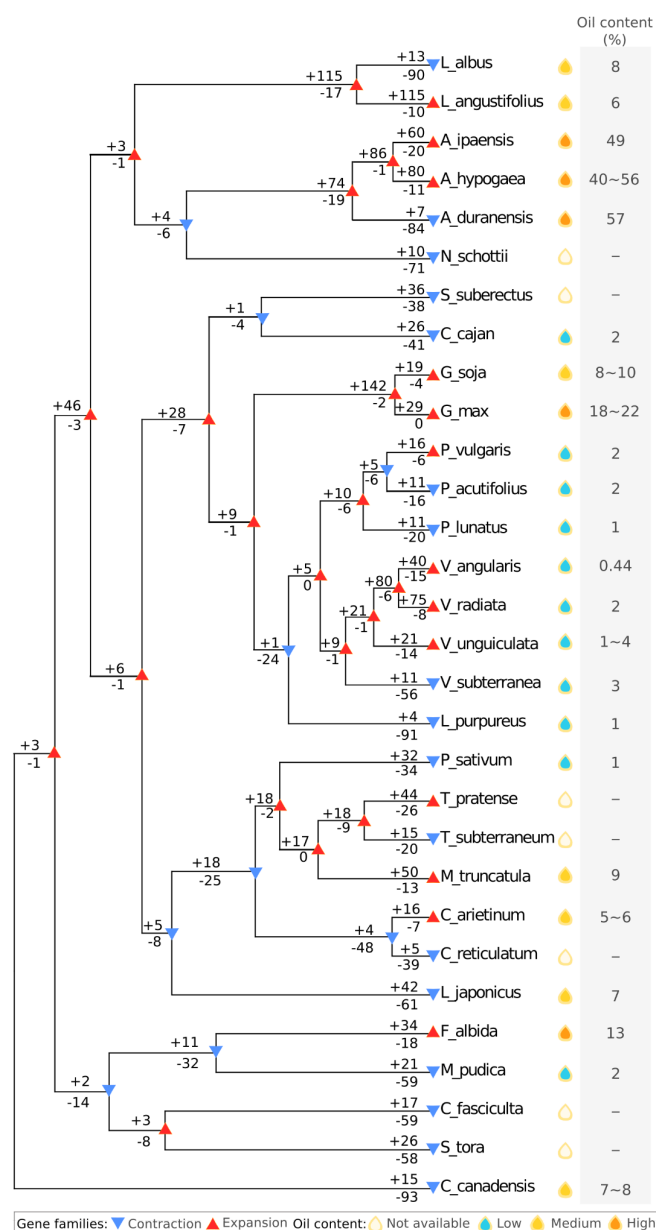


Figure 6. Species tree summarizing the number of orthogroups from gain families (~8 TPM) with significant (p-value < 0.05) expansion or contraction across 30 legume species. Gain families are those with more genes in soybean than in common bean. Red and blue triangles refer to nodes/leaves with more expansions and contractions, respectively. The drops represent seed oil content in each species: Low: below 4.9%; Medium: between 5% and 10%; High: more than 10%. Oil content source: *A. duranensis* (L. Huang et al. 2012); *A. hypogaea* (Shasidhar et al. 2017); *A. ipaensis* (Grosso, Nepote, and Guzmán 2000); *C. cajan* (Sharma, Nidhi, and Preeti 2011); *C. canadensis* (Duke and Ayensu 1985); *C. arietinum* (Zia-UI-Haq et al. 2007); *F. albida* (Hassan, Umar, and Yuguda 2007); *G. max* (Patil et al. 2018); *G. soja* (Patil et al. 2018); *L. purpureus* (Hossain et al. 2016); *L. japonicus* (Dam et al. 2009); *L. albus* (Bhardwaj, Hamama, and van Santen 2004); *L. angustifolius* (Lemus-Conejo et al. 2023); *M. truncatula* (Y. Song et al. 2017); *M. pudica* (Grygier et al. 2022); *P. acutifolius* (Bhardwaj, Hamama, and van Santen 2004); *P. lunatus* (Palupi et al. 2022); *P. vulgaris* (Sutivisedsak et al. 2011); *P. sativum* (Asen et al. 2023); *V. angularis* (Shweta and Katoch 2014); *V. radiata* (Zia-UI-Haq, Ahmad, and Iqbal 2008); *V. subterranea* (Minka and Bruneteau 2000); *V. unguiculata* (Perchuk et al. 2020). The CAFE plot was generated with cafeplotter (<https://github.com/moshi4/CafePlotter>).

3. Conclusion

The intricate history of plant WGD and the retention of multiple gene copies pose a considerable challenge in pinpointing the causative genes for specific traits. We employed a comprehensive approach for the investigation of gene families related to oil traits, using gene expression and co-expression, conservation and mutation rates. We identified soybean and common bean genes involved in TAG pathways, unveiled novel candidates and explored their expression and functional divergence in seeds, shedding light on their roles in lipid metabolism. Our findings do not only contribute to understanding the genetic mechanisms governing lipid metabolism, but also provide valuable leads for targeted genome editing for crop improvement and biotechnology (Figure 3, Supplementary Table 22).

4. Materials and methods

4.1 Selection of genes related to oil traits and enrichment analysis

We assembled a comprehensive dataset of soybean genes linked to oil traits from various sources: Aralip (McGlew et al. 2015), SoyCyc (v.9.0): diacylglycerol and triacylglycerol biosynthesis; and triacylglycerol degradation (Brown et al. 2021), and genes known for their involvement in lipid metabolism, obtained from the Mapman database, accessed via PLAZA Dicots 5.0 (Van Bel et al. 2022). These lists were supplemented with genes obtained through a systematic manual curation (Supplementary Table 1). This compilation resulted in a comprehensive collection of 2,176 soybean genes potentially associated with oil traits. The complete collection of homologous gene families were obtained from PLAZA 5.0, allowing the identification of 567 families containing potential oil genes. When considering the presence of homologs within these families, we expanded this set to 7,706 and 4,236 soybean and common bean candidate genes, respectively. Enrichment analyses for GO terms and conserved protein domains was performed in PhytoMine (<https://phytozome.jgi.doe.gov/phytomine/begin.do>), using Benjamini-Hochberg multiple testing correction (max p-value: 0.05) and the following background sets: genes in oil-candidate homologous families (Supplementary Table 4 and Supplementary Table 5); all genes expressed in seeds with ~1 TPM (Supplementary Table 11); gain-family genes with at least 8 TPM (Supplementary Table 13); all soybean protein-coding genes (Supplementary Table 19 and Supplementary Table 21).

4.2 RNA-seq data and gene expression analysis

We conducted a meticulous selection of RNA-Seq samples from the Soybean Expression Atlas (<https://soyatlas.venanciogroup.uenf.br/>) (Almeida-Silva, Pedrosa-Silva, and Venancio 2023). To ensure the selection of relevant samples, a filtering process was carried out using the following criteria to exclude samples that originated from: 1) indeterminate plant parts (e.g. whole plant, seedling, and unknown); 2) transgenic or mutant plants; 3) cultivars other than Williams 82 and; 4) specific treatments (e.g. exposure to biotic and abiotic stresses). Exceptions for criteria 2 and 3 above include samples from mutants and varieties that exhibited specific advantages in the context of oil biosynthesis (e.g. Seed_jack_GmZF351, Seed_jack_GmZF352, seed_Thorne_wt_r5_r6 and seed_gmOleo1). This systematic curation resulted in a list of 605 samples (Supplementary table 23). Gene expression estimates were retrieved in TPM. We used the median TPM to investigate the gene expression patterns across the diverse array of samples. We also retrieved the tissue specificity index

Tau (τ) (Kryuchkova-Mostacci and Robinson-Rechavi 2017; Yanai et al. 2005) available in the Soybean Expression Atlas (<https://soyAtlas.venanciogroup.uenf.br/>). Expression data of *P. vulgaris* was obtained from public data PVGEA (O'Rourke et al. 2014). Expression levels were classified as low (TPM between 1 and 5), medium (TPM between 5 and 10), and high (TPM greater than 10).

4.3 Gene coexpression analysis

We used the R package BioNERO (Almeida-Silva and Venancio 2022) to construct a coexpression network of gene-family genes with at least 8 TPM in seeds. Data preprocessing included replacing missing values with 0, removal of non-expressed genes (median minimal expression = 1), removal of outlying samples using the BioNero standard method (i.e. standardized connectivity - $Z.K < 2$), and adjusting for confounding artifacts to avoid spurious correlations. This process resulted in the exclusion of 35 samples. We used the WGCNA algorithm (Langfelder and Horvath 2008; B. Zhang and Horvath 2005) to compute the gene coexpression network. Hub genes were identified by combining two metrics: correlation of a gene to its module eigengene (module membership > 0.8) and sum of connection weights of a gene to all other genes in the module (degree; top 10% genes with highest degree). Network plots were generated using Cytoscape (Shannon et al. 2003).

4.4 Analysis of duplicated gene pairs

Protein sequences (.faa) and annotation data (.gff3) from Soybean (*G. max*, Wm82.a4.v1) and Common Bean (*P. vulgaris*, V2.1) were obtained from PLAZA 5.0 and Phytozome V12, respectively (Goodstein et al. 2012; Van Bel et al. 2022). Pairwise comparisons of soybean and common bean predicted proteins were conducted with Diamond 0.9.14 (Buchfink, Reuter, and Drost 2021). We used the DupGen Finder tool (Qiao et al. 2019) to classify gene duplication modes in one of five categories: dispersed (DD), proximal (PD), tandem (TD), transposed (TRD), and WGD (Supplementary Table 24). We computed nonsynonymous substitutions per nonsynonymous site (K_a), synonymous substitutions per synonymous site (K_s), and the K_a/K_s ratio for all identified gene pairs with the calculate_Ka_Ks_pipeline script (Qiao et al. 2019). K_s peaks were predicted using the doubletrouble R package (Almeida-Silva and Van de Peer 2024).

4.5 Analysis of legume orthogroups

A diverse collection of proteome datasets spanning 30 distinct legume species gathered from Phytozome (V12/V13) (Goodstein et al. 2012), PLAZA 5.0 (Van Bel et al. 2022), GigaDB (<http://gigadb.org/>), Kazuza genome database (Shirasawa et al. 2014), DRYAD (Vision 2010) and Legumepedia (Garg et al. 2022). When multiple splicing isoforms were present for a gene, only the longest isoform was retained. We used OrthoFinder 2.5.2 (Emms and Kelly 2019) to infer OGs using the parameters -S diamond-ultra-sens (Buchfink, Reuter, and Drost 2021), -M msa, and -T iqtree (iqtree.org). Orthofinder results were explored using the cogeqc R/Bioconductor package (Almeida-Silva and Van de Peer 2023). We used CAFE5 (Mendes et al. 2020) to find expansions and contractions in the OGs (-k 3). As input, we used a time-calibrated ultrametric species tree (Kumar et al. 2022) and the gene counts from each species in each OG. The time-calibrated species tree was calculated using the script make_ultrametric.py based on the SpeciesTree-rooted from Orthofinder and root age (-r 68) available in TimeTree (Kumar et al. 2022). Large OGs indicated by CAFE5 were removed. The key parameters lambda (λ) and mi (μ) were estimated by running CAFE 50 times. The

selection of optimal parameters was guided by the best maximum-likelihood estimate. CAFE results were visualized using the cafeplotter tool (<https://github.com/moshi4/CafePlotter>).

5. CRediT authorship contribution statement

Dayana K. Turquetti-Moraes: Conceptualization, Methodology, Validation, Formal analysis, Data curation, Investigation, Writing – original draft. Cláudio Benício Cardoso-Silva: selection of legume species and proteome data curation. Fabricio Almeida-Silva: Writing – review & editing, Methodology. Thiago M. Venancio: Conceptualization, Supervision, Project administration, Resources, Funding acquisition, Writing – original draft, Writing – review & editing.

6. Acknowledgements

This work was supported by Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brazil (CAPES, Finance Code 001) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). The funding agencies had no role in the design of the study and collection, analysis and interpretation of data and in writing.

7. Data and code availability

To ensure that all findings in this manuscript are reproducible, all data and code have been deposited on https://github.com/Dayana-Turquetti/Genes_related_to_oil_traits.

8. References

- Akoh, Casimir C., Guan-Chiun Lee, Yen-Chywan Liaw, Tai-Huang Huang, and Jei-Fu Shaw. 2004. "GDSL Family of Serine Esterases/Lipases." *Progress in Lipid Research* 43 (6): 534–52. <https://doi.org/10.1016/j.plipres.2004.09.002>.
- Alam, Intikhab, Khadija Batool, Yuanyuan Huang, Junjie Liu, and Liangfa Ge. 2022. "Developing Genetic Engineering Techniques for Control of Seed Size and Yield." *International Journal of Molecular Sciences* 23 (21): 13256. <https://doi.org/10.3390/ijms232113256>.
- Almeida-Silva, Fabricio, Kanhu C. Moharana, Fabricio B. Machado, and Thiago M. Venancio. 2020. "Exploring the Complexity of Soybean (Glycine Max) Transcriptional Regulation Using Global Gene Co-Expression Networks." *Planta* 252 (6): 104. <https://doi.org/10.1007/s00425-020-03499-8>.
- Almeida-Silva, Fabricio, Francisnei Pedrosa-Silva, and Thiago M. Venancio. 2023. "The Soybean Expression Atlas v2: A Comprehensive Database of over 5000 RNA-Seq Samples." *The Plant Journal* 116 (4): 1041–51. <https://doi.org/10.1111/tpj.16459>.
- Almeida-Silva, Fabricio, and Yves Van de Peer. 2023. "Assessing the Quality of Comparative Genomics Data and Results with the Cogeqc R/Bioconductor Package." *Methods in Ecology and Evolution* 14 (12): 2942–52. <https://doi.org/10.1111/2041-210X.14243>.
- Almeida-Silva, Fabricio, and Yves Van de Van de Peer. 2024. "Doubletrouble: An R/Bioconductor Package for the Identification, Classification, and Analysis of Gene and Genome Duplications." bioRxiv. <https://doi.org/10.1101/2024.02.27.582236>.
- Almeida-Silva, Fabricio, and Thiago M. Venancio. 2022. "BioNERO: An All-in-One R/Bioconductor Package for Comprehensive and Easy Biological Network Reconstruction." *Functional & Integrative Genomics* 22 (1): 131–36. <https://doi.org/10.1007/s10142-021-00821-9>.
- An, Yong-qiang Charles, Wolfgang Goettel, Qiang Han, Arthur Bartels, Zongrang Liu, and Wenyan

- Xiao. 2017. "Dynamic Changes of Genome-Wide DNA Methylation during Soybean Seed Development." *Scientific Reports* 7 (1): 12263. <https://doi.org/10.1038/s41598-017-12510-4>.
- Anshu, Aseem Kumar, Vineet Kumar, Anita Rani, Trupti Tayalkar, and Hamendra Singh Parmar. 2022. "Identification of Quantitative Trait Loci Governing Biosynthesis of Phospholipids in Soybean." *Crop and Pasture Science* 74 (3): 219–29. <https://doi.org/10.1071/CP22095>.
- Asen, Nancy D., Rotimi E. Aluko, Alex Martynenko, Alphonsus Utioh, and Pankaj Bhowmik. 2023. "Yellow Field Pea Protein (Pisum Sativum L.): Extraction Technologies, Functionalities, and Applications." *Foods* 12 (21): 3978. <https://doi.org/10.3390/foods12213978>.
- Azani, Nasim, Marielle Babineau, C. Donovan Bailey, Hannah Banks, Ariane Barbosa, Rafael Barbosa Pinto, James Boatwright, et al. 2017. "A New Subfamily Classification of the Leguminosae Based on a Taxonomically Comprehensive Phylogeny." *Taxon* 66 (February): 44–77.
- Aziziar, Zahra, Kianoosh Cheghamirza, Leila Zarei, and Ali Beheshti-Alagha. 2021. "Chemical and Morphological Characteristics of Common Bean Seed and Evaluating Genetic Advance in Commercial Classes." *Cellular and Molecular Biology* 67 (6): 89–99. <https://doi.org/10.14715/cmb/2021.67.6.13>.
- Bai, Shuangyi, James G. Wallis, Peter Denolf, Steven Engelen, Jesse D. Bengtsson, Michel Van Thournout, Jo Dierickx, Boris Haesendonckx, and John Browse. 2020. "The Biochemistry of Headgroup Exchange during Triacylglycerol Synthesis in Canola." *The Plant Journal* 103 (1): 83–94. <https://doi.org/10.1111/tpj.14709>.
- Bai, Yang, X Zhu, Guo X, Zhang W, Zhang G, Chen H, and Zhang Q. 2021. "Molecular Cloning and Functional Characterization of GmAAPTs from Soybean (Glycine Max)." *Plant Signaling & Behavior* 16 (1). <https://doi.org/10.1080/15592324.2020.1845048>.
- Bates, Philip D. 2016. "Understanding the Control of Acyl Flux through the Lipid Metabolic Network of Plant Oil Biosynthesis." *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids, Plant Lipid Biology*, 1861 (9, Part B): 1214–25. <https://doi.org/10.1016/j.bbalip.2016.03.021>.
- Bates, Philip D., and John Browse. 2012. "The Significance of Different Diacylglycerol Synthesis Pathways on Plant Oil Composition and Bioengineering." *Frontiers in Plant Science* 3. <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2012.00147>.
- Bates, Philip D., Timothy P. Durrett, John B. Ohlrogge, and Mike Pollard. 2009. "Analysis of Acyl Fluxes through Multiple Pathways of Triacylglycerol Synthesis in Developing Soybean Embryos." *Plant Physiology* 150 (1): 55–72. <https://doi.org/10.1104/pp.109.137737>.
- Bates, Philip D, Sten Stymne, and John Ohlrogge. 2013. "Biochemical Pathways in Seed Oil Synthesis." *Current Opinion in Plant Biology, Physiology and metabolism*, 16 (3): 358–64. <https://doi.org/10.1016/j.pbi.2013.02.015>.
- Bertioli, David John, Steven B. Cannon, Lutz Froenicke, Guodong Huang, Andrew D. Farmer, Ethalinda K. S. Cannon, Xin Liu, et al. 2016. "The Genome Sequences of Arachis Duranensis and Arachis Ipaensis, the Diploid Ancestors of Cultivated Peanut." *Nature Genetics* 48 (4): 438–46. <https://doi.org/10.1038/ng.3517>.
- Bhardwaj, Harbans L., Anwar A. Hamama, and Edzard van Santen. 2004. "Fatty Acids and Oil Content in White Lupin Seed as Affected by Production Practices." *Journal of the American Oil Chemists' Society* 81 (11): 1035–38. <https://doi.org/10.1007/s11746-004-1018-0>.
- Borisjuk, Ljudmilla, Thuy Ha Nguyen, Thomas Neuberger, Twan Rutten, Henning Tschiersch, Bernhard Claus, Ivo Feussner, et al. 2005. "Gradients of Lipid Storage, Photosynthesis and Plastid Differentiation in Developing Soybean Seeds." *New Phytologist* 167 (3): 761–76. <https://doi.org/10.1111/j.1469-8137.2005.01474.x>.
- Boyang, K. E., L. I. Wenlong, and ZHANG Caiying. 2023. "Expressions of SWEET Genes During Pod and Seed Developments and Under Different Stress Conditions in Soybean." *Journal of Agricultural Science & Technology (1008-0864)* 25 (8).
- Brown, Anne V, Shawn I Conners, Wei Huang, Andrew P Wilkey, David Grant, Nathan T Weeks, Steven B Cannon, Michelle A Graham, and Rex T Nelson. 2021. "A New Decade and New Data at

- SoyBase, the USDA-ARS Soybean Genetics and Genomics Database." *Nucleic Acids Research* 49 (D1): D1496–1501. <https://doi.org/10.1093/nar/gkaa1107>.
- Buchfink, Benjamin, Klaus Reuter, and Hajk-Georg Drost. 2021. "Sensitive Protein Alignments at Tree-of-Life Scale Using DIAMOND." *Nature Methods* 18 (4): 366–68. <https://doi.org/10.1038/s41592-021-01101-x>.
- Cai, Zhandong, Peiqi Xian, Yanbo Cheng, Yiwang Zhong, Yuan Yang, Qianghua Zhou, Tengxiang Lian, Qibin Ma, Hai Nian, and Liangfa Ge. 2023. "MOTHER-OF-FT-AND-TFL1 Regulates the Seed Oil and Protein Content in Soybean." *New Phytologist* 239 (3): 905–19. <https://doi.org/10.1111/nph.18792>.
- Carretero-Paulet, Lorenzo, and Yves Van de Peer. 2020. "The Evolutionary Conundrum of Whole-genome Duplication." *American Journal of Botany* 107 (8): 1101–5. <https://doi.org/10.1002/ajb2.1520>.
- Chang, Yue, Huan Liu, Min Liu, Xuezhu Liao, Sunil Kumar Sahu, Yuan Fu, Bo Song, et al. 2019. "The Draft Genomes of Five Agriculturally Important African Orphan Crops." *GigaScience* 8 (3): giy152. <https://doi.org/10.1093/gigascience/giy152>.
- Chen, Beibei, Gaoyang Zhang, Penghui Li, Jihong Yang, Liang Guo, Christoph Benning, Xuemin Wang, and Jian Zhao. 2020. "Multiple GmWRI1s Are Redundantly Involved in Seed Filling and Nodulation by Regulating Plastidic Glycolysis, Lipid Biosynthesis and Hormone Signalling in Soybean (Glycine Max)." *Plant Biotechnology Journal* 18 (1): 155–71. <https://doi.org/10.1111/pbi.13183>.
- Chen, Guanqun, John L. Harwood, M. Joanne Lemieux, Scot J. Stone, and Randall J. Weselake. 2022. "Acyl-CoA:Diacylglycerol Acyltransferase: Properties, Physiological Roles, Metabolic Engineering and Intentional Control." *Progress in Lipid Research* 88 (November): 101181. <https://doi.org/10.1016/j.plipres.2022.101181>.
- Chen, Ming, Gongshe Han, Charles R. Dietrich, Teresa M. Dunn, and Edgar B. Cahoon. 2006. "The Essential Nature of Sphingolipids in Plants as Revealed by the Functional Identification and Characterization of the Arabidopsis LCB1 Subunit of Serine Palmitoyltransferase." *The Plant Cell* 18 (12): 3576–93. <https://doi.org/10.1105/tpc.105.040774>.
- Cheng, Feng, Jian Wu, Xu Cai, Jianli Liang, Michael Freeling, and Xiaowu Wang. 2018. "Gene Retention, Fractionation and Subgenome Differences in Polyploid Plants." *Nature Plants* 4 (5): 258–68. <https://doi.org/10.1038/s41477-018-0136-7>.
- Dam, Svend, Brian S. Laursen, Jane H. Ørnfeldt, Bjarne Jochimsen, Hans Henrik Stærfeldt, Carsten Friis, Kasper Nielsen, et al. 2009. "The Proteome of Seed Development in the Model Legume Lotus Japonicus." *Plant Physiology* 149 (3): 1325–40. <https://doi.org/10.1104/pp.108.133405>.
- De Vega, Jose J., Sarah Ayling, Matthew Hegarty, Dave Kudrna, Jose L. Goicoechea, Åshild Ergon, Odd A. Rognli, et al. 2015. "Red Clover (*Trifolium Pratense* L.) Draft Genome Provides a Platform for Trait Improvement." *Scientific Reports* 5 (1): 17394. <https://doi.org/10.1038/srep17394>.
- Ding, Li-Na, Xiao-Juan Guo, Ming Li, Zheng-Li Fu, Su-Zhen Yan, Ke-Ming Zhu, Zheng Wang, and Xiao-Li Tan. 2019. "Improving Seed Germination and Oil Contents by Regulating the GDLS Transcriptional Level in Brassica Napus." *Plant Cell Reports* 38 (2): 243–53. <https://doi.org/10.1007/s00299-018-2365-7>.
- Duan, Zongbiao, Qing Li, Hong Wang, Xuemei He, and Min Zhang. 2023. "Genetic Regulatory Networks of Soybean Seed Size, Oil and Protein Contents." *Frontiers in Plant Science* 14. <https://www.frontiersin.org/articles/10.3389/fpls.2023.1160418>.
- Duke, James A., and Edward S. Ayensu. 1985. *Medicinal Plants of China*. Medicinal Plants of the World. Algonac, Mich.: Reference Publications.
- Emms, David M., and Steven Kelly. 2019. "OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics." *Genome Biology* 20 (1): 238. <https://doi.org/10.1186/s13059-019-1832-y>.
- Fang, Chao, Yanming Ma, Shiwen Wu, Zhi Liu, Zheng Wang, Rui Yang, Guanghui Hu, et al. 2017. "Genome-Wide Association Studies Dissect the Genetic Networks Underlying Agronomical

- Traits in Soybean." *Genome Biology* 18 (1): 161. <https://doi.org/10.1186/s13059-017-1289-9>.
- Flagel, Lex E., and Jonathan F. Wendel. 2009. "Gene Duplication and Evolutionary Novelty in Plants." *New Phytologist* 183 (3): 557–64. <https://doi.org/10.1111/j.1469-8137.2009.02923.x>.
- Freeling, Michael. 2009. "Bias in Plant Gene Content Following Different Sorts of Duplication: Tandem, Whole-Genome, Segmental, or by Transposition." *Annual Review of Plant Biology* 60 (1): 433–53. <https://doi.org/10.1146/annurev.arplant.043008.092122>.
- Freeling, Michael, Michael J Scanlon, and John E Fowler. 2015. "Fractionation and Subfunctionalization Following Genome Duplications: Mechanisms That Drive Gene Content and Their Consequences." *Current Opinion in Genetics & Development*, Genomes and evolution, 35 (December): 110–18. <https://doi.org/10.1016/j.gde.2015.11.002>.
- Garg, Vanika, Olga Dudchenko, Jinpeng Wang, Aamir W. Khan, Saurabh Gupta, Parwinder Kaur, Kai Han, et al. 2022. "Chromosome-Length Genome Assemblies of Six Legume Species Provide Insights into Genome Organization, Evolution, and Agronomic Traits for Crop Improvement." *Journal of Advanced Research*, Advances in Plant Genetics and Genomics, 42 (December): 315–29. <https://doi.org/10.1016/j.jare.2021.10.009>.
- Goodstein, David M., Shengqiang Shu, Russell Howson, Rochak Neupane, Richard D. Hayes, Joni Fazo, Therese Mitros, et al. 2012. "Phytozome: A Comparative Platform for Green Plant Genomics." *Nucleic Acids Research* 40 (D1): D1178–86. <https://doi.org/10.1093/nar/gkr944>.
- Grechkin, Alexander N. 2002. "Hydroperoxide Lyase and Divinyl Ether Synthase." *Prostaglandins & Other Lipid Mediators*, Molecular Biology of the Arachidonate Cascade, 68–69 (August): 457–70. [https://doi.org/10.1016/S0090-6980\(02\)00048-5](https://doi.org/10.1016/S0090-6980(02)00048-5).
- Griesmann, Maximilian, Yue Chang, Xin Liu, Yue Song, Georg Haberer, Matthew B. Crook, Benjamin Billault-Penneteau, et al. 2018. "Phylogenomics Reveals Multiple Losses of Nitrogen-Fixing Root Nodule Symbiosis." *Science* 361 (6398): eaat1743. <https://doi.org/10.1126/science.aat1743>.
- Grosso, Nelson R., Valeria Nepote, and Carlos A. Guzmán. 2000. "Chemical Composition of Some Wild Peanut Species (*Arachis* L.) Seeds." *Journal of Agricultural and Food Chemistry* 48 (3): 806–9. <https://doi.org/10.1021/jf9901744>.
- Grygier, Anna, Suryakant Chakradhari, Katarzyna Ratusz, Magdalena Rudzińska, Khageshwar Singh Patel, Danija Lazdiņa, and Paweł Górnaś. 2022. "Seven Underutilized Species of the Fabaceae Family with High Potential for Industrial Application as Alternative Sources of Oil and Lipophilic Bioactive Compounds." *Industrial Crops and Products* 186 (October): 115251. <https://doi.org/10.1016/j.indcrop.2022.115251>.
- Gupta, Ashutosh, Atanas G. Atanasov, Yutong Li, Neeraj Kumar, and Anupam Bishayee. 2022. "Chlorogenic Acid for Cancer Prevention and Therapy: Current Status on Efficacy and Mechanisms of Action." *Pharmacological Research* 186: 106505. <https://doi.org/10.1016/j.phrs.2022.106505>.
- Gupta, Sonal, Kashif Nawaz, Sabiha Parween, Riti Roy, Kamlesh Sahu, Anil Kumar Pole, Hitaishi Khandal, Rishi Srivastava, Swarup Kumar Parida, and Debasis Chattopadhyay. 2017. "Draft Genome Sequence of *Cicer Reticulatum* L., the Wild Progenitor of Chickpea Provides a Resource for Agronomic Trait Improvement." *DNA Research* 24 (1): 1–10. <https://doi.org/10.1093/dnares/dsw042>.
- Hake, Sarah, and Jeffrey Ross-Ibarra. 2015. "Genetic, Evolutionary and Plant Breeding Insights from the Domestication of Maize." *ELife* 4: e05861. <https://doi.org/10.7554/eLife.05861>.
- 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. <https://doi.org/10.1111/pbi.12615>.
- Hassan, L. G., K. J. Umar, and H. Yuguda. 2007. "Nutritional Evaluation Of *Faidherbia Albida* Seeds And Pulp As Source Of Feeds For Livestock," September.

- <https://tspace.library.utoronto.ca/handle/1807/55388>.
- Hemmati, Shiva, and Hassan Seradj. 2016. "Justicidin B: A Promising Bioactive Lignan." *Molecules (Basel, Switzerland)* 21 (7): 820. <https://doi.org/10.3390/molecules21070820>.
- Hooker, Julia C., Nour Nissan, Doris Luckert, Gerardo Zapata, Anfu Hou, Ramona M. Mohr, Aaron J. Glenn, et al. 2022. "GmSWEET29 and Paralog GmSWEET34 Are Differentially Expressed between Soybeans Grown in Eastern and Western Canada." *Plants* 11 (18): 2337. <https://doi.org/10.3390/plants11182337>.
- Hossain, Shahdat, Rashed Ahmed, Sujan Bhowmick, Abdullah Al Mamun, and Michio Hashimoto. 2016. "Proximate Composition and Fatty Acid Analysis of Lablab Purpureus (L.) Legume Seed: Implicates to Both Protein and Essential Fatty Acid Supplementation." *SpringerPlus* 5 (1): 1899. <https://doi.org/10.1186/s40064-016-3587-1>.
- Huang, Jinghua, Qibin Ma, Zhandong Cai, Qiuju Xia, Shuxian Li, Jia Jia, Li Chu, Tengxiang Lian, Hai Nian, and Yanbo Cheng. 2020. "Identification and Mapping of Stable QTLs for Seed Oil and Protein Content in Soybean [Glycine Max (L.) Merr.]." *Journal of Agricultural and Food Chemistry* 68 (23): 6448–60. <https://doi.org/10.1021/acs.jafc.0c01271>.
- Huang, Li, Huifang Jiang, Xiaoping Ren, Yuning Chen, Yingjie Xiao, Xinyan Zhao, Mei Tang, Jiaquan Huang, Hari D. Upadhyaya, and Boshou Liao. 2012. "Abundant Microsatellite Diversity and Oil Content in Wild Arachis Species." *PLOS ONE* 7 (11): e50002. <https://doi.org/10.1371/journal.pone.0050002>.
- 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. <https://doi.org/10.1038/s41467-019-14197-9>.
- Jing, Guangqin, Daoping Tang, Yao Yao, Youke Su, Yue Shen, Yang Bai, Wen Jing, et al. 2021. "Seed Specifically Over-Expressing DGAT2A Enhances Oil and Linoleic Acid Contents in Soybean Seeds." *Biochemical and Biophysical Research Communications* 568 (September): 143–50. <https://doi.org/10.1016/j.bbrc.2021.06.087>.
- Jusoh, Malinna, Saw Hong Loh, Tse Seng Chuah, Ahmad Aziz, and Thye San Cha. 2015. "Indole-3-Acetic Acid (IAA) Induced Changes in Oil Content, Fatty Acid Profiles and Expression of Four Fatty Acid Biosynthetic Genes in Chlorella Vulgaris at Early Stationary Growth Phase." *Phytochemistry* 111 (March): 65–71. <https://doi.org/10.1016/j.phytochem.2014.12.022>.
- Kanai, Masatake, Tetsuya Yamada, Makoto Hayashi, Shoji Mano, and Mikio Nishimura. 2019. "Soybean (Glycine Max L.) Triacylglycerol Lipase GmSDP1 Regulates the Quality and Quantity of Seed Oil." *Scientific Reports* 9 (1): 8924. <https://doi.org/10.1038/s41598-019-45331-8>.
- Kang, Sang-Ho, Ramesh Prasad Pandey, Chang-Muk Lee, Joon-Soo Sim, Jin-Tae Jeong, Beom-Soon Choi, Myunghee Jung, et al. 2020. "Genome-Enabled Discovery of Anthraquinone Biosynthesis in Senna Tora." *Nature Communications* 11 (1): 5875. <https://doi.org/10.1038/s41467-020-19681-1>.
- Kang, Yang Jae, Sue K. Kim, Moon Young Kim, Puji Lestari, Kil Hyun Kim, Bo-Keun Ha, Tae Hwan Jun, et al. 2014. "Genome Sequence of Mungbean and Insights into Evolution within Vigna Species." *Nature Communications* 5 (1): 5443. <https://doi.org/10.1038/ncomms6443>.
- Kim, Sang-Chul, and Xuemin Wang. 2020. "Phosphatidic Acid: An Emerging Versatile Class of Cellular Mediators." *Essays in Biochemistry* 64 (3): 533–46. <https://doi.org/10.1042/EBC20190089>.
- Kreplak, Jonathan, Mohammed-Amin Madoui, Petr Cápál, Petr Novák, Karine Labadie, Grégoire Aubert, Philipp E. Bayer, et al. 2019. "A Reference Genome for Pea Provides Insight into Legume Genome Evolution." *Nature Genetics* 51 (9): 1411–22. <https://doi.org/10.1038/s41588-019-0480-1>.
- Kryuchkova-Mostacci, Nadezda, and Marc Robinson-Rechavi. 2017. "A Benchmark of Gene Expression Tissue-Specificity Metrics." *Briefings in Bioinformatics* 18 (2): 205–14. <https://doi.org/10.1093/bib/bbw008>.

- 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): msac174. <https://doi.org/10.1093/molbev/msac174>.
- Langfelder, Peter, and Steve Horvath. 2008. "WGCNA: An R Package for Weighted Correlation Network Analysis." *BMC Bioinformatics* 9 (1): 559. <https://doi.org/10.1186/1471-2105-9-559>.
- Lavin, Matt, Patrick S Herendeen, and Martin F Wojciechowski. 2005. "Evolutionary Rates Analysis of Leguminosae Implicates a Rapid Diversification of Lineages during the Tertiary." *Systematic Biology* 54 (4): 575–94. <https://doi.org/10.1080/10635150590947131>.
- Lee, Junghoon, Ruth Welti, William T. Schapaugh, and Harold N. Trick. 2011. "Phospholipid and Triacylglycerol Profiles Modified by PLD Suppression in Soybean Seed." *Plant Biotechnology Journal* 9 (3): 359–72. <https://doi.org/10.1111/j.1467-7652.2010.00562.x>.
- Lemus-Conejo, Ana, Fernando Rivero-Pino, Sergio Montserrat-de la Paz, and Maria C. Millan-Linares. 2023. "Nutritional Composition and Biological Activity of Narrow-Leafed Lupins (*Lupinus Angustifolius* L.) Hydrolysates and Seeds." *Food Chemistry* 420 (September): 136104. <https://doi.org/10.1016/j.foodchem.2023.136104>.
- Li, Jiajia, Jinghui Zhao, Yinghui Li, Yali Gao, Sunan Hua, Muhammad Nadeem, Genlou Sun, et al. 2019. "Identification of a Novel Seed Size Associated Locus *SW9-1* in Soybean." *The Crop Journal* 7 (4): 548–59. <https://doi.org/10.1016/j.cj.2018.12.010>.
- Li, Runzhi, Keshun Yu, and David F. Hildebrand. 2010. "DGAT1, DGAT2 and PDAT Expression in Seeds and Other Tissues of Epoxy and Hydroxy Fatty Acid Accumulating Plants." *Lipids* 45 (2): 145–57. <https://doi.org/10.1007/s11745-010-3385-4>.
- Li, Si-Yu, Qing Zhang, Yuan-Hang Jin, Ji-Xin Zou, Yu-Sheng Zheng, and Dong-Dong Li. 2020. "A MADS-Box Gene, EgMADS21, Negatively Regulates EgDGAT2 Expression and Decreases Polyunsaturated Fatty Acid Accumulation in Oil Palm (*Elaeis Guineensis* Jacq.)." *Plant Cell Reports* 39 (11): 1505–16. <https://doi.org/10.1007/s00299-020-02579-z>.
- Li, Zheng, Michael T.W. McKibben, Geoffrey S. Finch, Paul D. Blischak, Brittany L. Sutherland, and Michael S. Barker. 2021. "Patterns and Processes of Diploidization in Land Plants." *Annual Review of Plant Biology* 72 (1): 387–410. <https://doi.org/10.1146/annurev-arplant-050718-100344>.
- Li-Beisson, Yonghua, Jens Neunzig, Youngsook Lee, and Katrin Philipp. 2017. "Plant Membrane-Protein Mediated Intracellular Traffic of Fatty Acids and Acyl Lipids." *Current Opinion in Plant Biology*, 40 Cell Biology 2017, 40 (December): 138–46. <https://doi.org/10.1016/j.pbi.2017.09.006>.
- Liu, Changkai, Heng Chen, Qintan Yu, Haidong Gu, Yansheng Li, Bingjie Tu, Hengyou Zhang, Qiuying Zhang, and Xiaobing Liu. 2023. "Identification of Quantitative Trait Loci and Candidate Genes for Seed Sucrose and Soluble Sugar Concentrations in Soybean." *Crop Science* 63 (5): 2976–92. <https://doi.org/10.1002/csc2.21080>.
- Liu, Hongbo, Linyan Wei, Jinbo Zhu, Bingxin Zhang, Yi Gan, and Yueping Zheng. 2022. "Identification of GmGPATs and Their Effect on Glycerolipid Biosynthesis through Seed-Specific Expression in Soybean." *Molecular Biology Reports* 49 (10): 9585–92. <https://doi.org/10.1007/s11033-022-07852-w>.
- Liu, Jin-Yang, Pei Li, Ya-Wen Zhang, Jian-Fang Zuo, Guo Li, Xu Han, Jim M. Dunwell, and Yuan-Ming Zhang. 2020. "Three-Dimensional Genetic Networks among Seed Oil-Related Traits, Metabolites and Genes Reveal the Genetic Foundations of Oil Synthesis in Soybean." *The Plant Journal* 103 (3): 1103–24. <https://doi.org/10.1111/tpj.14788>.
- Liu, Jin-Yang, Ya-Wen Zhang, Xu Han, Jian-Fang Zuo, Zhibin Zhang, Haihong Shang, Qijian Song, and Yuan-Ming Zhang. 2020. "An Evolutionary Population Structure Model Reveals Pleiotropic Effects of GmPDAT for Traits Related to Seed Size and Oil Content in Soybean." *Journal of Experimental Botany* 71 (22): 6988–7002. <https://doi.org/10.1093/jxb/eraa426>.
- Liu, Peng, Tian Xie, Xinyue Wu, Gongshe Han, Sita D. Gupta, Zike Zhang, Jian Yue, et al. 2023.

- "Mechanism of Sphingolipid Homeostasis Revealed by Structural Analysis of Arabidopsis SPT-ORM1 Complex." *Science Advances* 9 (13): eadg0728. <https://doi.org/10.1126/sciadv.adg0728>.
- Liu, Wennuan, David F. Hildebrand, and Glenn B. Collins. 1995. "Auxin-Regulated Changes of Fatty Acid Content and Composition in Soybean Zygotic Embryo Cotyledons." *Plant Science* 106 (1): 31–42. [https://doi.org/10.1016/0168-9452\(95\)04067-5](https://doi.org/10.1016/0168-9452(95)04067-5).
- Liu, Yun-Feng, Qing-Tian Li, Xiang Lu, Qing-Xin Song, Sin-Man Lam, Wan-Ke Zhang, Biao Ma, et al. 2014. "Soybean GmMYB73 Promotes Lipid Accumulation in Transgenic Plants." *BMC Plant Biology* 14 (March): 73. <https://doi.org/10.1186/1471-2229-14-73>.
- Lonardi, Stefano, María Muñoz-Amatriáin, Qihua Liang, Shengqiang Shu, Steve I. Wanamaker, Sassoum Lo, Jaakko Tanskanen, et al. 2019. "The Genome of Cowpea (*Vigna Unguiculata* [L.] Walp.)." *The Plant Journal* 98 (5): 767–82. <https://doi.org/10.1111/tpj.14349>.
- Lu, Chaofu, Zhanguo Xin, Zhonghai Ren, Martine Miquel, and John Browse. 2009. "An Enzyme Regulating Triacylglycerol Composition Is Encoded by the ROD1 Gene of Arabidopsis." *Proceedings of the National Academy of Sciences* 106 (44): 18837–42. <https://doi.org/10.1073/pnas.0908848106>.
- Lu, Long, Wei Wei, Qing-Tian Li, Xiao-Hua Bian, Xiang Lu, Yang Hu, Tong Cheng, et al. 2021. "A Transcriptional Regulatory Module Controls Lipid Accumulation in Soybean." *The New Phytologist* 231 (2): 661–78. <https://doi.org/10.1111/nph.17401>.
- Lu, Qing, Haifen Li, Yanbin Hong, Guoqiang Zhang, Shijie Wen, Xingyu Li, Guiyuan Zhou, et al. 2018. "Genome Sequencing and Analysis of the Peanut B-Genome Progenitor (*Arachis Ipaensis*)."
Frontiers in Plant Science 9. <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2018.00604>.
- Lu, Sijia, Chao Fang, Jun Abe, Fanjiang Kong, and Baohui Liu. 2022. "Current Overview on the Genetic Basis of Key Genes Involved in Soybean Domestication." *ABIOTECH* 3 (2): 126–39. <https://doi.org/10.1007/s42994-022-00074-5>.
- Lu, Xiang, Qing-Tian Li, Qing Xiong, Wei Li, Ying-Dong Bi, Yong-Cai Lai, Xin-Lei Liu, et al. 2016. "The Transcriptomic Signature of Developing Soybean Seeds Reveals the Genetic Basis of Seed Trait Adaptation during Domestication." *The Plant Journal: For Cell and Molecular Biology* 86 (6): 530–44. <https://doi.org/10.1111/tpj.13181>.
- Manan, Sehrish, Beibei Chen, Guangbiao She, Xiaochun Wan, and Jian Zhao. 2017. "Transport and Transcriptional Regulation of Oil Production in Plants." *Critical Reviews in Biotechnology* 37 (5): 641–55. <https://doi.org/10.1080/07388551.2016.1212185>.
- Marchive, Chloé, Krisztina Nikovics, Alexandra To, Loïc Lepiniec, and Sébastien Baud. 2014. "Transcriptional Regulation of Fatty Acid Production in Higher Plants: Molecular Bases and Biotechnological Outcomes." *European Journal of Lipid Science and Technology* 116 (10): 1332–43. <https://doi.org/10.1002/ejlt.201400027>.
- McGlew, Kathleen, Vincent Shaw, Meng Zhang, Ryeo Jin Kim, Weili Yang, Basil Shorrosh, Mi Chung Suh, and John Ohlrogge. 2015. "An Annotated Database of Arabidopsis Mutants of Acyl Lipid Metabolism." *Plant Cell Reports* 34 (4): 519–32. <https://doi.org/10.1007/s00299-014-1710-8>.
- Meinke, D. W., J. Chen, and R. N. Beachy. 1981. "Expression of Storage-Protein Genes during Soybean Seed Development." *Planta* 153 (2): 130–39. <https://doi.org/10.1007/BF00384094>.
- Mendes, Fábio K., Dan Vanderpool, Ben Fulton, and Matthew W. Hahn. 2020. "CAFE 5 Models Variation in Evolutionary Rates among Gene Families." *Bioinformatics (Oxford, England)*, December, btaa1022. <https://doi.org/10.1093/bioinformatics/btaa1022>.
- Minka, Samuel R, and Maud Bruneteau. 2000. "Partial Chemical Composition of Bambara Pea [*Vigna Subterranea* (L.) Verde]." *Food Chemistry* 68 (3): 273–76. [https://doi.org/10.1016/S0308-8146\(99\)00186-7](https://doi.org/10.1016/S0308-8146(99)00186-7).
- Moghaddam, Samira Mafi, Atena Oladzad, Chushin Koh, Larissa Ramsay, John P. Hart, Sujana Mamidi, Genevieve Hoopes, et al. 2021. "The Tepary Bean Genome Provides Insight into Evolution and Domestication under Heat Stress." *Nature Communications* 12 (1): 2638.

- <https://doi.org/10.1038/s41467-021-22858-x>.
- Moharana, Kanhu C., and Thiago M. Venancio. 2020. "Polyploidization Events Shaped the Transcription Factor Repertoires in Legumes (Fabaceae)." *The Plant Journal: For Cell and Molecular Biology* 103 (2): 726–41. <https://doi.org/10.1111/tpj.14765>.
- Morley, Stewart A., Fangfang Ma, Mazen Alazem, Cheryl Frankfater, Hochul Yi, Tessa Burch-Smith, Tom Elmo Clemente, Veena Veena, Hanh Nguyen, and Doug K. Allen. 2023. "Expression of Malic Enzyme Reveals Subcellular Carbon Partitioning for Storage Reserve Production in Soybeans." *New Phytologist* n/a (n/a). <https://doi.org/10.1111/nph.18835>.
- Nguyen, Quoc Thien, Anna Kisiala, Peter Andreas, R. J. Neil Emery, and Suresh Narine. 2016. "Soybean Seed Development: Fatty Acid and Phytohormone Metabolism and Their Interactions." *Current Genomics* 17 (3): 241–60.
- Niu, Yuan, Guoliang Zhang, Faxiang Wan, and Yuan-Ming Zhang. 2020. "Integration of RNA-Seq Profiling with Genome-Wide Association Study Predicts Candidate Genes for Oil Accumulation in Soybean." *Crop and Pasture Science* 71 (12): 996. <https://doi.org/10.1071/CP20358>.
- O'Rourke, Jamie A., Luis P. Iniguez, Fengli Fu, Bruna Bucciarelli, Susan S. Miller, Scott A. Jackson, Philip E. McClean, et al. 2014. "An RNA-Seq Based Gene Expression Atlas of the Common Bean." *BMC Genomics* 15 (1): 866. <https://doi.org/10.1186/1471-2164-15-866>.
- Palupi, H.T., T. Estiasih, Yunianta, and A. Sutrisno. 2022. "Physicochemical and Protein Characterization of Lima Bean (*Phaseolus Lunatus* L) Seed." *Food Research* 6 (1): 168–77. [https://doi.org/10.26656/fr.2017.6\(1\).107](https://doi.org/10.26656/fr.2017.6(1).107).
- Pan, Xue, Fred Y. Peng, and Randall J. Weselake. 2015. "Genome-Wide Analysis of PHOSPHOLIPID:DIACYLGLYCEROL ACYLTRANSFERASE (PDAT) Genes in Plants Reveals the Eudicot-Wide PDAT Gene Expansion and Altered Selective Pressures Acting on the Core Eudicot PDAT Paralog." *Plant Physiology* 167 (3): 887–904. <https://doi.org/10.1104/pp.114.253658>.
- Panchy, Nicholas, Melissa Lehti-Shiu, and Shin-Han Shiu. 2016. "Evolution of Gene Duplication in Plants1[OPEN]." *Plant Physiology* 171 (4): 2294–2316. <https://doi.org/10.1104/pp.16.00523>.
- Patil, Gunvant, Tri D. Vuong, Sandip Kale, Babu Valliyodan, Rupesh Deshmukh, Chengsong Zhu, Xiaolei Wu, et al. 2018. "Dissecting Genomic Hotspots Underlying Seed Protein, Oil, and Sucrose Content in an Interspecific Mapping Population of Soybean Using High-Density Linkage Mapping." *Plant Biotechnology Journal* 16 (11): 1939–53. <https://doi.org/10.1111/pbi.12929>.
- Pattee, H. E., D. K. Salunkhe, S. K. Sathe, N. R. Reddy, and Robert L. Ory. 1983. "Legume Lipids." *CRC Critical Reviews in Food Science and Nutrition* 17 (2): 97–139. <https://doi.org/10.1080/10408398209527345>.
- Perchuk, Irina, Tatyana Shelenga, Maria Gurkina, Elena Miroshnichenko, and Marina Burlyaeva. 2020. "Composition of Primary and Secondary Metabolite Compounds in Seeds and Pods of Asparagus Bean (*Vigna Unguiculata* (L.) Walp.) from China." *Molecules* 25 (17): 3778. <https://doi.org/10.3390/molecules25173778>.
- Pham, Anh-Tung, J. Grover Shannon, and Kristin D. Bilyeu. 2012. "Combinations of Mutant FAD2 and FAD3 Genes to Produce High Oleic Acid and Low Linolenic Acid Soybean Oil." *Theoretical and Applied Genetics* 125 (3): 503–15. <https://doi.org/10.1007/s00122-012-1849-z>.
- Pyc, Michal, Satinder K. Gidda, Damien Seay, Nicolas Esnay, Franziska K. Kretschmar, Yingqi Cai, Nathan M. Doner, et al. 2021. "LDIP Cooperates with SEIPIN and LDAP to Facilitate Lipid Droplet Biogenesis in Arabidopsis." *The Plant Cell* 33 (9): 3076–3103. <https://doi.org/10.1093/plcell/koab179>.
- Qi, Zhaoming, Chaocheng Guo, Haiyang Li, Hongmei Qiu, Hui Li, CholNam Jong, Guolong Yu, et al. 2023. "Natural Variation in Fatty Acid 9 Is a Determinant of Fatty Acid and Protein Content." *Plant Biotechnology Journal* n/a (n/a). <https://doi.org/10.1111/pbi.14222>.
- Qiao, Xin, Qionghou Li, Hao Yin, Kaijie Qi, Leitong Li, Runze Wang, Shaoling Zhang, and Andrew H.

- Paterson. 2019. "Gene Duplication and Evolution in Recurring Polyploidization–Diploidization Cycles in Plants." *Genome Biology* 20 (1): 38. <https://doi.org/10.1186/s13059-019-1650-2>.
- Qin, Shuangshuang, Lingqing Wu, Kunhua Wei, Ying Liang, Zhijun Song, Xiaolei Zhou, Shuo Wang, et al. 2019. "A Draft Genome for *Spatholobus Suberectus*." *Scientific Data* 6 (1): 113. <https://doi.org/10.1038/s41597-019-0110-x>.
- Rackis, J.J., D.J. Sessa, and D.H. Honig. 1979. "Flavor Problems of Vegetable Food Proteins." *Journal of the American Oil Chemists' Society* 56 (3): 262–71. <https://doi.org/10.1007/BF02671470>.
- Rahman, Shanjida, Shahidul Islam, Zitong Yu, Maoyun She, Eviatar Nevo, and Wujun Ma. 2020. "Current Progress in Understanding and Recovering the Wheat Genes Lost in Evolution and Domestication." *International Journal of Molecular Sciences* 21 (16): 5836. <https://doi.org/10.3390/ijms21165836>.
- Ren, Ren, Haifeng Wang, Chunce Guo, Ning Zhang, Liping Zeng, Yamao Chen, Hong Ma, and Ji Qi. 2018. "Widespread Whole Genome Duplications Contribute to Genome Complexity and Species Diversity in Angiosperms." *Molecular Plant, Genome Biology*, 11 (3): 414–28. <https://doi.org/10.1016/j.molp.2018.01.002>.
- Renny-Byfield, Simon, and Jonathan F. Wendel. 2014. "Doubling down on Genomes: Polyploidy and Crop Plants." *American Journal of Botany* 101 (10): 1711–25. <https://doi.org/10.3732/ajb.1400119>.
- Sandhu, Devinder, Jessie L. Alt, Curtis W. Scherder, Walter R. Fehr, and Madan K. Bhattacharyya. 2007. "Enhanced Oleic Acid Content in the Soybean Mutant M23 Is Associated with the Deletion in the Fad2-1a Gene Encoding a Fatty Acid Desaturase." *Journal of the American Oil Chemists' Society*. <http://dx.doi.org/10.1007/s11746-007-1037-5>.
- Santamaría, María Estrella, Mercedes Diaz-Mendoza, Isabel Diaz, and Manuel Martinez. 2014. "Plant Protein Peptidase Inhibitors: An Evolutionary Overview Based on Comparative Genomics." *BMC Genomics* 15 (1): 812. <https://doi.org/10.1186/1471-2164-15-812>.
- Sato, Shusei, Yasukazu Nakamura, Takakazu Kaneko, Erika Asamizu, Tomohiko Kato, Mitsuteru Nakao, Shigemi Sasamoto, et al. 2008. "Genome Structure of the Legume, *Lotus Japonicus*." *DNA Research* 15 (4): 227–39. <https://doi.org/10.1093/dnares/dsn008>.
- Schaefer, Robert J., Jean-Michel Michno, and Chad L. Myers. 2017. "Unraveling Gene Function in Agricultural Species Using Gene Co-Expression Networks." *Biochimica Et Biophysica Acta. Gene Regulatory Mechanisms* 1860 (1): 53–63. <https://doi.org/10.1016/j.bbagr.2016.07.016>.
- Schmutz, Jeremy, Steven B. Cannon, Jessica Schlueter, Jianxin Ma, Therese Mitros, William Nelson, David L. Hyten, et al. 2010. "Genome Sequence of the Palaeopolyploid Soybean." *Nature* 463 (7278): 178–83. <https://doi.org/10.1038/nature08670>.
- Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. 2003. "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks." *Genome Research* 13 (11): 2498–2504. <https://doi.org/10.1101/gr.1239303>.
- Sharma, Sheel, Agarwal Nidhi, and Verma Preeti. 2011. "Pigeon Pea (*Cajanus Cajan* L.): A Hidden Treasure of Regime Nutrition." 2011. <https://www.indianjournals.com/ijor.aspx?target=ijor:jfeb&volume=1&issue=2&article=002>.
- Shasidhar, Yaduru, Manish K. Vishwakarma, Manish K. Pandey, Pasupuleti Janila, Murali T. Variath, Surendra S. Manohar, Shyam N. Nigam, Baozhu Guo, and Rajeev K. Varshney. 2017. "Molecular Mapping of Oil Content and Fatty Acids Using Dense Genetic Maps in Groundnut (*Arachis Hypogaea* L.)." *Frontiers in Plant Science* 8 (May): 794. <https://doi.org/10.3389/fpls.2017.00794>.
- Shen, Yanting, Shan Hong, and Yonghui Li. 2022. "Pea Protein Composition, Functionality, Modification, and Food Applications: A Review." *Advances in Food and Nutrition Research* 101: 71–127. <https://doi.org/10.1016/bs.afnr.2022.02.002>.
- Shirasawa, Kenta, Sachiko Isobe, Satoshi Tabata, and Hideki Hirakawa. 2014. "Kazusa Marker

- DataBase: A Database for Genomics, Genetics, and Molecular Breeding in Plants." *Breeding Science* 64 (3): 264–71. <https://doi.org/10.1270/jsbbs.64.264>.
- Shoemaker, Randy C, Jessica Schlueter, and Jeff J Doyle. 2006. "Paleopolyploidy and Gene Duplication in Soybean and Other Legumes." *Current Opinion in Plant Biology*, Genome studies and molecular genetics: Part 1: Model legumes / edited by Nevin D Young and Randy C Shoemaker; Part 2: Maize genomics / edited by Susan R Wessler. Plant biotechnology / edited by John Salmeron and Luis R Herrera-Estrella, 9 (2): 104–9. <https://doi.org/10.1016/j.pbi.2006.01.007>.
- Shu, Yongjun, Diansi Yu, Dan Wang, Donglin Guo, and Changhong Guo. 2013. "Genome-Wide Survey and Expression Analysis of the MADS-Box Gene Family in Soybean." *Molecular Biology Reports* 40 (6): 3901–11. <https://doi.org/10.1007/s11033-012-2438-6>.
- Shweta, and Rajan Katoch. 2014. "Comparison of Nutritional Composition of Vigna Spp. Prevalent in the Mid-Hill Region of Himachal Pradesh" 27 (January): 202–7.
- Soltis, Douglas E., Victor A. Albert, Jim Leebens-Mack, Charles D. Bell, Andrew H. Paterson, Chunfang Zheng, David Sankoff, Claude W. de Pamphilis, P. Kerr Wall, and Pamela S. Soltis. 2009. "Polyploidy and Angiosperm Diversification." *American Journal of Botany* 96 (1): 336–48. <https://doi.org/10.3732/ajb.0800079>.
- Song, Hui, Pengfei Wang, Changsheng Li, Suoyi Han, Javier Lopez-Baltazar, Xinyou Zhang, and Xingjun Wang. 2016. "Identification of Lipoxygenase (LOX) Genes from Legumes and Their Responses in Wild Type and Cultivated Peanut upon Aspergillus Flavus Infection." *Scientific Reports* 6 (1): 35245. <https://doi.org/10.1038/srep35245>.
- Song, Qing-Xin, Qing-Tian Li, Yun-Feng Liu, Feng-Xia Zhang, Biao Ma, Wan-Ke Zhang, Wei-Qun Man, et al. 2013. "Soybean GmbZIP123 Gene Enhances Lipid Content in the Seeds of Transgenic Arabidopsis Plants." *Journal of Experimental Botany* 64 (14): 4329–41. <https://doi.org/10.1093/jxb/ert238>.
- Song, Youhong, Liang He, Xin-Ding Wang, Nathan Smith, Simon Wheeler, Manohar L. Garg, and Ray J. Rose. 2017. "Regulation of Carbon Partitioning in the Seed of the Model Legume Medicago Truncatula and Medicago Orbicularis: A Comparative Approach." *Frontiers in Plant Science* 8. <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2017.02070>.
- Stefanović, Saša, Bernard E. Pfeil, Jeffrey D. Palmer, and Jeff J. Doyle. 2009. "Relationships among Phaseoloid Legumes Based on Sequences from Eight Chloroplast Regions." *Systematic Botany* 34 (1): 115–28.
- Su, Hong-Gang, Xiao-Hong Zhang, Ting-Ting Wang, Wen-Liang Wei, Yan-Xia Wang, Jun Chen, Yong-Bin Zhou, et al. 2020. "Genome-Wide Identification, Evolution, and Expression of GDSL-Type Esterase/Lipase Gene Family in Soybean." *Frontiers in Plant Science* 11. <https://www.frontiersin.org/articles/10.3389/fpls.2020.00726>.
- Sun, RuHao, Li Gao, Zhiqi Mi, Yusheng Zheng, and Dongdong Li. 2020. "CnMADS1, a MADS Transcription Factor, Positively Modulates Cell Proliferation and Lipid Metabolism in the Endosperm of Coconut (Cocos Nucifera L.)." *Planta* 252 (5): 83. <https://doi.org/10.1007/s00425-020-03490-3>.
- Sutivisedsak, Nongnuch, Bryan R. Moser, Brajendra K. Sharma, Roque L. Evangelista, Huai N. Cheng, William C. Lesch, Robert R. Tangsrud, and Atanu Biswas. 2011. "Physical Properties and Fatty Acid Profiles of Oils from Black, Kidney, Great Northern, and Pinto Beans." *Journal of the American Oil Chemists' Society* 88 (2): 193–200. <https://doi.org/10.1007/s11746-010-1669-8>.
- Tang, Haibao, Vivek Krishnakumar, Shelby Bidwell, Benjamin Rosen, Agnes Chan, Shiguo Zhou, Laurent Gentzbittel, et al. 2014. "An Improved Genome Release (Version Mt4.0) for the Model Legume Medicago Truncatula." *BMC Genomics* 15 (1): 312. <https://doi.org/10.1186/1471-2164-15-312>.
- Tasma, I. Made, Volker Brendel, Steven A. Whitham, and Madan K. Bhattacharyya. 2008. "Expression and Evolution of the Phosphoinositide-Specific Phospholipase C Gene Family in Arabidopsis Thaliana." *Plant Physiology and Biochemistry: PPB* 46 (7): 627–37.

- <https://doi.org/10.1016/j.plaphy.2008.04.015>.
- Taurino, Marco, Sara Costantini, Stefania De Domenico, Francesco Stefanelli, Guillermo Ruano, María Otilia Delgadillo, José Juan Sánchez-Serrano, Maite Sanmartín, Angelo Santino, and Enrique Rojo. 2018. "SEIPIN Proteins Mediate Lipid Droplet Biogenesis to Promote Pollen Transmission and Reduce Seed Dormancy." *Plant Physiology* 176 (2): 1531–46. <https://doi.org/10.1104/pp.17.01430>.
- Thorne, Robert F. 2002. "How Many Species of Seed Plants Are There?" *TAXON* 51 (3): 511–12. <https://doi.org/10.2307/1554864>.
- Torabi, Sepideh, Arjun Sukumaran, Sangeeta Dhaubhadel, Sarah E. Johnson, Peter LaFayette, Wayne A. Parrott, Istvan Rajcan, and Milad Eskandari. 2021. "Effects of Type I Diacylglycerol O-Acyltransferase (DGAT1) Genes on Soybean (Glycine Max L.) Seed Composition." *Scientific Reports* 11 (1): 2556. <https://doi.org/10.1038/s41598-021-82131-5>.
- Turquetti-Moraes, Dayana K., Kanhu C. Moharana, Fabricio Almeida-Silva, Francisnei Pedrosa-Silva, and Thiago M. Venancio. 2022. "Integrating Omics Approaches to Discover and Prioritize Candidate Genes Involved in Oil Biosynthesis in Soybean." *Gene* 808 (January): 145976. <https://doi.org/10.1016/j.gene.2021.145976>.
- Valliyodan, Babu, Steven B. Cannon, Philipp E. Bayer, Shengqiang Shu, Anne V. Brown, Longhui Ren, Jerry Jenkins, et al. 2019. "Construction and Comparison of Three Reference-Quality Genome Assemblies for Soybean." *The Plant Journal* 100 (5): 1066–82. <https://doi.org/10.1111/tpj.14500>.
- Van de Peer, Yves, Eshchar Mizrahi, and Kathleen Marchal. 2017. "The Evolutionary Significance of Polyploidy." *Nature Reviews Genetics* 18 (7): 411–24. <https://doi.org/10.1038/nrg.2017.26>.
- Van Bel, Michiel, Francesca Silvestri, Eric M Weitz, Lukasz Kreft, Alexander Botzki, Frederik Coppens, and Klaas Vandepoele. 2022. "PLAZA 5.0: Extending the Scope and Power of Comparative and Functional Genomics in Plants." *Nucleic Acids Research* 50 (D1): D1468–74. <https://doi.org/10.1093/nar/gkab1024>.
- Vision, Todd. 2010. "The Dryad Digital Repository: Published Evolutionary Data as Part of the Greater Data Ecosystem." *Nature Precedings*, June, 1–1. <https://doi.org/10.1038/npre.2010.4595.1>.
- Wakelam, Michael J. O. 1998. "Diacylglycerol – When Is It an Intracellular Messenger?" *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1436 (1): 117–26. [https://doi.org/10.1016/S0005-2760\(98\)00123-4](https://doi.org/10.1016/S0005-2760(98)00123-4).
- Wang, Jie, Huaqin Kuang, Zhihui Zhang, Yongqing Yang, Long Yan, Mengchen Zhang, Shikui Song, and Yuefeng Guan. 2020. "Generation of Seed Lipooxygenase-Free Soybean Using CRISPR-Cas9." *The Crop Journal*, Crop genome editing: a big step toward breeding by design, 8 (3): 432–39. <https://doi.org/10.1016/j.cj.2019.08.008>.
- Wang, Shoudong, Shulin Liu, Jie Wang, Kengo Yokosho, Bin Zhou, Ya-Chi Yu, Zhi Liu, et al. 2020. "Simultaneous Changes in Seed Size, Oil Content and Protein Content Driven by Selection of SWEET Homologues during Soybean Domestication." *National Science Review* 7 (11): 1776–86. <https://doi.org/10.1093/nsr/nwaa110>.
- Wang, Xinyu, Qingying Li, Qi Zhang, Shiyu Huang, Jingyao Yu, Hongtao Qin, Huidong Qi, et al. 2019. "Identification of Soybean Genes Related to Fatty Acid Content Based on a Soybean Genome Collinearity Analysis." *Plant Breeding* 138 (6): 696–707. <https://doi.org/10.1111/pbr.12733>.
- Xu, Yang, Kristian Mark P. Caldo, Dipasmita Pal-Nath, Jocelyn Ozga, M. Joanne Lemieux, Randall J. Weselake, and Guanqun Chen. 2018. "Properties and Biotechnological Applications of Acyl-CoA:Diacylglycerol Acyltransferase and Phospholipid:Diacylglycerol Acyltransferase from Terrestrial Plants and Microalgae." *Lipids* 53 (7): 663–88. <https://doi.org/10.1002/lipd.12081>.
- Yanai, Itai, Hila Benjamin, Michael Shmoish, Vered Chalifa-Caspi, Maxim Shklar, Ron Ophir, Arren Bar-Even, et al. 2005. "Genome-Wide Midrange Transcription Profiles Reveal Expression Level Relationships in Human Tissue Specification." *Bioinformatics* 21 (5): 650–59. <https://doi.org/10.1093/bioinformatics/bti042>.
- Yang, Hongyan, Wubin Wang, Qingyuan He, Shihua Xiang, Dong Tian, Tuanjie Zhao, and Junyi Gai.

2019. "Identifying a Wild Allele Conferring Small Seed Size, High Protein Content and Low Oil Content Using Chromosome Segment Substitution Lines in Soybean." *Theoretical and Applied Genetics* 132 (10): 2793–2807. <https://doi.org/10.1007/s00122-019-03388-z>.
- Yang, Kai, Zhixi Tian, Chunhai Chen, Longhai Luo, Bo Zhao, Zhuo Wang, Lili Yu, et al. 2015. "Genome Sequencing of Adzuki Bean (*Vigna Angularis*) Provides Insight into High Starch and Low Fat Accumulation and Domestication." *Proceedings of the National Academy of Sciences* 112 (43): 13213–18. <https://doi.org/10.1073/pnas.1420949112>.
- Yang, Wenyu, Geliang Wang, Jia Li, Philip D. Bates, Xuemin Wang, and Doug K. Allen. 2017. "Phospholipase D ζ Enhances Diacylglycerol Flux into Triacylglycerol." *Plant Physiology* 174 (1): 110–23. <https://doi.org/10.1104/pp.17.00026>.
- Yao, Yanjie, Qingbo You, Guozhan Duan, Jianjun Ren, Shanshan Chu, Junqing Zhao, Xia Li, Xinan Zhou, and Yongqing Jiao. 2020. "Quantitative Trait Loci Analysis of Seed Oil Content and Composition of Wild and Cultivated Soybean." *BMC Plant Biology* 20 (1): 51. <https://doi.org/10.1186/s12870-019-2199-7>.
- Yeap, Wan-Chin, Fong-Chin Lee, Dilip Kumar Shabari Shan, Hamidah Musa, David Ross Appleton, and Harikrishna Kulaveerasingam. 2017. "WRI1-1, ABI5, NF-YA3 and NF-YC2 Increase Oil Biosynthesis in Coordination with Hormonal Signaling during Fruit Development in Oil Palm." *The Plant Journal* 91 (1): 97–113. <https://doi.org/10.1111/tpj.13549>.
- Zeng, Xuanrui, Hailun Liu, Hongyang Du, Sujing Wang, Wenming Yang, Yingjun Chi, Jiao Wang, Fang Huang, and Deyue Yu. 2018. "Soybean MADS-Box Gene GmAGL1 Promotes Flowering via the Photoperiod Pathway." *BMC Genomics* 19 (1): 51. <https://doi.org/10.1186/s12864-017-4402-2>.
- Zhang, Bin, and Steve Horvath. 2005. "A General Framework for Weighted Gene Co-Expression Network Analysis." *Statistical Applications in Genetics and Molecular Biology* 4 (1). <https://doi.org/10.2202/1544-6115.1128>.
- Zhang, Min, Shulin Liu, Zhao Wang, Yaqin Yuan, Zhifang Zhang, Qianjin Liang, Xia Yang, et al. 2022. "Progress in Soybean Functional Genomics over the Past Decade." *Plant Biotechnology Journal* 20 (2): 256–82. <https://doi.org/10.1111/pbi.13682>.
- Zhao, Jiangzhe, Rongrong Bi, Shuxiang Li, Dan Zhou, Yang Bai, Guangqin Jing, Kewei Zhang, and Wenhua Zhang. 2019. "Genome-Wide Analysis and Functional Characterization of Acyl-CoA:Diacylglycerol Acyltransferase from Soybean Identify GmDGAT1A and 1B Roles in Oil Synthesis in Arabidopsis Seeds." *Journal of Plant Physiology* 242 (November): 153019. <https://doi.org/10.1016/j.jplph.2019.153019>.
- Zhao, Shihui, Fan Yan, Yajing Liu, Monan Sun, Ying Wang, Jingwen Li, Xinsheng Zhang, Xuguang Yang, and Qingyu Wang. 2023. "Genome-Wide Identification and Expression Analysis of Diacylglycerol Acyltransferase Genes in Soybean (*Glycine Max*)." *PeerJ* 11 (March): e14941. <https://doi.org/10.7717/peerj.14941>.
- Zhou, Zhengkui, Yu Jiang, Zheng Wang, Zhiheng Gou, Jun Lyu, Weiyu Li, Yanjun Yu, et al. 2015. "Resequencing 302 Wild and Cultivated Accessions Identifies Genes Related to Domestication and Improvement in Soybean." *Nature Biotechnology* 33 (4): 408–14. <https://doi.org/10.1038/nbt.3096>.
- Zhuang, Yongbin, Xutong Wang, Xianchong Li, Junmei Hu, Lichuan Fan, Jacob B. Landis, Steven B. Cannon, et al. 2022. "Phylogenomics of the Genus *Glycine* Sheds Light on Polyploid Evolution and Life-Strategy Transition." *Nature Plants* 8 (3): 233–44. <https://doi.org/10.1038/s41477-022-01102-4>.
- Zia-Ul-Haq, Muhammad, Mansoor Ahmad, and Shahid Iqbal. 2008. "Characteristics of Oil from Seeds of 4 Mungbean [*Vigna Radiata* (L.) Wilczek] Cultivars Grown in Pakistan." *Journal of the American Oil Chemists' Society* 85 (9): 851–56. <https://doi.org/10.1007/s11746-008-1269-z>.
- Zia-Ul-Haq, Muhammad, Mansoor Ahmad, Shahid Iqbal, Shakeel Ahmad, and Hakoomat Ali. 2007. "Characterization and Compositional Studies of Oil from Seeds of Desi Chickpea (*Cicer Arietinum* L.) Cultivars Grown in Pakistan." *Journal of the American Oil Chemists' Society* 84

(12): 1143–48. <https://doi.org/10.1007/s11746-007-1136-3>.