IMPACTO DA POLIPLOIDIZAÇÃO NOS GENES ASSOCIADOS À PRODUÇÃO DE ÓLEO EM LEGUMINOSAS

DAYANA KELLY TURQUETTI DE MORAES

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE
DARCY RIBEIRO - UENF

CAMPOS DOS GOYTACAZES - RJ
JULHO 2024

IMPACTO DA POLIPLOIDIZAÇÃO NOS GENES ASSOCIADOS À PRODUÇÃO DE ÓLEO EM LEGUMINOSAS

DAYANA KELLY TURQUETTI DE MORAES

Tese apresentada ao Centro de Biociências e Biotecnologia, da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Doutora em Biotecnologia Vegetal

Orientador: Prof. Thiago Motta Venancio

CAMPOS DOS GOYTACAZES - RJ

JULHO 2024

FICHA CATALOGRÁFICA
UENF - Bibliotecas
Elaborada com os dados fornecidos pela autora.

M827 Moraes, Dayana Kelly Turquetti de.

Impacto da poliploidização nos genes associados à produção de óleo em leguminosas / Dayana Kelly Turquetti de Moraes. - Campos dos Goytacazes, RJ, 2024.

103 f. : il.
Inclui bibliografia.

Tese (Doutorado em Biotecnologia Vegetal) - Universidade Estadual do Norte Fluminense Darcy Ribeiro, Centro de Biociências e Biotecnologia, 2024.
Orientador: Thiago Motta Venancio.

1. Metabolismo lipídico. 2. Duplicação gênica . 3. Soja. I. Universidade Estadual do Norte Fluminense Darcy Ribeiro. II. Título.

CDD - 660.6

IMPACTO DA POLIPLOIDIZAÇÃO NOS GENES ASSOCIADOS À PRODUÇÃO DE ÓLEO EM LEGUMINOSAS

DAYANA KELLY TURQUETTI DE MORAES

Tese apresentada ao Centro de Biociências e Biotecnologia, da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Doutora em Biotecnologia Vegetal

Aprovada em 22 de julho de 2024.

Comissão examinadora:

gov.br Documento assinado digitalmente
RENATO VICENTINI DOS SANTOS
Data: 19/08/2024 11:04:51-0300
Verifique em https://validar.iti.gov.br

Dr. Renato Vicentini dos Santos - Unicamp

gov.br Documento assinado digitalmente
VANILDO SILVEIRA
Data: 14/08/2024 16:04:38-0300
Verifique em https://validar.iti.gov.br

Dr. Vanildo Silveira - UENF

gov.br Documento assinado digitalmente
MESSIAS GONZAGA PEREIRA
Data: 14/08/2024 15:02:42-0300
Verifique em https://validar.iti.gov.br

Dr. Messias Gonzaga Pereira - UENF

gov.br Documento assinado digitalmente
THIAGO MOTTA VENANCIO
Data: 14/08/2024 09:55:39-0300
Verifique em https://validar.iti.gov.br

Dr. Thiago Motta Venancio - UENF (Orientador)

Aos grandes amores da minha vida:

meu pai Sérgio, minha mãe Vera e minha irmã Dayene,

dedico.

AGRADECIMENTOS

Primeiramente agradeço a Deus por ser a proteção, o equilíbrio e a direção dos meus passos. De outro modo, sem Ele, não teria a base necessária para trilhar o caminho que me trouxe até aqui. Depois, agradeço por todo o amor, zelo, carinho, dedicação e incentivo da minha família, alicerce fundamental da minha jornada. Em especial, aos meus pais Sérgio Eduardo de Moraes e Vera Lúcia Helena Turquetti de Moraes, devo toda a minha gratidão. Agradeço à minha irmã Dayene Turquetti de Moraes que me proporcionou a amizade e o companheirismo indispensáveis, tornando a caminhada mais leve.

Agradeço ao meu orientador Dr. Thiago Motta Venancio pelo incentivo e confiança na condução das análises, na escrita dos trabalhos e pelas partilhas de vida e visão de futuro. Sou grata ainda pelas ricas trocas de conhecimento e vivência com os meus amigos Francisnei Pedrosa, Késia Dias, Kanhu Moharana, Hemanuel Passareli, Fabrício Almeida, Fabrício Brum, Rajesh Gazara, Isabella Pinheiro, Filipe Matteoli, Cláudio Benício Cardoso, Sarah Jacobs, Kevelin Xavier, Ana Luiza Karl, Gabriel Quintanilha, Daniel Correa, Rebeca Brinco, Amanda Forte, Felipe Rimes, Diego Peña e Hector Vasquez. Cada um de vocês, com suas diferentes personalidades, me forjaram uma pessoa melhor emocional e profissionalmente.

A todos os professores do Programa de Pós Graduação em Biotecnologia Vegetal da UENF, devo minha gratidão e admiração. Em especial, agradeço à Dra. Antônia Elenir Amâncio Oliveira, ao Dr. Antônio Teixeira do Amaral Júnior e ao professor Luiz-Eduardo Del-Bem da Universidade Federal de Minas Gerais (UFMG) por suas contribuições como membros da banca de qualificação do meu doutorado. Além disso, agradeço aos membros da banca de defesa de tese, Dr. Renato Vicentini dos Santos, Dr. Messias Gonzaga Pereira e Dr. Vanildo Silveira pelo aceite e sugestões de melhoria no documento final.

Agradeço ao Programa de Pós-Graduação em Biotecnologia Vegetal da UENF por ter aberto portas para o meu desenvolvimento profissional e pela oportunidade de ter contribuído com o Núcleo de Estudos em Biotecnologia Vegetal - GerminAção, no período de 2022.2 a 2023.1. Agradeço ainda à atual coordenadora do PPGBV, Dra. Claudete Santa Catarina, por todo o apoio na organização e participação em eventos científicos e pela atenção dedicada às demandas discentes do programa. Por fim, agradeço à UENF e aos seus cursos de educação à distância do consórcio CEDERJ, ponte que me proporcionou experiências frutíferas e me conduziu à conclusão desta longa jornada em minha formação acadêmica. Agradeço também às agências de fomento CAPES, FAPERJ e CNPQ que forneceram os recursos necessários para que a minha formação fosse possível.

SUMÁRIO

RESUMO

Com o crescimento da população mundial e o aumento da industrialização, a demanda por óleos vegetais está se expandindo rapidamente. Além de serem componentes nutricionais essenciais na dieta humana e animal, os óleos vegetais são importantes matérias-primas para a produção industrial de lubrificantes, biodiesel e cosméticos. Eles são ricos em ácidos graxos insaturados e essenciais. Entre as fontes mais relevantes de óleo vegetal estão a soja, palma, milho, canola, oliva, girassol, colza e o amendoim. Há mais de vinte anos, a publicação do primeiro genoma de planta - *Arabidopsis,* um organismo modelo - impulsionou a pesquisa básica sobre vias metabólicas e mecanismos de biossíntese de óleo. Dada a relevância socioeconômica do óleo de soja, foram realizados muitos estudos ao longo das décadas para investigar os genes-chave envolvidos na regulação e síntese do óleo, visando melhorar sua qualidade e aumentar seu conteúdo nos grãos. Embora importantes genes, como fatores de transcrição e aqueles envolvidos no transporte de lipídeos e na biossíntese de óleo, tenham sido identificados e até mesmo manipulados para aumentar o conteúdo e melhorar a qualidade do óleo, ainda existem aspectos que não são completamente compreendidos. Por exemplo, o impacto dos eventos de duplicação gênica em genes chave associados ao metabolismo lipídico. Esta tese está estruturada em uma introdução geral, seguida por dois capítulos. O primeiro, aborda uma revisão de literatura sobre os principais genes envolvidos em vias essenciais para a biossíntese de óleo em soja e o segundo consiste em um artigo científico que, essencialmente, discute os processos evolutivos que conduziram ao enriquecimento de óleo em soja.

Palavras chave: metabolismo lipídico, duplicação gênica, soja.

ABSTRACT

With the growth of the global population and the increase in industrialization, the demand for vegetable oils is rapidly expanding. In addition to being essential nutritional components in human and animal diets, vegetable oils are important raw materials for the industrial production of lubricants, biodiesel, and cosmetics. They are rich in unsaturated and essential fatty acids. Among the most relevant sources of vegetable oil are soybean, palm, corn, canola, olive, sunflower, rapeseed, and peanut. Over twenty years ago, the publication of the first plant genome - of Arabidopsis, a model organism - initiated basic research on metabolic pathways and oil biosynthesis mechanisms. Given the socio economic relevance of soybean oil, many studies have been conducted over the decades to investigate the key genes involved in oil regulation and synthesis, aiming to improve its quality and increase its content in the grains. Although important genes, such as transcription factors and those involved in lipid transport and oil biosynthesis, have been identified and even manipulated to increase oil content and improve quality, there are still aspects that are not completely understood. For example, the impact of gene duplication events on key genes associated with lipid metabolism. This thesis is structured with a general introduction, followed by two chapters. The first addresses a literature review on the main genes involved in essential pathways for oil biosynthesis, especially in soybeans, and the second consists of a scientific article that essentially discusses the evolutionary processes that led to the enrichment of oil in soybeans.

Keywords: lipid metabolism, gene duplication, soybean.

ÍNDICE DE FIGURAS

ÍNDICE DE TABELAS

ÍNDICE DE ABREVIATURAS E SIGLAS

[16:0]	Ácido palmítico
[18:0]	Ácido esteárico
[18:1]	Ácido oleico
[18:2]	Ácido linoleico
[18:3]	Ácido linolênico
2-MAG	*2-monoacilglycerol acylhydrolase*
AAPT	*Alcohol phosphotransferase*
ABCA9	*ATP-Binding Cassette, Sub-Family A9*
ABI3	*Abscisic acid insensitive 3*
ACBPs	*Acyl-CoA binding proteins*
ACCase	*Acetyl-Coenzyme A carboxylase*
Acetil-CoA	*Acetyl-coenzyme A*
ACP	*Acyl carrier protein*
AGIs	Ácidos graxos insaturados
AGLs	Ácidos graxos livres
AGmi	Ácido graxo monoinsaturado
AGpi	Ácido graxo poli-insaturado
AGs	Ácidos graxos
AGsat	Ácido graxo saturado
AT	Acetil-CoA–ACP transferase
ATP	Adenosina trifosfato
BASS2	*Bile acid sodium symporter family protein 2* - transportador de piruvato
BC	Biotina Carboxilase
BCCP	Proteína carreadora de biotina carboxilase
CoA	*Coenzyme A*
COs	Corpos oleaginosos
CPT	*Cytidine diphosphate-choline:diacylglycerol cholinephosphotransferase*
CRDR	Ciclo de Condensação, Redução, Desidratação, Redução
CT	*Carboxyl transferase*
DAG	*Diacylglycerol*
DAG1	DAG formado a partir da via de Kennedy
DAG2	DAG derivado da fosfatidilcolina
DGAT	*Acyl-coa: diacylglycerol acyltransferase*
DH	Desidratase
EIM	Espaço intermembrana
ER	Enoil-ACP redutase
FA	*Fatty Acid*

FAD2	*Delta(12)-fatty-acid desaturase*
FAD3	*Omega-3 fatty acid desaturase*
FAE	*Fatty Acid Elongase*
FAS	*Fatty Acid Synthase*
FATs	*Fatty Acyl-ACP thioesterase*
FAX	*Fatty acid export*
FUS3	*FUSCA3*
G3P	*Glycerol-3-phosphate*
G6P	*Glucose-6-phosphate*
Gma	*Glycine max*
GmbZIP123	*Basic leucine zipper 123*
GmDof	*DNA binding with one finger*
GmDREBL	*Dehydration-Responsive Element Binding-like*
GmMFT	*Mother of FT and TFL1*
GmZF351	*Zinc Finger 351*
GmZF392	*Zinc Finger 392*
GPAT	*Acyl-CoA: Glycerol-3-phosphate acyltransferase*
GPT	*Glucose-6-phosphate translocase*
HD	Beta-hidroxiacil-ACP desidratase
HSL	H*ormone-sensitive lipase*
KASs	*3-ketoacil-ACP synthases*
KR	Beta-cetoacil-ACP redutase
Ks	Taxa de substituição sinônima por sítio sinônimo entre parálogos
KS	*ß-ketoacyl-ACP synthase*
LACs9	*Long-chain acyl-CoA synthetase 9*
LEC	*Leafy cotyledon*
LM	Lipídeos de membrana
LPA	L*ysophosphatidic acid*
LPAAT	*Acyl-CoA: lysophosphatidic acid acyltransferase*
LPC	*Lysophosphatidylcholine*
LPCAT	*Lysophosphatidylcholine acyltransferase*
ME	Membrana externa
MI	Membrana interna
MT	Malonil-ACP transferase
PA	P*hosphatidic acid*
PAP	*Phosphatidic Acid Phosphatase*
PC	*Phosphatidylcholine*
PDAT	*Phospholipid: diacylglycerol acyltransferase*
PDCT	P*hosphatidylcholine diacylglycerol cholinephosphotransferase*

PDH	*Pyruvate dehydrogenase*
PDHm	*Mitochondrial pyruvate dehydrogenase*
PDHp	*Plastidial pyruvate dehydrogenase*
PEP	*Phosphoenolpyruvate*
PK	Pyruvate kinase
PKc	*Cytoplasmic pyruvate kinase*
PKp	*Plastidial pyruvate kinase*
PLC	*Phospholipase C*
PLD	*Phospholipase D*
PPT	*Phosphoenolpyruvate/phosphate translocator*
Pvu	*Phaseolus vulgaris*
Pyr	*Pyruvate*
R1 a R8	Estágios reprodutivos durante o desenvolvimento da semente de soja
RE	Retículo endoplasmático
SAD	*Stearoyl-ACP desaturase*
SE	*Seed expression*
Seed-1	*Stage 1 seeds, 50 mg*
Seed-2	*Stage 2 seeds, 150 mg*
Seed-H	*Heart stage seeds, 7 mg*
Suc	*Sucrose*
TAG	*Triacylglycerol*
TAG Deg	*TAG degradation pathway*
TAG Syn	*TAG synthesis pathway*
TCA	*Tricarboxylic acid cycle*
TF	*Transcription factor*
TGL1	T*riacylglycerol lipase*
TPM	*Transcripts per million*
WGD	*Whole Genome Duplication*
WRI1	*WRINKLED 1*

1 - INTRODUÇÃO GERAL

1.1 Poliploidização

A poliploidização ou duplicação do genoma inteiro (do inglês, *Whole Genome Duplication* - WGD), envolve a fusão de dois ou mais conjuntos completos de cromossomos no núcleo celular. Este fenômeno pode ocorrer tanto dentro de uma única espécie, conhecido como autopoliploidia (e.g. xx para xxxx), quanto entre diferentes espécies próximas, denominado alopoliploidia (e.g. xx + yy para xxyy). Em geral, os poliploides podem surgir pela duplicação de células somáticas ou pela fusão de gametas citologicamente não reduzidos, sendo essa última a forma mais comum na natureza (Otto, 2007). Gametas não reduzidos surgem devido a anomalias durante a meiose (Figura 1), onde a redução do número de cromossomos não ocorre (Ramsey & Ramsey, 2014).

A meiose em autopoliploides e alopoliploides (Figura 1) apresenta diferenças significativas devido às suas distintas origens e organização cromossômica. Em autopoliploides, a meiose é caracterizada por complicações durante o pareamento cromossômico. Isso ocorre porque a presença de múltiplos conjuntos homólogos pode levar a configurações de pareamento como bivalentes, trivalentes ou quadrivalentes. A segregação dos cromossomos homólogos pode ser desigual, resultando em células-filhas com um número variável de cromossomos, o que pode causar desequilíbrios genéticos. Como consequência, a meiose irregular em autopoliploides frequentemente leva à formação de gametas aneuploides, que têm um número anormal de cromossomos, afetando a viabilidade e a fertilidade dos organismos (Mason & Pires, 2015). Em contraste, nos alopoliploides, os conjuntos cromossômicos são homeólogos, ou seja, semelhantes mas não idênticos. Durante a meiose, os cromossomos homeólogos geralmente não se emparelham devido às diferenças nas sequências de DNA, resultando no pareamento preferencial de cromossomos dentro de cada conjunto parental. Essa característica leva a uma segregação cromossômica mais regular em alopoliploides do que em autopoliploides. Como resultado, uma maior proporção de gametas euploides (com o número correto de cromossomos) é produzida, promovendo a estabilidade genética e a fertilidade dos alopoliploides (Adams et al., 2003; Yant & Bomblies, 2017). Devido a maior regularidade da meiose, os alopoliploides geralmente são mais férteis do que os autopoliploides (Yant & Bomblies, 2017).

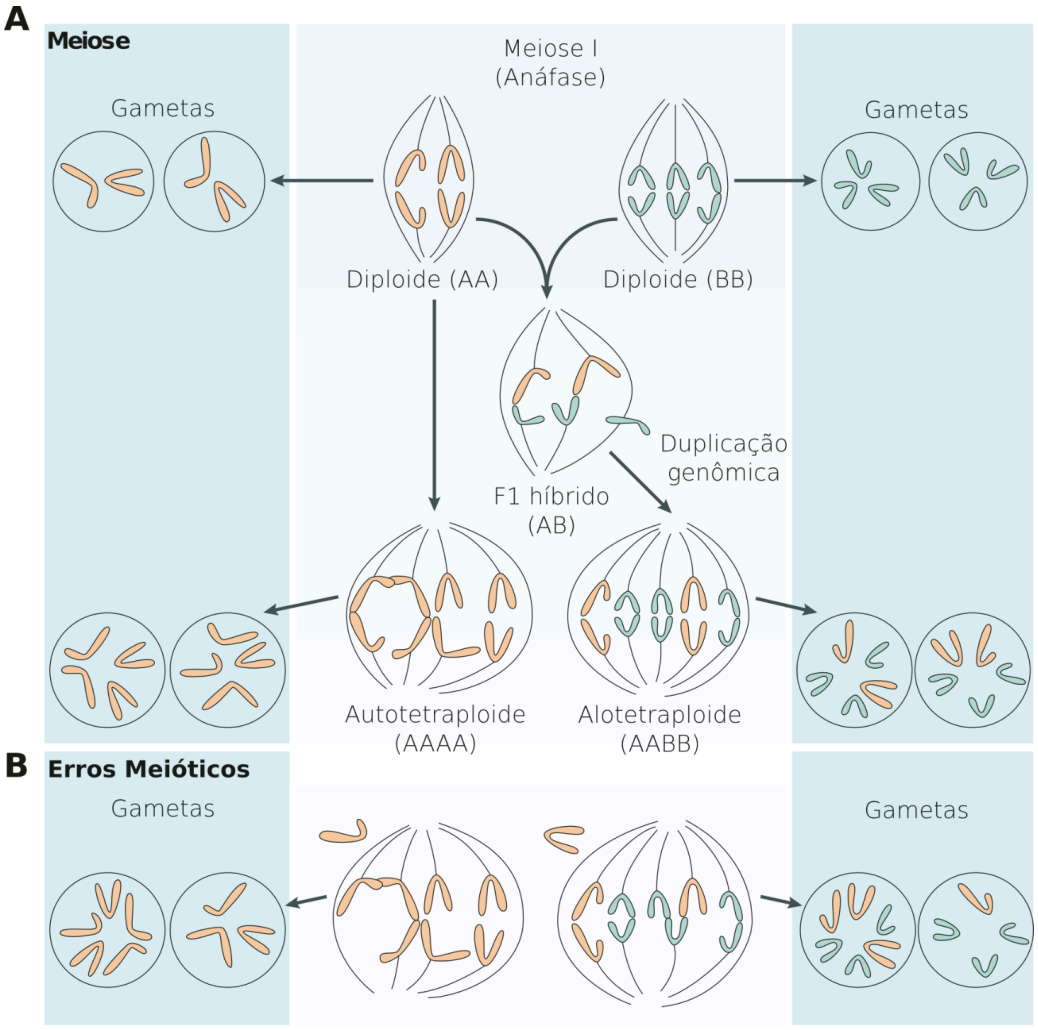


Figura 1. Poliploidização e irregularidades na divisão meiótica. **A**. Composição cromossomal e o comportamento de diploides formando poliploides. A imagem ilustra os estágios iniciais da anáfase I com a separação dos pares de cromossomos (painel central) e os gametas (painéis laterais). O emparelhamento dos cromossomos homólogos é defeituoso no híbrido F1 devido à divergência na estrutura e no número de cromossomos. O emparelhamento é restaurado pela duplicação do genoma, o que produz um alopoliploide, onde os dois conjuntos de cromossomos homeólogos se emparelham independentemente. No autopoliploide, o emparelhamento envolve frequentes multivalentes, uma vez que quatro cromossomos de cada tipo estão presentes. **B**. Irregularidades meióticas resultando em cromossomos retardatários, i.e., que não se movem corretamente para os pólos opostos da célula (à esquerda) e gametas aneuploides (à direita). A estrutura de duas cromátides na anáfase foi omitida para simplificar a representação. Adaptado de Comai (Comai, 2005).

Após cada evento de WGD, deleções, divergências e rearranjos dos genes duplicados restabelecem o estado diploide do genoma, ou seja, há uma exigência de equilíbrio genético (Lynch & Conery, 2000; Qiao et al., 2019). Sob a seleção natural, as duplicações gênicas podem seguir diferentes destinos, a saber: pseudogenização (perda de função), neofuncionalização (aquisição de novas funções) ou subfuncionalização (divisão das funções originais entre as cópias gênicas duplicadas) (Cheng et al., 2018; Freeling et al., 2015). Tanto a especialização imediata (Adams et al., 2003) quanto a subfuncionalização (Lynch & Force, 2000) podem impedir que genes duplicados acumulem

mutações que eliminem sua função, uma vez que cada cópia gênica é necessária em algum momento ou local do organismo. Por exemplo, genes pleiotrópicos podem passar a tolerar mais mutações sobre as quais a seleção natural pode agir para favorecer novas funções. Ademais, análises moleculares demonstram que eventos genômicos pós WGD incluem mudanças genéticas e epigenéticas, mudanças na expressão gênica e variações fenotípicas (Gaeta & Pires, 2010; K. D. Kim et al., 2015; L. Wang et al., 2021). O pareamento de homólogos e a recombinação são os mecanismos chave que contribuem para a variabilidade genética[1]. Consequentemente, a compreensão dos efeitos de eventos de WGD sobre os processos fisiológicos e bioquímicos das plantas é essencial para elucidar sua influência na biologia vegetal.

Na história evolutiva das plantas a poliploidia tem sido considerada um importante mecanismo para especiação e também para a adaptabilidade em mudanças ambientais (Shimizu-Inatsugi et al., 2017). De acordo com Van de Peer et al. (2009), a vantagem competitiva dos poliploides sobre seus progenitores diploides está associada principalmente à segregação transgressiva, que envolve a formação de fenótipos extremos e um aumento do vigor (Van de Peer et al., 2009). Os métodos para identificar eventos de WGD são muitos e variados, por exemplo: distribuição da taxa de substituição sinônima por sítio sinônimo entre parálogos (K_s) onde um pico na distribuição é interpretado como um evento de WGD (Lynch & Conery, 2000; Vanneste et al., 2013), filogenômica (Jiao et al., 2011), tamanho do genoma, cariótipo, análises de número de cópias de genes (Lynch & Conery, 2000; Vanneste et al., 2013) e sintenia (Tang et al., 2008). Por meio de abordagens filogenéticas, identificou-se que várias linhagens de plantas diferentes experimentaram eventos de WGD independentemente em torno da fronteira cretáceo-paleógeno (Figura 2). Todas as plantas com sementes experimentaram pelo menos um evento de WGD em sua história evolutiva, caracterizando uma ancestralidade paleopoliploide (Fawcett et al., 2009; Renny-Byfield & Wendel, 2014). Além disso, outro evento de WGD é compartilhado por todas as plantas com flores (angiospermas) (Jiao et al., 2011), e um WGD adicional é compartilhado por vários clados principais de plantas com flores, incluindo as monocotiledôneas (Jiao et al., 2014), eudicotiledôneas (Jiao et al., 2012), Asteraceae (C.-H. Huang et al., 2016), Brassicales (Kagale et al., 2014), Gramíneas (Estep et al., 2014) e as Fabaceae (leguminosas) (Cannon et al., 2015). Esses resultados levantam a hipótese de que a duplicação do genoma pode ter conferido resistência à extinção para certas linhagens de plantas, permitindo que

[1]A variabilidade genética só pode alimentar a evolução de uma população poliploide se os indivíduos conseguirem sobreviver às mutações genômicas. Se a taxa de mutação for muito alta e os efeitos de aptidão das mutações forem muito graves, a extinção será o resultado provável.

genomas poliploides se tornassem dominantes após episódios de extinção em massa (Figura 2).

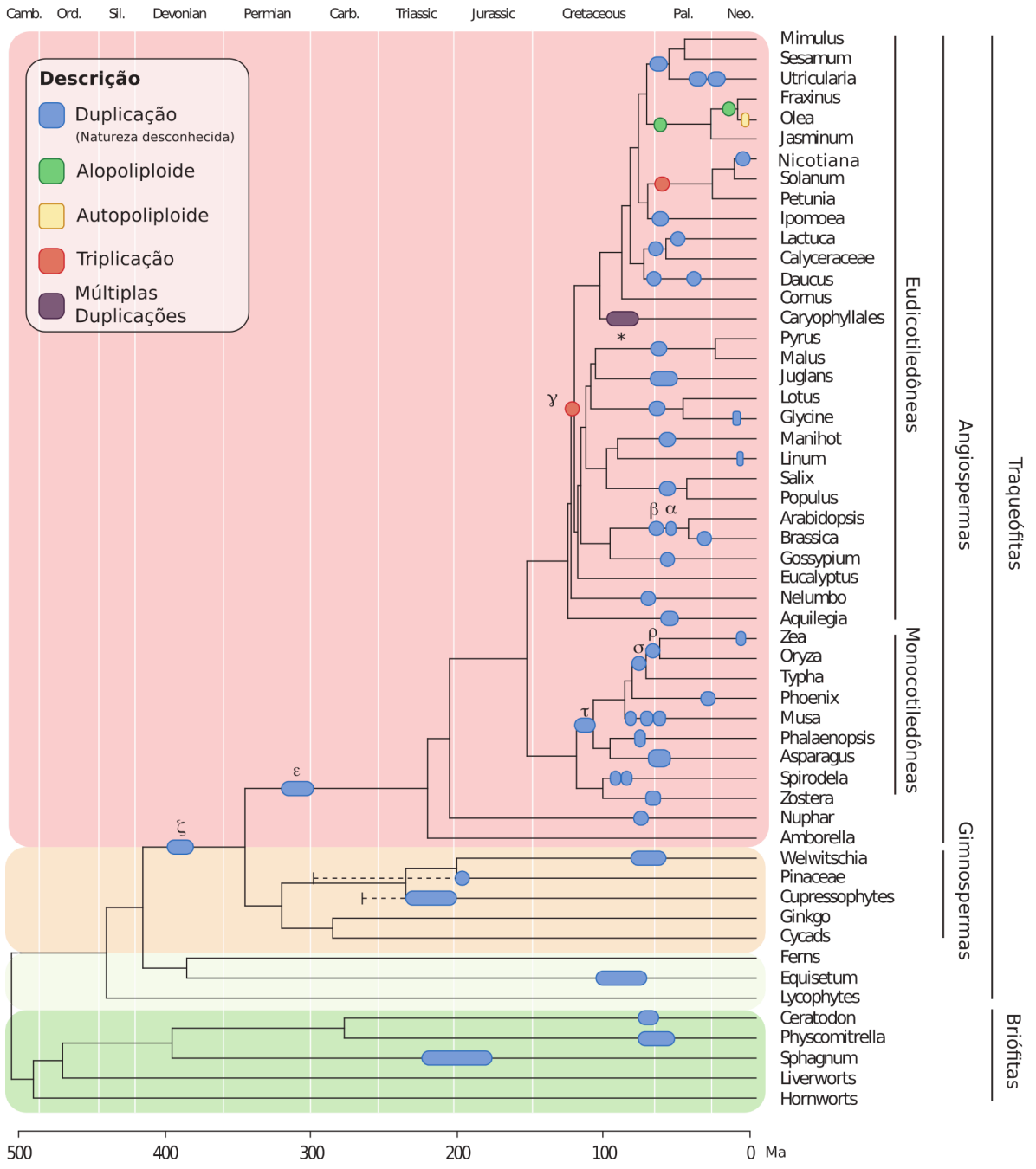


Figura 2. Eventos de duplicação do genoma inteiro no reino vegetal (Plantae). Eventos de duplicação de origem desconhecida são mostrados em azul, triplicações em vermelho, eventos de autopoliploidia em amarelo e eventos de alopoliploidia em verde. O asterisco associado a *Caryophyllales* representa 26 eventos de WGD independentes. Os eventos de duplicação nomeados são mostrados ao lado de sua letra grega correspondente. Abreviações: **Camb**., Cambriano; **Carb**., Carbonífero; **Ord**., Ordoviciano; **Neo**., Neogênico; **Pal**., Paleozoico; **Sil**., Siluriano. Imagem adaptada de Clark & Donoghue (Clark & Donoghue, 2018).

1.2 As leguminosas

As leguminosas pertencem à família Fabaceae, terceiro maior grupo das angiospermas, com cerca de 770 gêneros e 19.500 espécies com grande variação morfológica e ecológica (Azani et al., 2017). Alguns representantes da família são: amendoim (*Arachis hypogaea*), feijão (*Phaseolus vulgaris*), grão-de-bico (*Cicer arietinum*) e soja (*Glycine max*). Além da relevância nutricional e econômica (Grdeń & Jakubczyk, 2023; Rotundo et al., 2024), as leguminosas são conhecidas pelo estabelecimento de interações simbióticas com bactérias diazotróficas do gênero *Rizhobium* (Roy et al., 2020). Essa associação desempenha um papel crucial na fixação biológica de nitrogênio, característica de grande importância agronômica, uma vez que o nitrogênio é um macronutriente essencial para o crescimento e desenvolvimento das plantas (Herridge et al., 2008).

A soja corresponde a mais da metade da produção mundial de oleaginosas destinadas à alimentação humana e animal, além de ser importante em aplicações industriais e farmacêuticas, bem como na produção de biodiesel (Chaudhary et al., 2015; Zhou et al., 2015). É uma leguminosa amplamente cultivada, notável por seu alto teor proteico (40%) e por conter todos os aminoácidos essenciais, além de ser rica em fibras, vitaminas (e.g. do complexo B), minerais (e.g. ferro, cálcio e magnésio) e possuir compostos bioativos como as isoflavonas, que têm propriedades antioxidantes que podem trazer benefícios à saúde (P. Qin et al., 2022). Ademais, destaca-se pela alta concentração de óleo (22%) (Messina & Messina, 2010), característica que vem sendo amplamente estudada visando melhorar a qualidade do óleo no grão (H. Song et al., 2023). A depender da cultivar e das condições de cultivo, as concentrações de óleo podem variar de 8.3% a 27.9% (Wilson, 2004). Atualmente, o Brasil é o maior produtor e exportador de soja, seguido por Estados Unidos e Argentina. As estimativas nacionais de área plantada são de 45 milhões de hectares e a produção da soja, safra 2023/2024, chega a mais de 147 milhões de toneladas com produtividade média de 3.229 kg/hectare (Conab, 2024).

Há evidências de que a soja tenha passado por dois eventos de WGD: o mais remoto há 58 milhões de anos (compartilhado por todas as leguminosas - Tabela 1) e o mais recente tendo ocorrido há 13 milhões de anos (exclusivo do gênero *Glycine*) (Gill et al., 2009). Nessa perspectiva, Schmutz et al. (2010), sequenciaram os 20 cromossomos da soja (Tabela 1) com uma montagem que cobriu aproximadamente 85% do genoma estimado (955 Mb), reportando que 75% dos genes codificadores de proteína do genoma de referência (cv. Williams 82) estão presentes em duas ou mais cópias (Schmutz et al., 2010). Anos depois, o sequenciamento do genoma do feijão (*P. vulgaris*) revelou que 91% dos seus genes estão localizados em blocos sintênicos com a soja (Schmutz et al., 2014).

Essas duas espécies pertencem a grupos irmãos, que divergiram há aproximadamente 19 milhões de anos e possuem composições distintas de macronutrientes em seus grãos. Por exemplo, a soja possui aproximadamente 40% de proteínas, 25% de carboidratos e 22% de lipídeos, enquanto o feijão contém cerca de 21% de proteínas, 60% de carboidratos e 2% de lipídeos (Celmeli et al., 2018; Zhou et al., 2019). Pelo evento de WGD exclusivo do gênero *Glycine*, a comparação destes genomas é uma grande oportunidade para o estudo evolutivo dos efeitos de WGD na expressão e divergência funcional de genes homólogos associados a características específicas de interesse agronômico, por exemplo, o metabolismo lipídico.

Tabela 1. Ploidia e registro do último evento de duplicação do genoma inteiro (WGD) em seis espécies de leguminosas

Espécie	Ploidia (n)	WGD (ma)	Referências
Arachis hypogaea	20	3.5	(Seijo et al., 2007)
Glycine max	20	13	(Schmutz et al., 2010)
Phaseolus vulgaris	11	58	(Schmutz et al., 2014)
Medicago truncatula	8	58	(Tang et al., 2014)
Cicer arietinum	8	58	(Varshney et al., 2013)
Pisum sativum	7	58	(Kreplak et al., 2019)

2 - CAPÍTULO 1. REVISÃO DE LITERATURA

BIOSSÍNTESE DE ÓLEO EM SOJA

2.1 Desenvolvimento e acúmulo de reservas na semente de soja

De acordo com o modelo proposto por Fehr e Caviness, os estágios de desenvolvimento da soja são classificados em vegetativo e reprodutivo (Fehr & Caviness, 1977). No estágio vegetativo, a planta de soja passará pelo desenvolvimento das folhas, tornando-se autossuficiente na produção de energia por meio da fotossíntese (Figura 3.A). No estágio reprodutivo, que varia de R1 a R8, a planta é induzida ao florescimento por estímulos ambientais, englobando a floração, o desenvolvimento das vagens, das sementes e a maturação (Figura 3.B).

Em R1 há o início da floração, onde é observada a formação de uma flor em qualquer nó; em R2 observa-se a floração completa; em R3 há o início da formação das vagens; em R4 observa-se a vagem completa. O estágio R5 marca o início da formação das sementes, onde destacam-se os estágios intermediários que acompanham a evolução do acúmulo de reservas. Tal acúmulo é medido de acordo com o percentual de granação máxima nas vagens, a saber: R5.1 – até 10%, as sementes se desligam da planta mãe, cessando a translocação de fotoassimilados; R5.2 – entre 10 e 25%; R5.3 – entre 25 e 50%; R5.4 – entre 50 e 75% e R5.5 – entre 75 e 100%. Durante esses estágios o óleo acumula-se rapidamente, especialmente nos cotilédones, acompanhado por um grande aumento no tamanho das sementes (Q.-X. Song et al., 2013; Y. Song et al., 2017). No estágio R6 é possível observar o pleno enchimento das sementes ainda verdes que ocupam toda a cavidade da vagem. O estágio R7 marca o início da maturidade onde observam-se vagens com sua cor madura e em R8 observa-se a maturidade completa, onde 95% das vagens alcançam sua cor madura (Fehr & Caviness, 1977).

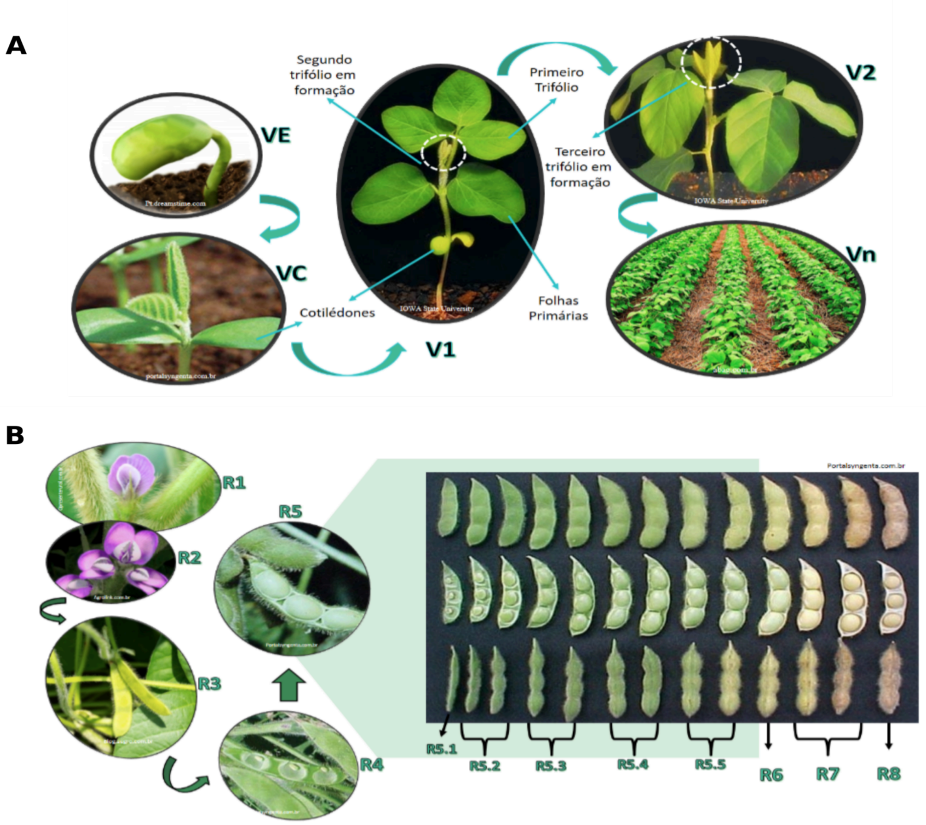


Figura 3. Estágios fenológicos da soja baseados em Fehr e Caviness (1977). **A**. Estágio vegetativo. Em VE os cotilédones estão acima da superfície do solo pelo desdobramento da alça do hipocótilo. É o momento de início do desdobramento e expansão das folhas primárias. Em VC os cotilédones já estão bem desenvolvidos e fornecem nutrientes para a sobrevivência da plântula. Em V1 é possível notar as folhas primárias expandidas e paralelas ao solo. Em um estágio já tardio de V1, como na imagem, é possível verificar o primeiro trifólio bem desenvolvido e o início do desenvolvimento do segundo trifólio. Em V2, na figura em estágio tardio, pode-se observar o pleno desenvolvimento do primeiro e do segundo trifólio e o início do desenvolvimento do terceiro trifólio. Em Vn há a representação de uma plantação de soja no final do estágio vegetativo. A depender da variedade, a planta de soja pode formar até 20 trifólios (V21), esses estágios não são mostrados na figura. **B**. Estágio reprodutivo de R1 a R8. Em R1 e R2 há o período de floração. Em R3 e R4 há formação e pleno desenvolvimento da vagem. Em R5 estão mostrados os subestágios que culminam com o acúmulo de reservas, conhecido como o período de enchimento das sementes que termina em R6. R7 e R8 são os períodos de início da maturação e maturação plena, respectivamente, indicando o ponto da colheita. Abreviações: **VE**, estágio vegetativo de emergência; **VC**, estágio vegetativo cotiledonar; **V1**, primeiro nó; **V2**, segundo nó; **Vn**, enésimo nó; **R1 a R8**, diferentes estágios do desenvolvimento reprodutivo. Fonte: imagens de IOWA State University e Portal Syngenta. Adaptada.

A maturação ou duração do enchimento das sementes é de grande interesse para as culturas agrícolas como a soja. Durante esta fase, uma rede regulatória é ativada para acumular produtos de armazenamento (e.g. proteínas, carboidratos e lipídeos) em vários níveis (Duan et al., 2023). Essas reservas se acumulam durante a maturação das sementes para nutrir a planta jovem durante a germinação, oferecendo um alto valor nutricional e químico. Como resultado, muitas pesquisas têm focado no processo de maturação visando o melhoramento da produção de grãos (Duan et al., 2023; Finkelstein, 2010; Manan et al., 2017; Roscoe et al., 2015; Wobus & Weber, 1999). O metabolismo lipídico é transcricionalmente regulado durante o desenvolvimento da semente. Por exemplo: WRI1,

LEC1, LEC2, FUS3 e ABI3 são os principais fatores de transcrição (TFs, do inglês *transcription factors*) envolvidos na regulação do desenvolvimento da semente e no acúmulo de óleo (Apêndice 1, Figura 4) (Santos-Mendoza et al., 2008). Além deles, outros TFs como GmDof11, GmDof4, GmZF351, GmZF392, GmbZIP123, GmMFT e GmDREBL também foram descritos como importantes reguladores da síntese de lipídeos (Cai et al., 2023; Q.-T. Li et al., 2017; Y.-F. Liu et al., 2014; L. Lu et al., 2021; Manan et al., 2017; Miao et al., 2020; Y.-Q. Zhang et al., 2016). A ação conjunta desses TFs resulta em fluxo de carbono para a síntese de AGs e, consequentemente, leva à produção de triacilglicerol (TAG) (Figura 4).

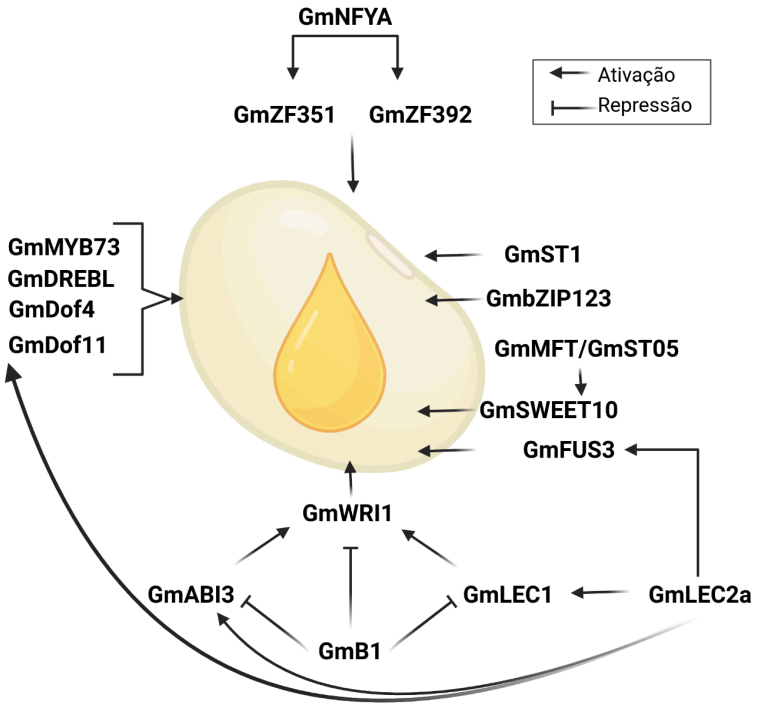


Figura 4. Rede regulatória com os principais fatores de transcrição e genes codificadores de proteínas associados ao acúmulo de óleo em soja. O identificador dos genes apresentados, bem como suas respectivas publicações estão disponíveis no Apêndice 1. Imagem adaptada de Duan et al. (Duan et al., 2023). Criado com BioRender.com.

Além dos processos regulatórios, a composição nutricional da semente de soja é uma característica complexa também influenciada por vários genes e pelo ambiente (Chaudhary et al., 2015; Collakova et al., 2013). A publicação do genoma de referência da soja (da cultivar Williams 82) em 2010 (Schmutz et al., 2010) alavancou o desenvolvimento da genômica funcional da soja (M. Zhang et al., 2022). Apesar disso, alterar a composição ou o conteúdo nutricional desta leguminosa é um processo desafiador, pois, além das possíveis interações genéticas, epigenéticas e ambientais, cerca de 75% dos genes da soja

estão duplicados (Schmutz et al., 2010). Estudos realizados em *Arabidopsis thaliana* demonstraram o envolvimento de mais de 600 genes no metabolismo lipídico, incluindo ao menos 120 reações enzimáticas (Li-Beisson et al., 2013). A partir de comparações com *A. thaliana*, estimou-se que a soja tenha de duas a três vezes mais genes relacionados à sinalização lipídica e à síntese de lipídeos de membrana (LM), enquanto que 63% mais genes em soja, acredita-se, estão envolvidos na síntese *de novo* de ácidos graxos (AGs) nos plastídeos (S. Yang et al., 2019).

2.2 Lipídeos em soja

Os lipídios são compostos hidrofóbicos presentes em todas as células e que desempenham diversas funções biológicas essenciais. Por exemplo, são a alternativa biológica mais eficiente no armazenamento de energia na forma de TAG - molécula altamente reduzida (Bates et al., 2013; Cagliari et al., 2011), compõem membranas celulares (Reszczyńska & Hanaka, 2020), além de participarem em processos de sinalização, reconhecimento e comunicação celular (Michaelson et al., 2016). Adicionalmente, os lipídeos atuam como pigmentos fotossensíveis e âncoras hidrofóbicas para proteínas (Hemsley, 2015; James & Nichols, 1966).

Entre os lipídeos celulares, os AGs são os mais abundantes. São ácidos carboxílicos com cadeias hidrocarbonadas, classificados de acordo com o número de átomos de carbono e o grau de insaturação. Podem ser de cadeia curta (2 a 6 carbonos), média (8 a 12 carbonos), longa (14 a 18 carbonos) ou muito longa (maior que 18 carbonos) e podem ser saturados - i.e., sem nenhuma ligação dupla (AGsat, e.g. ácido palmítico [16:0] e esteárico [18:0]), monoinsaturados - i.e., com uma ligação dupla (AGmi, e.g. ácido oleico [18:1]) ou poli-insaturados - i.e., duas ou mais ligações duplas (AGpi, e.g. ácido linoleico [18:2] e linolênico [18:3]) (Figura 5.A) (Voelker & Kinney, 2001).

O óleo de soja é gerado e armazenado nas sementes principalmente como AGs, TAGs e tocoferóis (vitamina E) (Chu et al., 2023; A. Liu et al., 2022). Em soja há maior concentração de AGs de cadeia longa e insaturados, a saber, 5 a 11% [18:3], 43 a 56% [18:2], 15 a 30% [18:1] e 11 a 26% de AGsat (Clemente & Cahoon, 2009; Martin et al., 2008; Nguyen et al., 2016). É a composição desses AGs que determinará a qualidade do óleo e, portanto, esses cinco AGs mais comuns foram alvos de alterações por meio de engenharia genética, no óleo de soja, visando atender às demandas dos consumidores finais (Fehr, 2007). Por exemplo, um óleo de soja rico em AGmi [18:1] pode ser mais vantajoso do que o óleo convencional devido à sua maior estabilidade oxidativa. Do mesmo modo, aumentar o

teor de ômega-3 [18:3] e reduzir o conteúdo de ômega-6 [18:2] contribui para a prevenção da obesidade e de doenças cardiovasculares (Yeom et al., 2020).

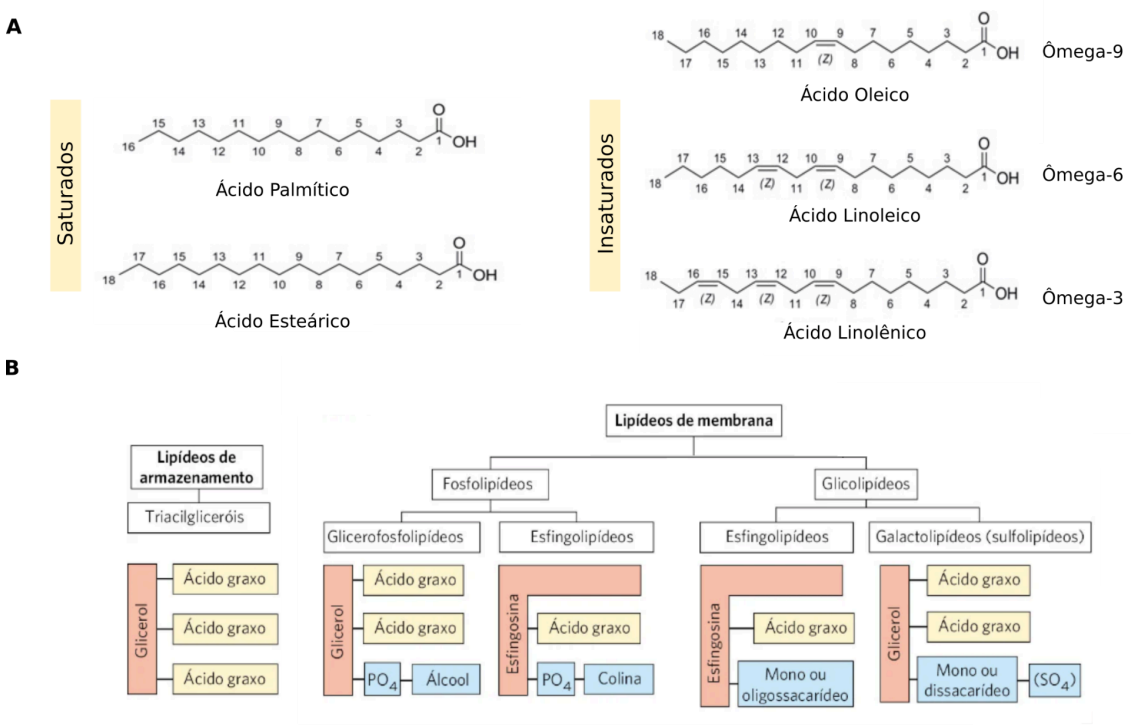


Figura 5. Principais tipos de ácidos graxos e demais lipídeos. **A**. Estrutura e composição dos ácidos graxos saturados (palmítico e esteárico) e insaturados (oleico,linoleico e linolênico). **B**. Representação da constituição de lipídeos de armazenamento e lipídeos de membrana. Todos os lipídeos representados têm glicerol ou esfingosina como esqueleto (em vermelho), que estão ligados a um ou mais grupos alquila de cadeia longa (amarelo) e uma cabeça polar (azul). Em TAGs, glicerofosfolipídeos, galactolipídeos e sulfolipídeos, os grupos alquilas são ácidos graxos em ligação éster. Os esfingolipídeos têm um único ácido graxo em ligação amida com o esqueleto de esfingosina. Nos fosfolipídios, a cabeça polar está unida por ligação fosfodiéster, enquanto os glicolipídeos têm uma ligação glicosídica direta entre o açúcar da cabeça polar e o esqueleto de glicerol. Fonte: (Nelson & Cox, 2010), p.366. Adaptado.

Os LM têm AGs ligados às posições sn-1 e sn-2 do esqueleto de glicerol, com uma cabeça polar na posição sn-3 (Figura 5.B). Por outro lado, quando as três posições do glicerol estão esterificadas com AGs (Figura 5.B), o resultado é o TAG, a principal forma de armazenamento de energia em sementes de soja (Ohlrogge & Browse, 1995). Dependendo das necessidades do organismo, grande parte dos AGs sintetizados será incorporada nos componentes fosfolipídicos das membranas ou armazenada na forma de TAGs. Geralmente, o acúmulo de TAGs e outros nutrientes ocorre na fase R5 de maturação da semente (Figura 3), e a energia armazenada em TAGs será importante para a síntese de carboidratos durante a germinação das sementes e para o início do desenvolvimento das plântulas (Fehr & Caviness, 1977; Graham, 2008; Ohlrogge & Browse, 1995; Theodoulou & Eastmond, 2012). Além disso, os TAGs atuam como uma proteção contra AGs citotóxicos e

outros intermediários lipídicos, desempenhando um importante papel na manutenção da homeostase lipídica intracelular e na sobrevivência das células (Fan et al., 2013, 2014). As vias de síntese de LM e de TAGs começam com a formação de ésteres acil graxo de glicerol. Assim, os TAGs de muitas sementes compartilham os mesmos grupos acil encontrados em LM (Figura 5.A).

2.3 Biossíntese de óleo em soja

A sacarose é a principal fonte de carbono para a síntese de TAG em plantas superiores. Em sementes e outros tecidos não fotossinteticamente ativos, a sacarose é quebrada por invertases e convertida a piruvato por meio da glicólise (Bourgis et al., 2011). O piruvato é o precursor para a produção de moléculas de acetil-Coenzima A (acetil-CoA) destinadas à síntese de AGs (Baud & Lepiniec, 2010). Nesse contexto, a piruvato quinase (PK), a piruvato desidrogenase (PDH) e a acetil-CoA carboxilase (ACCase) são as enzimas chave para a síntese de AGs (Figura 6.A). Ademais, a via das pentoses fosfato mantém a homeostase de carbono pelo suprimento de moléculas redutoras para os processos anabólicos como a biossíntese de AGs (Stincone et al., 2015).

Em sementes oleaginosas, um intenso tráfego de lipídeos passa por diferentes vias bioquímicas para a formação de AGs e armazenamento em TAGs na semente (Figura 6.A). A fosfatidilcolina (PC) desempenha papel central nessas vias como um substrato para as modificações de acil e como carreadora para o tráfego de grupos acil entre as organelas e os subdomínios de membrana (Bates et al., 2013). Em 2009, Bates et al. investigaram a origem dos AGs constituintes dos TAGs formados em soja (Bates et al., 2009) e concluíram que cerca de 90% dos TAGs são formados a partir de diacilglicerois (DAGs) derivados de PC, ou seja, por meio de vias alternativas às de *Kennedy* - esta última também conhecida como via clássica de síntese de TAG (Bates & Browse, 2012; Bates et al., 2009).

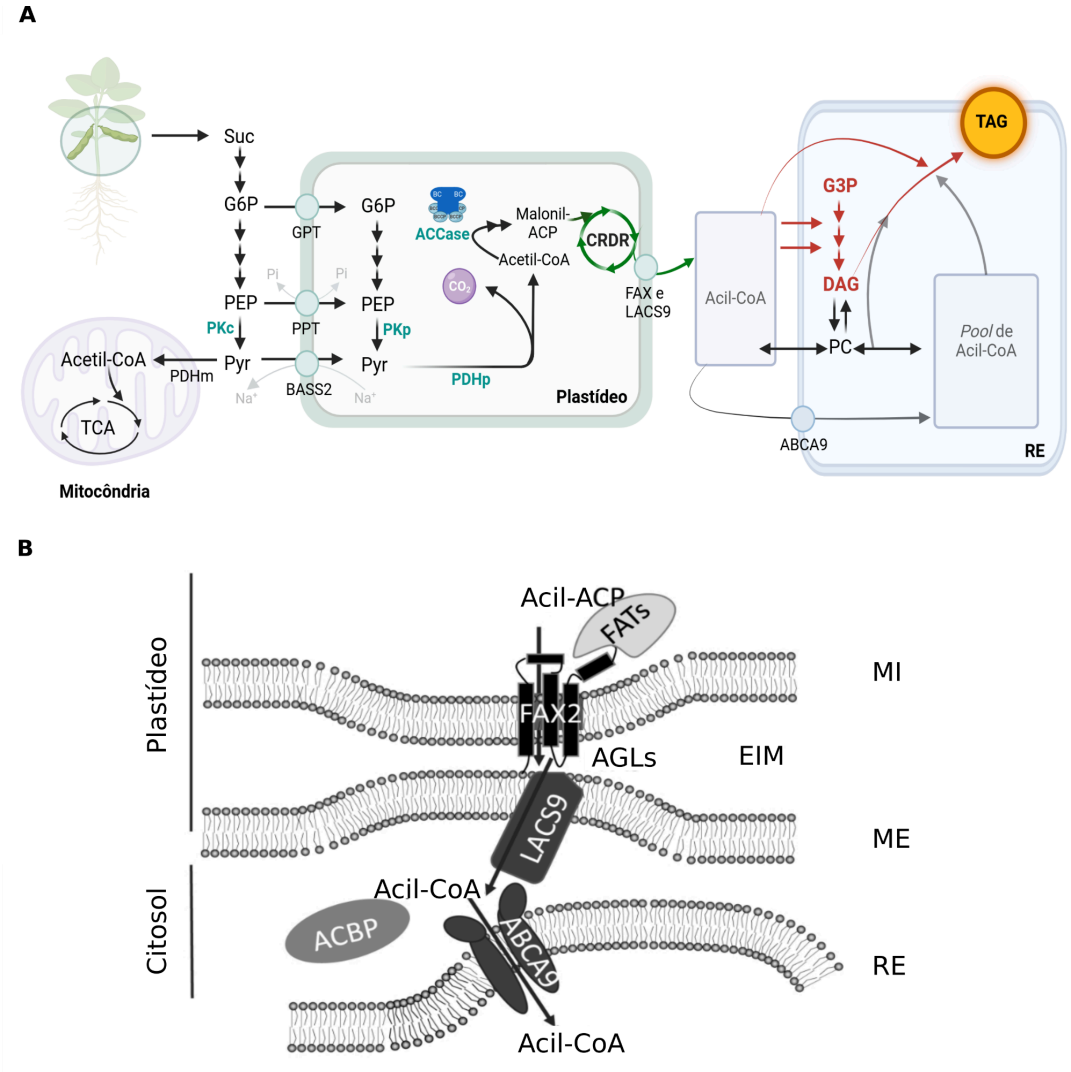


Figura 6. Diagrama simplificado das vias do metabolismo de óleo e exportação de ácidos graxos durante o desenvolvimento das sementes. **A**. A sacarose importada dos tecidos maternos é convertida em G6P. Em seguida, a G6P é metabolizada para PEP. No citosol, o PEP é convertido em Pyr ou entra no plastídio via PPT. Nos plastídios, o Pyr é usado como substrato para a biossíntese de AGs que podem ser direcionados para a via de biossíntese de TAG no RE. O Pyr citosólico é utilizado em muitas vias metabólicas, como a biossíntese de proteínas e o ciclo do TCA. Em vermelho está representada a via de Kennedy, em verde as enzimas chave para a síntese de AGs. Os círculos na membrana plastidial indicam os transportadores específicos para cada metabólito. Modificado de Lee et al. (E.-J. Lee et al., 2017) plastídio, acil-ACPs são hidrolisados por FATs, formando AGLs que são exportados pela FAX2 através da membrana interna do plastídeo para o espaço intermembrana. A acilação dos AGLs é catalisada por LACS9 antes de serem transportados através da ME plastidial. Os acil-CoAs exportados do plastídio são capturados por ACBPs no citosol ou direcionados ao RE para a síntese de TAG. Acredita-se que FATs, FAX2 e LACS9 formam um complexo por meio de interações físicas entre as proteínas para a exportação mais eficiente de AGs. Fonte: Tian et al. (Tian et al., 2019). Abreviações: **Suc**, sacarose; **G6P**, Glicose-6-fosfato; **PEP**, fosfoenolpiruvato; **Pyr**, piruvato; **PKc**, Pyr quinase citosólica; **PKp**, Pyr quinase plastidial; **PDHm**, pyr desidrogenase mitocondrial; **PDHp**, pyr desidrogenase plastidial; **TCA**, ciclo do ácido tricarboxílico; **GPT**, translocador de G6P; **PPT**, translocador de fosfoenolpiruvato/fosfato; **BASS2**, transportador de pyr; **ACCase**, acetil-coenzima A carboxilase; **ACP**, proteína carreadora de acil; **CRDR**, ciclo de condensação, redução, desidratação e redução catalisada pelo complexo enzimático da ácido graxo sintase; **FAX**, exportador de ácido graxo; **LACS9**, acil-CoA sintetase de cadeia longa 9; **G3P**, Glicerol-3-fosfato; **DAG**, diacilglicerol; **TAG**, triacilglicerol; **PC**, fosfatidilcolina; **ABCA9**, transportador dependente de ATP; **FATs**, tioesterase de ácido graxo; **AGLs**, ácidos graxos livres; **ACBP**, proteína ligante de acil-CoA; **MI**, membrana interna; **EIM**, espaço intermembrana; **ME**, membrana externa; **RE**, retículo endoplasmático.

O processo de formação de óleo na semente pode ser dividido em quatro etapas, a saber: i. síntese *de novo* de AGs nos plastídeos; ii. alongamento e edição de acil no plastídeo e no RE; iii. montagem de TAG no RE e iv. formação de corpos oleaginosos (COs). Cada uma dessas etapas (Figura 6) envolve um grande número de genes constituindo um complexo processo metabólico sensível às variantes ambientais (Yang et al., 2019). Resumidamente, o início da biossíntese de AGs envolve a acetil-CoA e reações bioquímicas nos plastídeos que produzem 16:0-proteína carreadora de acil (ACP, do inglês *Acyl Carrier Protein*). Em seguida, 16:0-ACP é modificado por reações enzimáticas, resultando em longas cadeias de acil-CoA que são armazenadas em um *pool* de acil-CoA no retículo endoplasmático (RE). Simultaneamente, C18:0 e C18:1 são ligados a malonil-CoA por meio de reações sequenciais, similar à síntese de AGs nos plastídeos, produzindo AGs insaturados de cadeia longa e/ou muito longa (a depender da espécie), que são armazenados no *pool* de fosfatidilcolina (PC, do inglês *Phospohatidylcholine*) para processos subsequentes. Os produtos armazenados no *pool* de acil-CoA e PC são conduzidos à via de *Kennedy* ou à via complexa, levando à formação de TAG (Figura 6) (Bates et al., 2013).

2.3.1 Síntese de novo e alongamento de cadeias acil

A síntese *de novo* de AGs é realizada no estroma dos plastídeos. Nessa organela será determinado o comprimento das cadeias acil de até 18 carbonos e o nível de AGSat que irão compor o óleo depositado nas sementes. A síntese e o alongamento das cadeias acil são realizadas pela adição sucessiva de duas unidades de carbono: os dois primeiros são provenientes do acetil-CoA e os seguintes do malonil-CoA, os blocos de construção dos AGs (Xu et al., 2018). O malonil-CoA é uma molécula de três carbonos que é produzida pela carboxilação de acetil-CoA pela forma heteromérica da ACCase, um complexo enzimático com quatro subunidades - biotina carboxilase (BC), proteína carreadora de biotina (BCCP), α- carboxil transferase (α-CT) e β-carboxil transferase (β-CT) (Apêndice 2, Figura 7), essencial para a síntese *de novo* de AGs (Salie & Thelen, 2016). Essa enzima utiliza bicarbonato (HCO_3^-) como fonte de CO_2, adenosina trifosfato (ATP), acetil-CoA e biotina (como um cofator) para produzir o malonil-CoA (Figura 7).

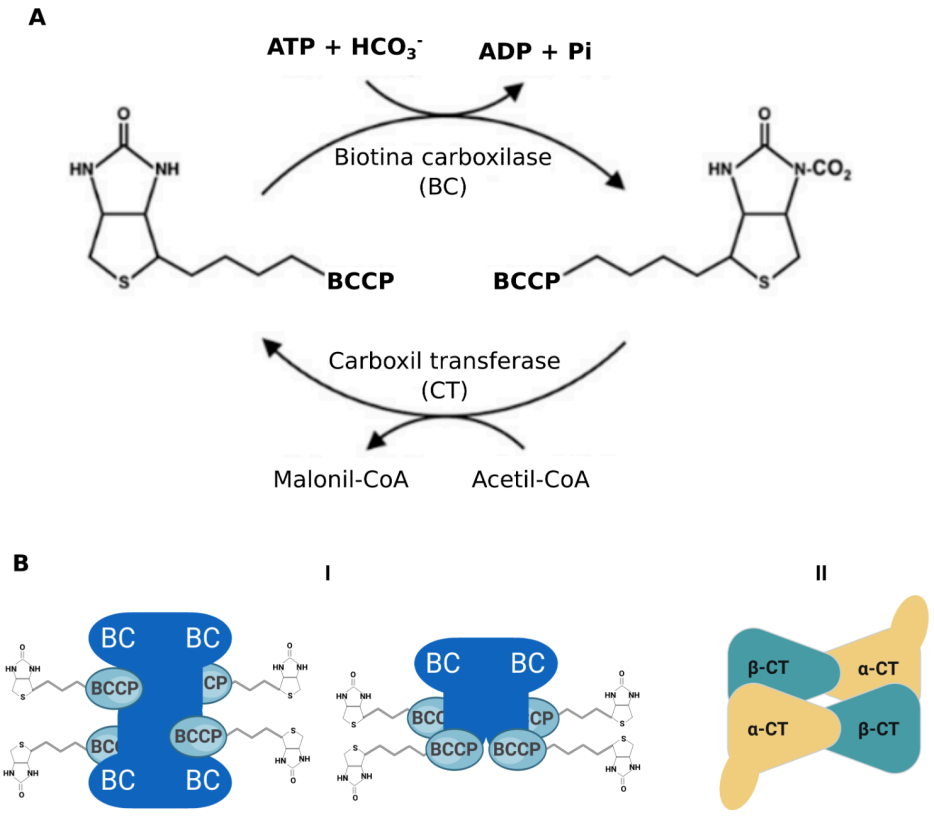


Figura 7. Mecanismo de reação da Acetil-CoA Carboxilase (ACCase) e esquema dos sub complexos da ACCase heteromérica de plantas. **A**. A carboxilação da porção de biotina dependente de ATP com a proteína carreadora de biotina (BCCP) é catalisada pela biotina carboxilase (BC). O grupo carboxila ativado é então transferido para a acetil-coenzima A pela carboxil transferase (CT) para produzir malonil-coenzima A. O acoplamento do sítio ativo entre as meias reações de BC e CT é facilitado pelo braço flexível de biotina-lisina dentro do holo-BCCP. **B**. A estrutura total da ACCase heteromérica em plantas ainda é incerta, mas sabe-se que consiste em dois sub complexos separados. **I**. Há evidências de um subcomplexo BC–BCCP consistindo de uma proporção de 1:1 (esquerda) ou uma proporção de 1:2 (direita). Em ambos os casos, o subcomplexo contém quatro cofatores de biotina disponíveis para a carboxilação. **II**. a meia reação de CT é um heterotetrâmero contendo duas subunidades α-CT e duas β-CT. Adaptado de Salie e Thelen (Salie & Thelen, 2016).

O processo de síntese de AGs é catalisado pelo complexo enzimático conhecido por ácido graxo sintase (FAS, do inglês *Fatty Acid Synthase*) que em plantas é composto por sete cadeias polipeptídicas (Figura 8). Um dos componentes deste complexo é a proteína carreadora de acila (ACP, do inglês *Acil Carrier Protein*) que tem como grupo prostético a fosfopanteteína (vitamina B5), também encontrada na CoA. Esse processo envolve quatro reações, a saber: condensação de *Claisen*, redução, desidratação e redução novamente (CRDR), que alongam a cadeia acila por meio da condensação de duas unidades de carbono por vez. Cada uma das quatro reações, apresentadas na Figura 8, é catalisada pela atividade de uma enzima específica da FAS (Apêndice 2). As reações de condensação são catalisadas por enzimas conhecidas como 3-cetoacil-ACP sintases (KASs, do inglês *3-ketoacyl-ACP synthases*). A primeira reação de condensação (acil + malonil) é catalisada pela KAS III, as subsequentes (C4:C16) pela KAS I e de 16:0 para 18:0 é catalisada pela

KAS II (Ohlrogge & Browse, 1995). Os AGs de cadeia muito longa são gerados por um sistema enzimático do RE, chamado elongase de AGs (FAE, do inglês *Fatty Acid Elongase*) (Harwood, 1988, 1996), que usa substratos de AGs preexistentes, gerados pela FAS (J. B. Ohlrogge & Jaworski, 1997). Tanto os sistemas FAS *de novo* quanto os FAE utilizam iterações cíclicas CRDR para alongar uma cadeia acila em dois átomos de carbono por ciclo. Em contraste com o sistema FAS, que utiliza intermediários acila ligados à ACP, o sistema FAE utiliza intermediários acila ligados à CoA nessas reações (Harwood, 1988; Li-Beisson et al., 2013).

A síntese se inicia pela transferência de um grupo acetil para um resíduo –SH da ß-acetil-ACP sintase (KS) da FAS e de um grupo malonil para a ACP. O grupo malonil reage com o grupo acetil liberando uma molécula de CO_2. Esse CO_2 liberado é o mesmo que foi adicionado ao acetil-CoA, formando o malonil-CoA (Figura 7). Portanto, apesar de essencial, o CO_2 não faz parte do AG formado. Essa descarboxilação é importante para tornar a reação termodinamicamente favorável e irreversível. A reação de condensação transfere o grupo acil para a ACP (Figura 8, etapa 1). O próximo passo envolve uma redução dependente de NADPH+H^+ para formar 3-hidroxibutiril-ACP a partir de 3-cetobutiril-ACP (Figura 8, etapa 2). Subsequentemente, uma desidratase (DH) remove uma molécula de água (H_2O) da molécula recém-formada, resultando em um trans-Δ2-butenoil-ACP (Figura 8, etapa 3). O último passo do ciclo é uma segunda redução dependente de NADPH + H^+ para desfazer a dupla ligação proveniente da saída de H_2O na desidratação, consequentemente produzindo o tioéster butiril-ACP saturado de quatro carbonos (Figura 8, etapa 4). Para que o alongamento da cadeia prossiga, o grupo butila formado é transferido para o resíduo -SH da KS e outro grupo malonil é transferido para a ACP. O ciclo se repete por mais sete ou oito vezes, usando o malonil como doador de 2 carbonos para o alongamento da cadeia, e finda por produzir um AG 16:0 ou 18:0, respectivamente (Ohlrogge & Browse, 1995). Além disso, ainda no plastídeo, é produzido o AGmi 18:1.

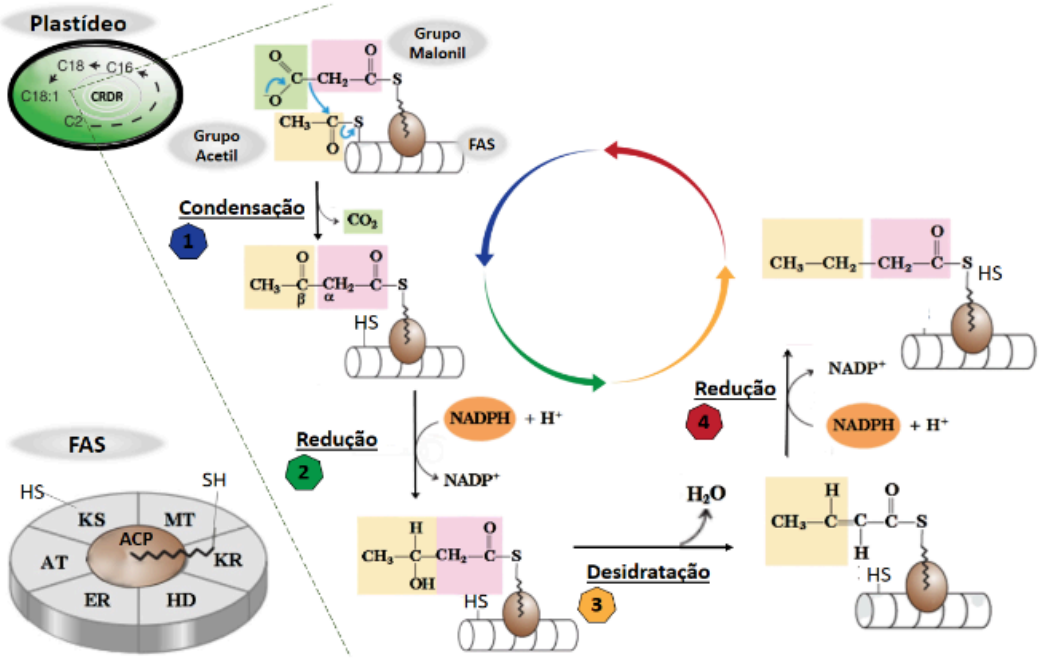


Figura 8. Ciclo de síntese e alongamento dos ácidos graxos nos plastídeos. Em **1**. Ocorre a condensação do grupo acetil com o malonil, liberando CO2, catalisada pela ß-cetoacil sintase III (**KAS III**). **2**. Redução do 3-hidroxiacil-ACP. **3**. Desidratação de um enoil-ACP **4**. Segunda redução para formar uma molécula saturada de quatro carbonos 4:0-ACP. Subsequentes ciclos de reações de condensação de 4:0-ACP com o malonil-ACP até 16:0-ACP tipicamente são catalisados por uma enzima conhecida como **KAS I**. A condensação do 16:0-ACP com o malonil-ACP é catalisada pela **KAS II**. Outras abreviações: **FAS**, ácido graxo sintase; **KS**, beta-cetoacil-ACP sintase; **MT**, malonil-ACP transferase; **KR**, beta-cetoacil-ACP redutase; **HD**, beta-hidroxiacil-ACP desidratase; **ER**, enoil-ACP redutase; **AT**, acetil-CoA–ACP transferase; **ACP**, proteína carreadora de acil. Fonte: Nelson e Cox, 2010. Princípios de Bioquímica de Lehninger, 5ª edição, capítulo 21. Modificado.

2.3.2 Edição de acil: dessaturação, terminação e liberação

Ainda no plastídeo, muitos dos grupos acil 18:0-ACP sofrem dessaturação no carbono 9, catalisada pela ∆9-desaturase também conhecida como desaturase de estearoil-ACP (SAD, do inglês *stearoyl-ACP desaturase*), para formar 18:1-ACP. Normalmente, duas acil-ACP tioesterases diferentes, codificadas pelos genes parálogos *FATA* e *FATB*, hidrolisam a acil-ACP para produzir AG livres (AGLs). *FATA* hidrolisa predominantemente o 18:1-ACP, com menor atividade sobre 18:0-ACP e 16:0-ACP. *FATB* geralmente hidrolisa AGSat com cadeias carbônicas variando de C14-C18-ACPs, preferencialmente 16:0-ACP, embora também hidrolise 18:1-ACP. Ao que tudo indica, *FATA* e *FATB* interrompem a reação de alongamento da cadeia quando esta chega a um determinado número de carbonos (Bates et al., 2013). Portanto, os AGLs 16:0 e 18:1 são os principais produtos da síntese de AGs plastidial e suas proporções relativas são determinadas pelas atividades de *FATA*, *FATB*, *SAD* e *KASII* (Apêndice 2, Figura 8) (Bates et al., 2013; Li-Beisson et al., 2013; Zhou et al., 2021).

Os AGLs devem seguir para o RE onde sofrerão mais insaturações e/ou serão incorporados ao TAG. No entanto, devido a sua hidrofobicidade, os AGs não podem se mover livremente no ambiente celular aquoso (Li et al., 2016) e necessitam de transportadores que auxiliem no tráfego entre plastídeo e RE (Manan, Chen, et al., 2017). Assim, o transporte de AGs do plastídeo ao RE é um pré-requisito para o acúmulo de óleo na semente (Tian et al., 2019). Desse modo, os AGLs devem ser convertidos em acil-CoA antes de serem liberados no citosol (Figura 6.B). Proteínas da família acil-CoA sintetase de cadeia longa (LACs9, do inglês *long chain acyl-CoA synthetase*), localizadas na membrana do envelope externo catalisam boa parte da atividade plastidial na formação de acil-CoA, regulando sua concentração (Jessen et al., 2015). Acredita-se que a exportação de acil-CoAs esterificados do plastídeo/citosol ocorra por meio de proteínas da família FAX (do inglês, *fatty acid export*) com o auxílio de outras proteínas (R. Chen et al., 2023; N. Li et al., 2020; Tian et al., 2019) e os acil-CoAs são importados para o RE através das proteínas de ligação ao acil-CoA (ACBPs, do inglês *acyl-CoA binding proteins*) localizadas no citosol ou na membrana do RE das células de semente (Benning, 2009; S. Kim et al., 2013). Apesar da família FAX ser bem conhecida em Arabidopsis, não há muitas informações sobre ela na soja. Além disso, moléculas de PC podem participar do transporte de AGs do cloroplasto para o RE através do ciclo de edição de acil durante o acúmulo de óleo na semente (Tjellström et al., 2012).

Antes da acilação do glicerol-3-fosfato (G3P) para formar ácido fosfatídico (PA, do inglês *phosphatidic acid*), DAG ou fosfolipídios, os AGs nascentes são primeiro incorporados à PC, reação catalisada pela LPCAT (lisofosfatidilcolina aciltransferase, do inglês *lysophosphatidylcholine acyltransferase*), na membrana do envelope externo do plastídeo (Tjellström et al., 2012). A síntese de AGpis de cadeia longa ou de AGs raros geralmente ocorre no PA ou no *pool* de acyl-CoA, onde os AGs plastidiais nascentes, principalmente: 16:0, 18:0 e 18:1, sofrerão outras modificações como mais alongamento e dessaturação. O alongamento de cadeias acil é catalisado por uma elongase ligada ao RE com acil-CoA como substrato. Enquanto esterificado à PC, 18:1 pode ser desaturado para 18:2 e 18:3 por duas desaturases de AGs especializadas associadas à membrana microssomal: FAD2 e FAD3, que introduzem uma ligação dupla na configuração cis nas posições Δ-12 e Δ-15 dos AGs na posição sn-2 da PC, respecitivamente. Após a modificação, os AGs na posição sn-2 da PC podem ser liberados pela atividade reversa da LPCAT (Bates et al., 2013). A desacilação da PC e a reacilação da lisofosfatidilcolina (LPC, do inglês *lysophosphatidylcholine*) marcam a edição de acil que troca AG da PC com o *pool* de acil-CoA sem a síntese ou degradação líquida de PC. Esse ciclo permite a formação de DAG com AGpis. O DAG representa um importante ponto de ramificação entre a síntese de

TAG e LM. A reação que converte DAG para TAG, catalisada pela diacilglicerol acil transferase (DGAT), é única na via de biossíntese de TAG (Bates & Browse, 2012; Bates et al., 2013; Li-Beisson et al., 2013).

2.3.3 Formação de TAG

Ao chegarem no citoplasma os AGLs são esterificados à CoA e agora servem como substratos para a formação de lipídeos. Além disso, os AGs modificados pela edição de acil são movidos para fora da PC e, eventualmente, também podem ser incorporados em TAGs por meio de diferentes rotas (Figura 9). A síntese de TAG ocorre no RE e provavelmente também envolve reações nos COs.

Em sua forma simples, a síntese *de novo* de TAG dependente de acil-CoA, também conhecida como *via de Kennedy* ou via glicerol fosfato (Figura 9, setas verdes), ocorre a partir da incorporação sequencial de AGs no esqueleto de G3P, envolvendo quatro passos enzimáticos. Por outro lado, há a via complexa, onde o TAG pode ser sintetizado por diferentes vias (Figura 9, setas azuis e roxas), como a independente de acil-CoA que transfere grupos acil da PC para DAG pela ação da fosfolipídeo diacilglicerol aciltransferase (PDAT, do inglês *phospholipid: diacylglycerol acyltransferase*) e por outras vias onde a PC desempenha um papel fundamental (Bates et al., 2013; Fan et al., 2014; R. Li et al., 2010; Ohlrogge & Browse, 1995; Pan et al., 2015).

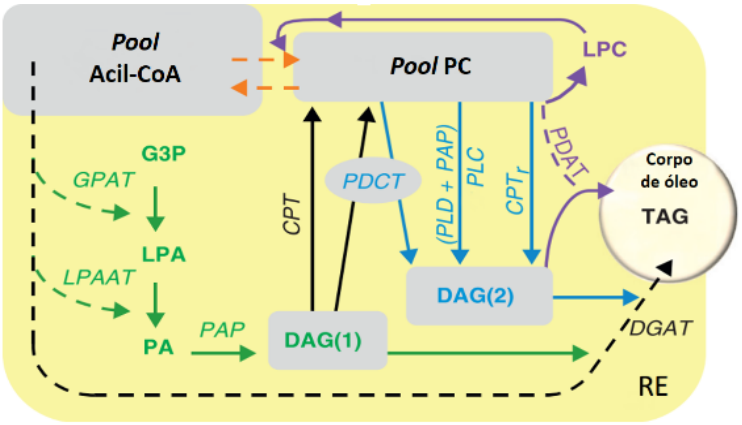


Figura 9. Diferentes vias da síntese de triacilglicerol. Linhas verdes: síntese de novo de TAG; Linhas azuis: síntese de diacilglicerol derivado de fosfatidilcolina (DAG(2)); Linhas laranjas: edição de acil; Linhas roxas: ação da fosfolipídeo: diacilglicerol aciltransferase (PDAT). DAG(1): proveniente da síntese de novo. Abreviações: **G3P**: Glicerol-3-fosfato; **GPAT**, glicerol-3-fosfato aciltransferase; **LPA**, ácido lisofosfatídico; **LPAAT**, ácido lisofosfatídico acil transferase; **PA**, ácido fosfatídico; **PAP**, fosfatase de ácido fosfatídico; **PC**, fosfatidilcolina; **CPT** e **PDCT**, colina transferases; **PLC** e **PLD**, fosfolipase C e D, respectivamente; **DGAT**, diacilglicerol aciltransferase; **TAG**, triacilglicerol; **RE**, retículo endoplasmático. Fonte: Bates et al. (Bates et al., 2013).

2.3.3.1 Via simples

A G3P aciltransferase (GPAT, do inglês *glycerol-3-phosphate acyltransferase*) ligada à membrana do RE inicia o processo transferindo a cadeia acil da CoA para a posição sn-1 do G3P, formando o ácido lisofosfatídico (LPA, do inglês *Lysophosphatidic acid*). Essa enzima parece ter uma baixa seletividade para as cadeias acil. A aciltransferase do LPA (LPAAT, do inglês *Lysophosphatidic acid acyltransferase*) catalisa a transferência da cadeia acila do éster-CoA para o sn-2, criando PA (Figura 9, setas verdes). Nas plantas, essa enzima tem maior afinidade por cadeias acila insaturadas. Depois da segunda acilação, a desfosforilação do PA resultante é catalisada pela fosfatase do ácido fosfatídico (PAP, do inglês *phosphatidic acid phosphatase*), formando DAG. A enzima final, DGAT, transfere grupos acil a partir da acil-CoA para o sn-3 do DAG e forma TAG (Bates et al., 2013).

2.3.3.2 Via complexa

Estudos com mutantes da DGAT demonstraram evidências da existência de vias alternativas à de *Kennedy* para a formação de TAG em plantas (Bates & Browse, 2012; Bates et al., 2009). A PC claramente funciona como um intermediário chave como substrato para as modificações de acil e como carreadora para o tráfego de grupos acil entre as organelas. Assim, há três mecanismos que possibilitam o fluxo de AG para a PC direcionada à síntese de TAG, a saber:

I. Edição de acil (Figura 9, setas laranjas);

II. transferência direta do AG a partir da PC para o DAG, pela ação da PDAT (Figura 9, setas roxas). Essa enzima pode transferir a cadeia acil do sn-2 a partir da PC para o DAG, formando liso-PC e TAG. A síntese de TAG pela PDAT1 é dependente da atividade de LPCAT para a regeneração da PC a partir de liso-PC;

III. utilização de DAG derivada da PC como substrato para a biossíntese de TAG (Figura 9, DAG(2), setas azuis).

2.3.4 Corpos oleaginosos

Uma vez sintetizadas, as moléculas de TAG se juntam para formar estruturas denominadas COs ou oleossomos. Esses COs são estruturas subcelulares circulares ou ovóides armazenados no citoplasma. São envoltos por uma monocamada de fosfolipídios com proteínas integrais de membrana, por exemplo, a oleosina (W. X. Liu et al., 2013; Tzen et

al., 1993). Há diferentes isoformas de oleosina e duas proteínas integrais menores, caleosina e esterosina que se acumulam sequencialmente ao longo do desenvolvimento da semente e têm um papel fundamental na estabilização dos COs durante a dessecação da semente (Y. Song et al., 2017). As oleosinas contêm um domínio hidrofóbico ligado ao corpo oleaginoso e dois domínios anfipáticos. Essas proteínas determinam o tamanho dos corpos oleaginosos e, assim, facilitam a mobilização de TAGs durante a germinação (Siloto et al., 2006). As caleosinas também parecem desempenhar um papel na mobilização de TAGs durante a germinação, possivelmente facilitando as interações com vacúolos. Recentemente, descobriu-se que essas proteínas apresentam diferentes estruturas moleculares que lhes conferem funções celulares além da estruturação de COs (Saadat, 2023). As esteroleosinas, além de um domínio de ancoragem nos COs, possuem uma desidrogenase de ligação ao esterol que pode desempenhar um papel na transdução de sinal (Lin et al., 2002). A soja possui pequenos e estáveis COs que compõem de 18% a 22% de sua massa total. Esses COs são gotículas organizadas de TAGs ricos em constituintes bioativos menores. Eles são envolvidos por uma monocamada de fosfolipídios nos quais estão embutidos sete oleosinas, duas caleosinas e uma esteroleosina (zaaboul et al., 2022).

Em eucariotos, os COs são derivados do RE. O modelo de biogênese de COs no RE e brotamento no citosol é o mais amplamente aceito (Y. Zhao et al., 2023). De acordo com este modelo, a biogênese dos COs é dividida em três etapas principais: nucleação, crescimento e brotamento. A nucleação refere-se ao acúmulo inicial de lipídios neutros (principalmente TAGs) que formam uma "lente" lipídica entre as duas monocamadas do RE. O crescimento refere-se à expansão de tamanho devido à absorção de TAGs sintetizados ou fusão de pequenos COs nascentes. O brotamento é o processo pelo qual a membrana do RE se projeta para o citoplasma para formar os COs nascentes (A. H. C. Huang, 2018; Thiam & Ikonen, 2021). Apesar destes estudos, a formação de COs em plantas permanece pouco compreendida.

3 - CAPÍTULO 2. ARTIGO[2]

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

Authors: Dayana K. Turquetti-Moraes, Cláudio Benício Cardoso-Silva, Fabricio Almeida-Silva, Thiago M. Venancio.

[2] **Supplementary Tables and Figures Statement:** The supplementary tables related to this chapter are available online at the following link: https://doi.org/10.1101/2024.05.02.592228. These tables provide additional and detailed information that complements the data presented in this article. Regarding the supplementary figures, they are available in the appendix of this document.

3.1 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.

3.2 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 (Aziziaram et al., 2021; Pattee et al., 1983; Shen et al., 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 et al., 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 et al., 2006).

Gene and genome duplications have been extensively associated with plant adaptation and diversification (Van de Peer et al., 2017; Zhuang et al., 2022). Gene duplication may occur through mechanisms such as WGDs or small-scale duplications (SSD) (Flagel & Wendel, 2009; Ren et al., 2018). WGDs also played a role in the domestication of plants that eventually became modern crops, such as wheat and maize (Carretero-Paulet & Van de Peer, 2020; Hake & 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 et al., 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 et al., 2015). Over time, gene families can gain or lose genes, generating a wealth of genetic material for adaptation (Cheng et al., 2018; Moharana & Venancio, 2020; Renny-Byfield & 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 (H. Song et al., 2016; J. Wang et al., 2020).

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, Chen, et al., 2017; Marchive et al., 2014; Meinke et al., 1981; Nguyen et al., 2016; Pham et al., 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.

3.3 Results and discussion

3.3.1 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.

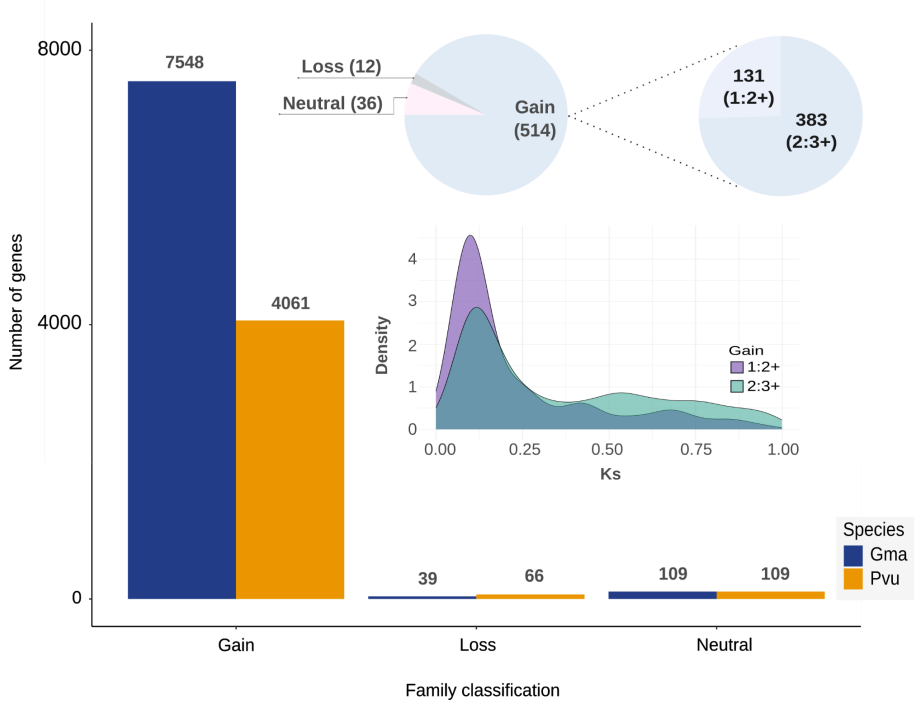


Figure 1. Number of soybean and common bean oil-related genes in different relative family size categories. Inset: number of families in each category. The two categories of gain families (1:2+ and 2:3+) are highlighted, alongside their respective K_s density plots. **Gain**: families with more genes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

3.3.2 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, 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, 2013; Bates & 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.

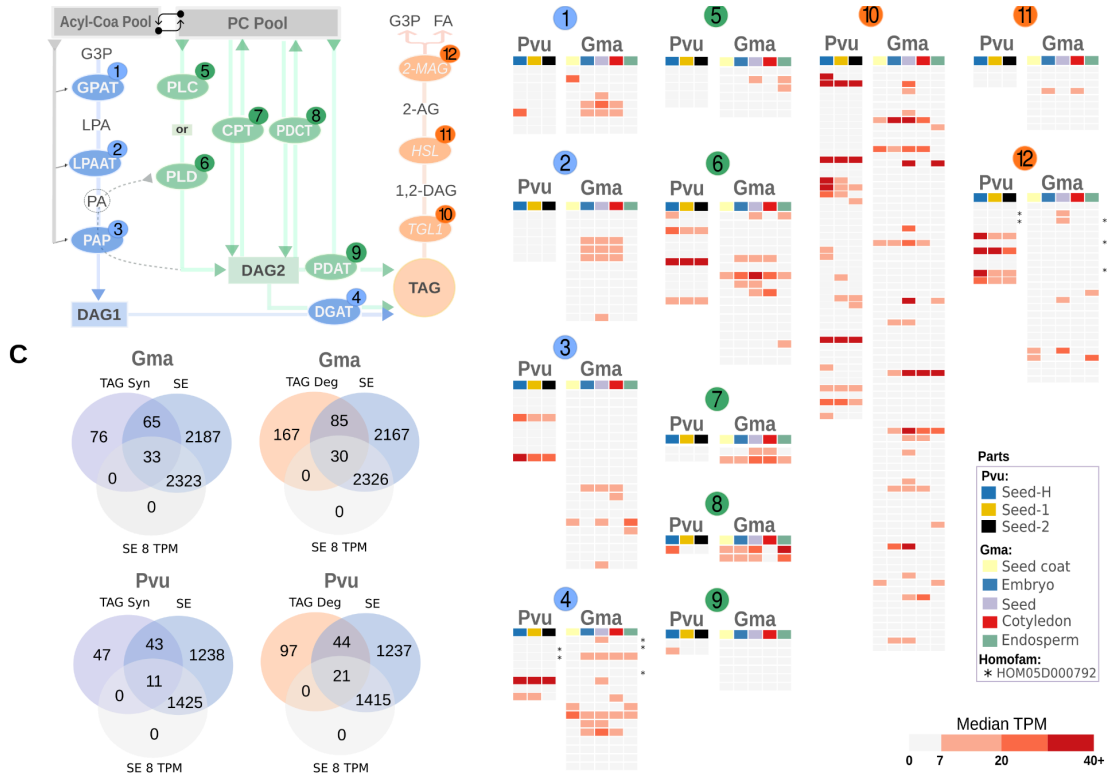


Figure 2. Major enzymes involved in DAG/TAG synthesis and degradation. **A**. Reactions involved in TAG degradation and formation from FAs. Blue: Kennedy pathway (gray arrows indicate the source of FA to feed the pathway). Green: complex pathway; the dashed arrow indicates a reaction PLD-PAP to form 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 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 (S.-C. Kim & 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 et al., 2013; J. Lee et al., 2011; Wakelam, 1998; W. Yang et al., 2017). 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), PLDα1 (Glyma.13G364900) is a promising target considering its high expression in cotyledons (~20 TPM). Further, suppression of PLDα in soybean results in decreased levels of polyunsaturated FAs in TAG (J. Lee et al., 2011), suggesting a flow from PC to TAG via PLD-PAP without PDCT. In addition, GmPLDα3 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 et al., 2015). PDATs often have

contrasting expression profiles in different plant species (Pan et al., 2015; Torabi et al., 2021). 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 et al., 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.

3.3.3 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 (Fang et al., 2017; Niu et al., 2020) (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 et al., 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. Lu et al., 2022; S. Wang et al., 2020). 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 (Boyang et al., 2023; Hooker et al., 2022; C. Liu et al., 2023; H. Yang et al., 2019). 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).

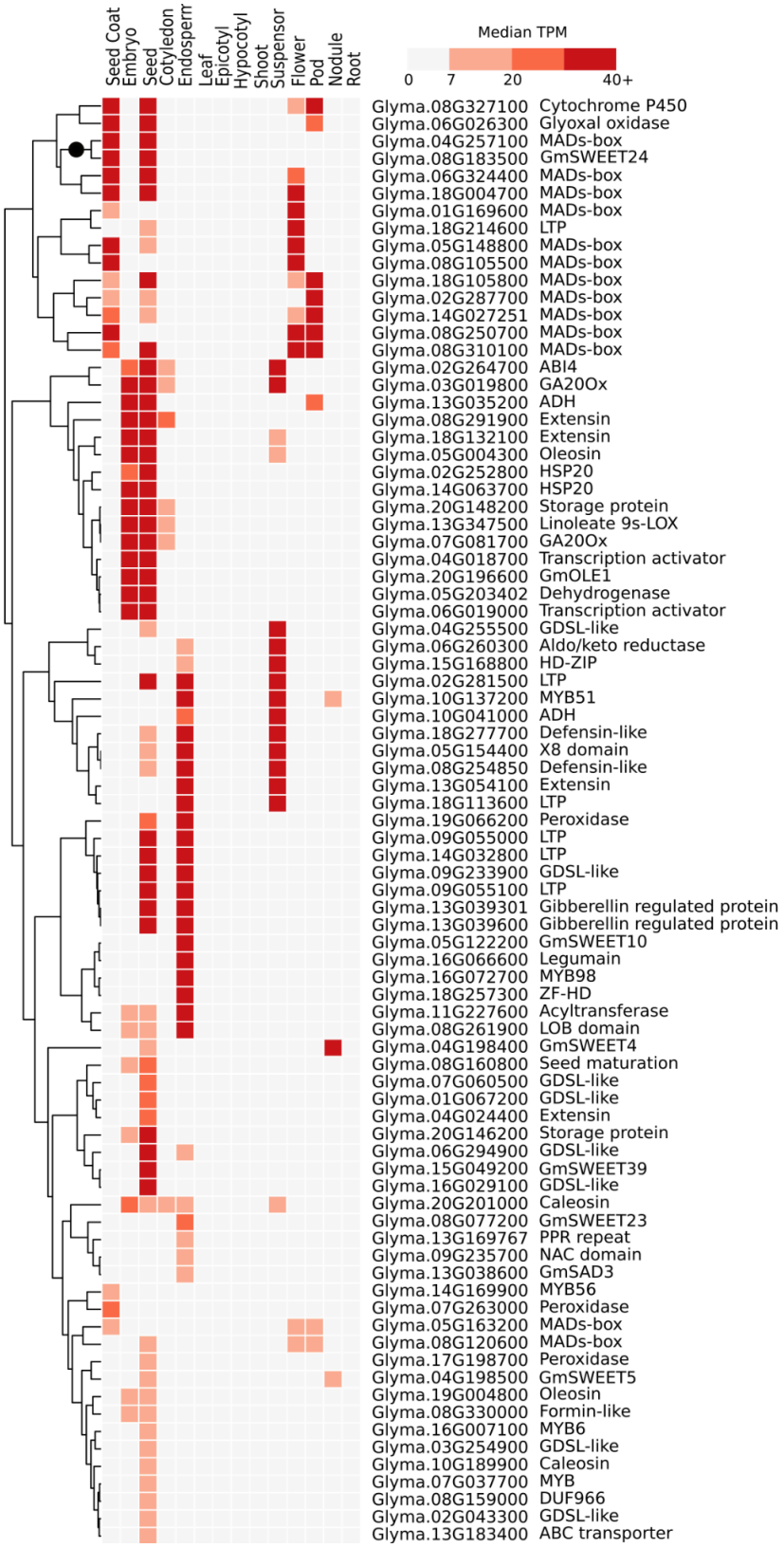


Figure 3. Expression of oil candidate genes with preferential expression in seeds. Tissue specificity metrics were obtained from the Soybean Expression Atlas (Almeida-Silva et al., 2023). Annotations were retrieved from Phytozome (V13) and Soybase.org. TPM: transcripts per million. Black circle marks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

3.3.4 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 & Venancio, 2022) (see methods for details) to compute a co-expression network (Almeida-Silva et al., 2020; Schaefer et al., 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, PLDα1: 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 (Cai et al., 2023; Fang et al., 2017; J. Huang et al., 2020; Zhou et al., 2015). 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).

3.3.5 Conservation of oil candidate genes in legumes

We employed Orthofinder (Emms & 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 & 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 (OG0035121: 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 (Jusoh et al., 2015; W. Liu et al., 1995). 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 1. Plant species used in this study, genome information and data source.

Species	Assembly	Genome size	Publication	Database
Arachis duranensis	V2.0	1.25 Gb	(Garg et al., 2022)	Legumepedia
Arachis hypogaea	V1.0	2.7 Gb	(Bertioli et al., 2016)	Phytozome 13
Arachis ipaensis	V1.0	1.56 Gb	(Q. Lu et al., 2018)	DRYAD
Cajanus cajan	V2.0	833 Mb	(Garg et al., 2022)	Legumepedia

Cercis canadensis	V1	342 Mb	(Griesmann et al., 2018)	GigaDB
Chamaecrista fasciculata	V1	429 Mb	(Griesmann et al., 2018)	GigaDB
Cicer arietinum	V2 .0	738 Mb	(Garg et al., 2022)	Legumepedia
Cicer reticulatum	_	416 Mb	(S. Gupta et al., 2017)	NCBI
Faidherbia albida	_	654 Mb	(Chang et al., 2019)	GigaDB
Glycine max	V4.0	978 Mb	(Schmutz et al., 2010)	Phytozome 13
Glycine soja	V1.1	985 Mb	(Valliyodan et al., 2019)	Phytozome 13
Lablab purpureus	_	615 Mb	(Chang et al., 2019)	GigaDB
Lotus japonicus	V3.0	472 Mb	(Sato et al., 2008)	Kazusa
Lupinus albus	V1	450 Mb	(Hufnagel et al., 2020)	Phytozome 13
Lupinus angustifolius	V1.0	924 Mb	(Hane et al., 2017)	NCBI
Medicago truncatula	V4.0	411 Mb	(Tang et al., 2014)	Phytozome 13
Mimosa pudica	V1	557 Mb	(Griesmann et al., 2018)	GigaDB
Nissolia schottii	V1	466 Mb	(Griesmann et al., 2018)	GigaDB
Phaseolus acutifolius	V1.0	512 Mb	(Moghaddam et al., 2021)	Phytozome 13
Phaseolus lunatus	V1	546 Mb	_	Phytozome 13
Phaseolus vulgaris	V2.0	537 Mb	_	Phytozome 13
Pisum sativum	V1	4 Gb	(Kreplak et al., 2019)	URGI/INRA
Senna tora	V1	547 Mb	(S.-H. Kang et al., 2020)	NCBI
Spatholobus suberectus	V1	793 Mb	(S. Qin et al., 2019)	NCBI
Trifolium pratense	V2	345 Mb	(De Vega et al., 2015)	Phytozome 13
Trifolium subterraneum	V2.0	540 Mb	(Garg et al., 2022)	Legumepedia
Vigna angularis	V1.1	466 Mb	(K. Yang et al., 2015)	NCBI
Vigna radiata	V1	463 Mb	(Y. J. Kang et al., 2014)	NCBI
Vigna subterranea	_	535 Mb	(Chang et al., 2019)	GigaDB
Vigna unguiculata	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 et al., 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 4). 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) (Morley et al., 2023; Q.-X. Song et al., 2013; Torabi et al., 2021; S. Wang et al., 2020). 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.

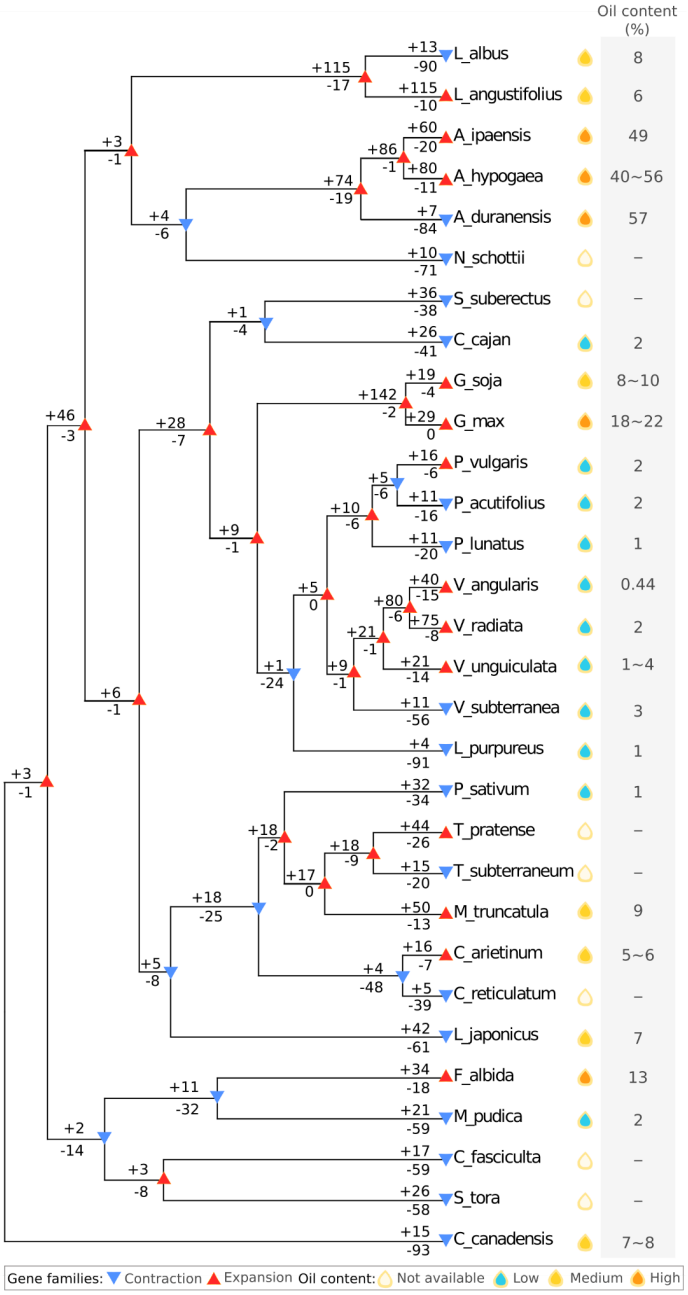


Figure 4. Species tree summarizing the number of orthogroups from gain families (~8 TPM) with significant (p-value < 0.05) expansion or contraction across 30 legume species. Gain families are those with more genes in soybean than in common bean. Red and blue triangles refer to nodes/leaves with more 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 et al., 2000); C. cajan (Sharma et al., 2011); C. canadensis (Duke & Ayensu, 1985); C. arietinum (Zia-Ul-Haq et al., 2007); F. albida (Hassan et al., 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 et al., 2004); L. angustifolius (Lemus-Conejo et al., 2023); M. truncatula (Y. Song, He, et al., 2017); M. pudica (Grygier et al., 2022); P. acutifolius (Bhardwaj et al., 2004); P. lunatus (Palupi et al., 2022); P. vulgaris (Sutivisedsak et al., 2011); P. sativum (Asen et al., 2023); V. angularis (Shweta & Katoch, 2014); V. radiata (Zia-Ul-Haq et al., 2008); V. subterranea (Minka & Bruneteau, 2000); V. unguiculata (Perchuk et al., 2020). The CAFE plot was generated with cafeplotter (https://github.com/moshi4/CafePlotter).

3.4 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).

3.5 Materials and methods

3.5.1 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).

3.5.2 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 et al., 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 & 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).

3.5.3 Gene coexpression analysis

We used the R package BioNERO (Almeida-Silva & Venancio, 2022) to construct a coexpression network of gain-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 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 & Horvath, 2008; B. Zhang & 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).

3.5.4 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 et al., 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 & Van de Peer, 2024).

3.5.5 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 & Kelly, 2019) to infer OGs using the parameters -S diamond-ultra-sens (Buchfink et al., 2021), -M msa, and -T iqtree (iqtree.org). Orthofinder results were explored using the cogeqc R/Bioconductor package (Almeida-Silva & 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).

3.6 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.

3.7 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.

3.8 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.

4 - CONSIDERAÇÕES FINAIS

A soja desempenha papel significativo na balança comercial brasileira, sendo uma crítica fonte primária de proteína e óleo vegetal. Este trabalho investigou o impacto dos eventos de WGD nas famílias gênicas relacionadas a óleo (FRO) na soja e no feijão comum, além de outras 28 espécies de leguminosas. Nossos resultados revelaram insights significativos sobre a evolução e a funcionalidade dos genes do metabolismo de óleo, destacando a complexidade das adaptações genéticas e a importância das duplicações gênicas na diversificação das funções biológicas.

Primeiramente, observamos que 91,5% das FRO se expandiram na soja em comparação com o feijão comum. Identificamos dois cenários principais de duplicação gênica: 25,5% das famílias exibiram uma relação de um gene no feijão para dois ou mais na soja (1:2+), enquanto 74,5% mostraram duplicações em ambas as espécies (2:3+). Esses padrões de duplicação indicam diferentes trajetórias evolutivas pós-WGD, com os genes de metabolismo lipídico retornando a uma única cópia após o WGD de aproximadamente 58 milhões de anos (ma) e duplicando novamente no WGD próximo a 13 ma, enquanto os genes regulatórios foram retidos como duplicatas após ambos os eventos de WGD.

Os picos observados nas distribuições de valores de K_s (substituições sinônimas por sítios sinônimos) correspondem às idades esperadas das duplicações WGD, reforçando a hipótese de que esses eventos tiveram um papel crucial na expansão e especialização das FRO na soja. Notavelmente, a expansão das famílias de biossíntese de TAG (~3:1) e a contração das famílias de degradação de TAG (~1.4:1) na soja, em comparação com o feijão, sugerem uma adaptação específica para aumentar o acúmulo de óleo na soja.

Além disso, a análise de expansão das FRO em outras 28 espécies de leguminosas revelou variações associadas aos níveis de óleo dessas espécies. Esse resultado indica que há grande complexidade no metabolismo lipídico de leguminosas, sendo necessário uma investigação a nível de espécie, como foi realizado com soja e feijão no presente trabalho. Nossas investigações possibilitam a seleção de candidatos prioritários na melhoria genética de culturas com foco no teor de óleo, orientando estratégias mais direcionadas e eficientes de edição gênica para o desenvolvimento de cultivares com maior valor econômico e nutricional.

A importância deste estudo se destaca ainda mais quando consideramos que os genes duplicados podem sofrer variações em suas funções. Duplicações gênicas fornecem matéria-prima para a evolução, permitindo que os genes duplicados adquiram novas funções (neofuncionalização) ou se especializem em funções preexistentes

(subfuncionalização). Essas variações podem levar a inovações adaptativas, como a melhoria da eficiência metabólica ou a resistência a estresses ambientais. No contexto agrícola, entender essas variações funcionais pode ser crucial para desenvolver plantas com características desejáveis, como maior teor de óleo, melhor qualidade nutricional e maior resistência a pragas e doenças.

Em síntese, este estudo não só lança luz sobre a evolução do enriquecimento de óleo em sementes de soja, mas também destaca a importância das duplicações gênicas e das adaptações específicas de espécies para a compreensão e aprimoramento do teor e qualidade do óleo em leguminosas. Os dados apresentados oferecem um conjunto de genes candidatos para futuras pesquisas e aplicações biotecnológicas voltadas para a otimização do conteúdo de óleo em soja, contribuindo para a segurança alimentar.

5- REFERÊNCIAS

Adams, K. L., Cronn, R., Percifield, R., & Wendel, J. F. (2003). Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences*, *100*(8), 4649–4654. https://doi.org/10.1073/pnas.0630618100

Akoh, C. C., Lee, G.-C., Liaw, Y.-C., Huang, T.-H., & Shaw, J.-F. (2004). GDSL family of serine esterases/lipases. *Progress in Lipid Research*, *43*(6), 534–552. https://doi.org/10.1016/j.plipres.2004.09.002

Alam, I., Batool, K., Huang, Y., Liu, J., & Ge, L. (2022). Developing Genetic Engineering Techniques for Control of Seed Size and Yield. *International Journal of Molecular Sciences*, *23*(21), Article 21. https://doi.org/10.3390/ijms232113256

Almeida-Silva, F., Moharana, K. C., Machado, F. B., & Venancio, T. M. (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, F., Pedrosa-Silva, F., & Venancio, T. M. (2023). The Soybean Expression Atlas v2: A comprehensive database of over 5000 RNA-seq samples. *The Plant Journal*, *116*(4), 1041–1051. https://doi.org/10.1111/tpj.16459

Almeida-Silva, F., & Van de Peer, Y. V. de. (2024). *doubletrouble: An R/Bioconductor package for the identification, classification, and analysis of gene and genome duplications* (p. 2024.02.27.582236). bioRxiv. https://doi.org/10.1101/2024.02.27.582236

Almeida-Silva, F., & Van de Peer, Y. (2023). Assessing the quality of comparative genomics data and results with the cogeqc R/Bioconductor package. *Methods in Ecology and Evolution*, *14*(12), 2942–2952. https://doi.org/10.1111/2041-210X.14243

Almeida-Silva, F., & Venancio, T. M. (2022). BioNERO: An all-in-one R/Bioconductor package for comprehensive and easy biological network reconstruction. *Functional & Integrative Genomics*, *22*(1), 131–136. https://doi.org/10.1007/s10142-021-00821-9

An, Y. C., Goettel, W., Han, Q., Bartels, A., Liu, Z., & Xiao, W. (2017). Dynamic Changes of Genome-Wide DNA Methylation during Soybean Seed Development. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-12510-4

Anshu, A. K., Kumar, V., Rani, A., Tayalkar, T., & Parmar, H. S. (2022). Identification of quantitative trait loci governing biosynthesis of phospholipids in soybean. *Crop and Pasture Science*, *74*(3), 219–229. https://doi.org/10.1071/CP22095

Asen, N. D., Aluko, R. E., Martynenko, A., Utioh, A., & Bhowmik, P. (2023). Yellow Field Pea Protein (Pisum sativum L.): Extraction Technologies, Functionalities, and

Applications. *Foods*, *12*(21), Article 21. https://doi.org/10.3390/foods12213978

Azani, N., Babineau, M., Bailey, C. D., Banks, H., Barbosa, A. R., Pinto, R. B., Boatwright, J. S., Borges, L. M., Brown, G. K., Bruneau, A., Candido, E., Cardoso, D., Chung, K.-F., Clark, R. P., Conceição, A. de S., Crisp, M., Cubas, P., Delgado-Salinas, A., Dexter, K. G., … Zimmerman, E. (2017). A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny: The Legume Phylogeny Working Group (LPWG). *TAXON*, *66*(1), 44–77. https://doi.org/10.12705/661.3

Aziziaram, Z., Cheghamirza, K., Zarei, L., & Beheshti-Alagha, A. (2021). Chemical and morphological characteristics of common bean seed and evaluating genetic advance in commercial classes. *Cellular and Molecular Biology*, *67*(6), Article 6. https://doi.org/10.14715/cmb/2021.67.6.13

Bai, S., Wallis, J. G., Denolf, P., Engelen, S., Bengtsson, J. D., Van Thournout, M., Dierickx, J., Haesendonckx, B., & Browse, J. (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, Y., Zhu, X., Guo X, Zhang W, Zhang G, Chen H, & 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, P., & Browse, J. (2012). The Significance of Different Diacylgycerol Synthesis Pathways on Plant Oil Composition and Bioengineering. *Frontiers in Plant Science*, *3*. https://www.frontiersin.org/articles/10.3389/fpls.2012.00147

Bates, P. 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*, *1861*(9, Part B), 1214–1225. https://doi.org/10.1016/j.bbalip.2016.03.021

Bates, P. D., & Browse, J. (2012). The Significance of Different Diacylgycerol 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, P. D., Durrett, T. P., Ohlrogge, J. B., & Pollard, M. (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, P. D., Stymne, S., & Ohlrogge, J. (2013). Biochemical pathways in seed oil synthesis. *Current Opinion in Plant Biology*, *16*(3), 358–364. https://doi.org/10.1016/j.pbi.2013.02.015

Baud, S., & Lepiniec, L. (2010). Physiological and developmental regulation of seed oil production. *Progress in Lipid Research*, *49*(3), 235–249. https://doi.org/10.1016/j.plipres.2010.01.001

Benning, C. (2009). Mechanisms of Lipid Transport Involved in Organelle Biogenesis in Plant Cells. *Annual Review of Cell and Developmental Biology*, *25*(Volume 25, 2009), 71–91. https://doi.org/10.1146/annurev.cellbio.042308.113414

Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K. S., Liu, X., Gao, D., Clevenger, J., Dash, S., Ren, L., Moretzsohn, M. C., Shirasawa, K., Huang, W., Vidigal, B., Abernathy, B., Chu, Y., Niederhuth, C. E., Umale, P., … Ozias-Akins, P. (2016). The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. *Nature Genetics*, *48*(4), Article 4. https://doi.org/10.1038/ng.3517

Bhardwaj, H. L., Hamama, A. A., & van Santen, E. (2004). Fatty acids and oil content in white lupin seed as affected by production practices. *Journal of the American Oil Chemists' Society*, *81*(11), 1035–1038. https://doi.org/10.1007/s11746-004-1018-0

Borisjuk, L., Nguyen, T. H., Neuberger, T., Rutten, T., Tschiersch, H., Claus, B., Feussner, I., Webb, A. G., Jakob, P., Weber, H., Wobus, U., & Rolletschek, H. (2005). Gradients of lipid storage, photosynthesis and plastid differentiation in developing soybean seeds. *New Phytologist*, *167*(3), 761–776. https://doi.org/10.1111/j.1469-8137.2005.01474.x

Bourgis, F., Kilaru, A., Cao, X., Ngando-Ebongue, G.-F., Drira, N., Ohlrogge, J. B., & Arondel, V. (2011). Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. *Proceedings of the National Academy of Sciences*, *108*(30), 12527–12532. https://doi.org/10.1073/pnas.1106502108

Boyang, K. E., Wenlong, L. I., & Caiying, Z. (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, A. V., Conners, S. I., Huang, W., Wilkey, A. P., Grant, D., Weeks, N. T., Cannon, S. B., Graham, M. A., & Nelson, R. T. (2021). A new decade and new data at SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Research*, *49*(D1), D1496–D1501. https://doi.org/10.1093/nar/gkaa1107

Buchfink, B., Reuter, K., & Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, *18*(4), Article 4. https://doi.org/10.1038/s41592-021-01101-x

Cagliari, A., Margis, R., Maraschin, F. dos S., Turchetto-Zolet, A. C., Loss, G., & Margis-Pinheiro, M. (2011). Biosynthesis of Triacylglycerols (TAGs) in plants and algae. *International Journal of Plant Biology*, *2*(1), Article 1. https://doi.org/10.4081/pb.2011.e10

Cai, Z., Xian, P., Cheng, Y., Zhong, Y., Yang, Y., Zhou, Q., Lian, T., Ma, Q., Nian, H., & Ge, L. (2023). MOTHER-OF-FT-AND-TFL1 regulates the seed oil and protein content in

soybean. *New Phytologist*, *239*(3), 905–919. https://doi.org/10.1111/nph.18792

Cannon, S. B., McKain, M. R., Harkess, A., Nelson, M. N., Dash, S., Deyholos, M. K., Peng, Y., Joyce, B., Stewart, C. N., Jr, Rolf, M., Kutchan, T., Tan, X., Chen, C., Zhang, Y., Carpenter, E., Wong, G. K.-S., Doyle, J. J., & Leebens-Mack, J. (2015). Multiple Polyploidy Events in the Early Radiation of Nodulating and Nonnodulating Legumes. *Molecular Biology and Evolution*, *32*(1), 193–210. https://doi.org/10.1093/molbev/msu296

Carretero-Paulet, L., & Van de Peer, Y. (2020). The evolutionary conundrum of whole-genome duplication. *American Journal of Botany*, *107*(8), 1101–1105. https://doi.org/10.1002/ajb2.1520

Celmeli, T., Sari, H., Canci, H., Sari, D., Adak, A., Eker, T., & Toker, C. (2018). The Nutritional Content of Common Bean (Phaseolus vulgaris L.) Landraces in Comparison to Modern Varieties. *Agronomy*, *8*(9), Article 9. https://doi.org/10.3390/agronomy8090166

Chang, Y., Liu, H., Liu, M., Liao, X., Sahu, S. K., Fu, Y., Song, B., Cheng, S., Kariba, R., Muthemba, S., Hendre, P. S., Mayes, S., Ho, W. K., Yssel, A. E. J., Kendabie, P., Wang, S., Li, L., Muchugi, A., Jamnadass, R., … Liu, X. (2019). The draft genomes of five agriculturally important African orphan crops. *GigaScience*, *8*(3), giy152. https://doi.org/10.1093/gigascience/giy152

Chaudhary, J., Patil, G. B., Sonah, H., Deshmukh, R. K., Vuong, T. D., Valliyodan, B., & Nguyen, H. T. (2015). Expanding Omics Resources for Improvement of Soybean Seed Composition Traits. *Frontiers in Plant Science*, *6*. https://doi.org/10.3389/fpls.2015.01021

Chen, B., Zhang, G., Li, P., Yang, J., Guo, L., Benning, C., Wang, X., & Zhao, J. (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), 155–171. https://doi.org/10.1111/pbi.13183

Chen, G., Harwood, J. L., Lemieux, M. J., Stone, S. J., & Weselake, R. J. (2022). Acyl-CoA:diacylglycerol acyltransferase: Properties, physiological roles, metabolic engineering and intentional control. *Progress in Lipid Research*, *88*, 101181. https://doi.org/10.1016/j.plipres.2022.101181

Chen, L., Zheng, Y., Dong, Z., Meng, F., Sun, X., Fan, X., Zhang, Y., Wang, M., & Wang, S. (2018). Soybean (Glycine max) WRINKLED1 transcription factor, GmWRI1a, positively regulates seed oil accumulation. *Molecular Genetics and Genomics*, *293*(2), 401–415. https://doi.org/10.1007/s00438-017-1393-2

Chen, M., Han, G., Dietrich, C. R., Dunn, T. M., & Cahoon, E. B. (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), 3576–3593. https://doi.org/10.1105/tpc.105.040774

Chen, R., Yamaoka, Y., Feng, Y., Chi, Z., Xue, S., & Kong, F. (2023). Co-Expression of Lipid Transporters Simultaneously Enhances Oil and Starch Accumulation in the Green Microalga Chlamydomonas reinhardtii under Nitrogen Starvation. *Metabolites*, *13*(1), Article 1. https://doi.org/10.3390/metabo13010115

Cheng, F., Wu, J., Cai, X., Liang, J., Freeling, M., & Wang, X. (2018). Gene retention, fractionation and subgenome differences in polyploid plants. *Nature Plants*, *4*(5), 258–268. https://doi.org/10.1038/s41477-018-0136-7

Chu, D., Zhang, Z., Hu, Y., Fang, C., Xu, X., Yuan, J., Zhang, J., Tian, Z., & Wang, G. (2023). Genome-wide scan for oil quality reveals a coregulation mechanism of tocopherols and fatty acids in soybean seeds. *Plant Communications*, *4*(5), 100598. https://doi.org/10.1016/j.xplc.2023.100598

Clark, J. W., & Donoghue, P. C. J. (2018). Whole-Genome Duplication and Plant Macroevolution. *Trends in Plant Science*, *23*(10), 933–945. https://doi.org/10.1016/j.tplants.2018.07.006

Clemente, T. E., & Cahoon, E. B. (2009). Soybean Oil: Genetic Approaches for Modification of Functionality and Total Content. *Plant Physiology*, *151*(3), 1030–1040. https://doi.org/10.1104/pp.109.146282

Collakova, E., Aghamirzaie, D., Fang, Y., Klumas, C., Tabataba, F., Kakumanu, A., Myers, E., Heath, L. S., & Grene, R. (2013). Metabolic and Transcriptional Reprogramming in Developing Soybean (Glycine max) Embryos. *Metabolites*, *3*(2), 347–372. https://doi.org/10.3390/metabo3020347

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nature Reviews Genetics*, *6*(11), 836–846. https://doi.org/10.1038/nrg1711

Conab. (2024). *Conab—Boletim da Safra de Grãos: 8º levantamento, safra 2023/2024*. http://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos

Dam, S., Laursen, B. S., Ornfelt, J. H., Jochimsen, B., Stærfeldt, H. H., Friis, C., Nielsen, K., Goffard, N., Besenbacher, S., Krusell, L., Sato, S., Tabata, S., Thogersen, I. B., Enghild, J. J., & Stougaard, J. (2009). The Proteome of Seed Development in the Model Legume Lotus japonicus. *Plant Physiology*, *149*(3), 1325–1340. https://doi.org/10.1104/pp.108.133405

De Vega, J. J., Ayling, S., Hegarty, M., Kudrna, D., Goicoechea, J. L., Ergon, A., Rognli, O. A., Jones, C., Swain, M., Geurts, R., Lang, C., Mayer, K. F. X., Rössner, S., Yates, S., Webb, K. J., Donnison, I. S., Oldroyd, G. E. D., Wing, R. A., Caccamo, M., … Skot, L. (2015). Red clover (Trifolium pratense L.) draft genome provides a platform for trait

improvement. *Scientific Reports*, *5*(1), Article 1. https://doi.org/10.1038/srep17394

Ding, L.-N., Guo, X.-J., Li, M., Fu, Z.-L., Yan, S.-Z., Zhu, K.-M., Wang, Z., & Tan, X.-L. (2019). Improving seed germination and oil contents by regulating the GDSL transcriptional level in Brassica napus. *Plant Cell Reports*, *38*(2), 243–253. https://doi.org/10.1007/s00299-018-2365-7

Duan, Z., Li, Q., Wang, H., He, X., & Zhang, M. (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

Duan, Z., Zhang, M., Zhang, Z., Liang, S., Fan, L., Yang, X., Yuan, Y., Pan, Y., Zhou, G., Liu, S., & Tian, Z. (2022). Natural allelic variation of GmST05 controlling seed size and quality in soybean. *Plant Biotechnology Journal*, *20*(9), 1807–1818. https://doi.org/10.1111/pbi.13865

Duke, J. A., & Ayensu, E. S. (1985). *Medicinal plants of China*. Reference Publications.

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y

Estep, M. C., McKain, M. R., Vela Diaz, D., Zhong, J., Hodge, J. G., Hodkinson, T. R., Layton, D. J., Malcomber, S. T., Pasquet, R., & Kellogg, E. A. (2014). Allopolyploidy, diversification, and the Miocene grassland expansion. *Proceedings of the National Academy of Sciences*, *111*(42), 15149–15154. https://doi.org/10.1073/pnas.1404177111

Fan, J., Yan, C., Roston, R., Shanklin, J., & Xu, C. (2014). Arabidopsis Lipins, PDAT1 Acyltransferase, and SDP1 Triacylglycerol Lipase Synergistically Direct Fatty Acids toward β-Oxidation, Thereby Maintaining Membrane Lipid Homeostasis. *The Plant Cell*, *26*(10), 4119–4134. https://doi.org/10.1105/tpc.114.130377

Fan, J., Yan, C., & Xu, C. (2013). Phospholipid:diacylglycerol acyltransferase-mediated triacylglycerol biosynthesis is crucial for protection against fatty acid-induced cell death in growing tissues of Arabidopsis. *The Plant Journal*, *76*(6), 930–942. https://doi.org/10.1111/tpj.12343

Fang, C., Ma, Y., Wu, S., Liu, Z., Wang, Z., Yang, R., Hu, G., Zhou, Z., Yu, H., Zhang, M., Pan, Y., Zhou, G., Ren, H., Du, W., Yan, H., Wang, Y., Han, D., Shen, Y., Liu, S., … Tian, Z. (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

Fawcett, J. A., Maere, S., & Van de Peer, Y. (2009). Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. *Proceedings of the National Academy of Sciences*, *106*(14), 5737–5742.

https://doi.org/10.1073/pnas.0900906106

Fehr, W., & Caviness, C. (1977). *Stages of soybean development*.

Fehr, W. R. (2007). Breeding for Modified Fatty Acid Composition in Soybean. *Crop Science*, *47*(S3), S-72-S-87. https://doi.org/10.2135/cropsci2007.04.0004IPBS

Finkelstein, R. R. (2010). The Role of Hormones during Seed Development and Germination. In P. J. Davies (Ed.), *Plant Hormones: Biosynthesis, Signal Transduction, Action!* (pp. 549–573). Springer Netherlands. https://doi.org/10.1007/978-1-4020-2686-7_24

Flagel, L. E., & Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytologist*, *183*(3), 557–564. https://doi.org/10.1111/j.1469-8137.2009.02923.x

Freeling, M. (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), 433–453. https://doi.org/10.1146/annurev.arplant.043008.092122

Freeling, M., Scanlon, M. J., & Fowler, J. E. (2015). Fractionation and subfunctionalization following genome duplications: Mechanisms that drive gene content and their consequences. *Current Opinion in Genetics & Development*, *35*, 110–118. https://doi.org/10.1016/j.gde.2015.11.002

Gaeta, R. T., & Chris Pires, J. (2010). Homoeologous recombination in allopolyploids: The polyploid ratchet. *New Phytologist*, *186*(1), 18–28. https://doi.org/10.1111/j.1469-8137.2009.03089.x

Garg, V., Dudchenko, O., Wang, J., Khan, A. W., Gupta, S., Kaur, P., Han, K., Saxena, R. K., Kale, S. M., Pham, M., Yu, J., Chitikineni, A., Zhang, Z., Fan, G., Lui, C., Valluri, V., Meng, F., Bhandari, A., Liu, X., … Varshney, R. K. (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*, *42*, 315–329. https://doi.org/10.1016/j.jare.2021.10.009

Gill, N., Findley, S., Walling, J. G., Hans, C., Ma, J., Doyle, J., Stacey, G., & Jackson, S. A. (2009). Molecular and Chromosomal Evidence for Allopolyploidy in Soybean. *Plant Physiology*, *151*(3), 1167–1174. https://doi.org/10.1104/pp.109.137935

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research*, *40*(D1), D1178–D1186. https://doi.org/10.1093/nar/gkr944

Graham, I. A. (2008). Seed Storage Oil Mobilization. *Annual Review of Plant Biology*, *59*(Volume 59, 2008), 115–142. https://doi.org/10.1146/annurev.arplant.59.032607.092938

Grdeń, P., & Jakubczyk, A. (2023). Health benefits of legume seeds. *Journal of the Science*

of Food and Agriculture, *103*(11), 5213–5220. https://doi.org/10.1002/jsfa.12585

Grechkin, A. N. (2002). Hydroperoxide lyase and divinyl ether synthase. *Prostaglandins & Other Lipid Mediators*, *68–69*, 457–470. https://doi.org/10.1016/S0090-6980(02)00048-5

Griesmann, M., Chang, Y., Liu, X., Song, Y., Haberer, G., Crook, M. B., Billault-Penneteau, B., Lauressergues, D., Keller, J., Imanishi, L., Roswanjaya, Y. P., Kohlen, W., Pujic, P., Battenberg, K., Alloisio, N., Liang, Y., Hilhorst, H., Salgado, M. G., Hocher, V., … Cheng, S. (2018). Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science*, *361*(6398), eaat1743. https://doi.org/10.1126/science.aat1743

Grosso, N. R., Nepote, V., & Guzmán, C. A. (2000). Chemical Composition of Some Wild Peanut Species (Arachis L.) Seeds. *Journal of Agricultural and Food Chemistry*, *48*(3), 806–809. https://doi.org/10.1021/jf9901744

Grygier, A., Chakradhari, S., Ratusz, K., Rudzińska, M., Patel, K. S., Lazdiņa, D., & Górnaś, P. (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*, 115251. https://doi.org/10.1016/j.indcrop.2022.115251

Guo, W., Chen, L., Chen, H., Yang, H., You, Q., Bao, A., Chen, S., Hao, Q., Huang, Y., Qiu, D., Shan, Z., Yang, Z., Yuan, S., Zhang, C., Zhang, X., Jiao, Y., Tran, L.-S. P., Zhou, X., & Cao, D. (2020). Overexpression of GmWRI1b in soybean stably improves plant architecture and associated yield parameters, and increases total seed oil production under field conditions. *Plant Biotechnology Journal*, *18*(8), 1639–1641. https://doi.org/10.1111/pbi.13324

Gupta, A., Atanasov, A. G., Li, Y., Kumar, N., & Bishayee, A. (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, S., Nawaz, K., Parween, S., Roy, R., Sahu, K., Kumar Pole, A., Khandal, H., Srivastava, R., Kumar Parida, S., & Chattopadhyay, D. (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, S., & Ross-Ibarra, J. (2015). Genetic, evolutionary and plant breeding insights from the domestication of maize. *ELife*, *4*, e05861. https://doi.org/10.7554/eLife.05861

Hane, J. K., Ming, Y., Kamphuis, L. G., Nelson, M. N., Garg, G., Atkins, C. A., Bayer, P. E., Bravo, A., Bringans, S., Cannon, S., Edwards, D., Foley, R., Gao, L., Harrison, M. J., Huang, W., Hurgobin, B., Li, S., Liu, C.-W., McGrath, A., … Singh, K. B. (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), 318–330. https://doi.org/10.1111/pbi.12615

Harwood, J. L. (1988). Fatty Acid Metabolism. *Annual Review of Plant Biology*, *39*(Volume 39, 1988), 101–138. https://doi.org/10.1146/annurev.pp.39.060188.000533

Harwood, J. L. (1996). Recent advances in the biosynthesis of plant fatty acids. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*, *1301*(1), 7–56. https://doi.org/10.1016/0005-2760(95)00242-1

Hassan, L. G., Umar, K. J., & Yuguda, H. (2007). *Nutritional Evaluation Of Faidherbia Albida Seeds And Pulp As Source Of Feeds For Livestock*. https://tspace.library.utoronto.ca/handle/1807/55388

Hemmati, S., & Seradj, H. (2016). Justicidin B: A Promising Bioactive Lignan. *Molecules (Basel, Switzerland)*, *21*(7), 820. https://doi.org/10.3390/molecules21070820

Hemsley, P. A. (2015). The importance of lipid modified proteins in plants. *New Phytologist*, *205*(2), 476–489. https://doi.org/10.1111/nph.13085

Herridge, D. F., Peoples, M. B., & Boddey, R. M. (2008). Global inputs of biological nitrogen fixation in agricultural systems. *Plant and Soil*, *311*(1), 1–18. https://doi.org/10.1007/s11104-008-9668-3

Hooker, J. C., Nissan, N., Luckert, D., Zapata, G., Hou, A., Mohr, R. M., Glenn, A. J., Barlow, B., Daba, K. A., Warkentin, T. D., Lefebvre, F., Golshani, A., Cober, E. R., & Samanfar, B. (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, S., Ahmed, R., Bhowmick, S., Mamun, A. A., & Hashimoto, M. (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, A. H. C. (2018). Plant Lipid Droplets and Their Associated Proteins: Potential for Rapid Advances. *Plant Physiology*, *176*(3), 1894–1918. https://doi.org/10.1104/pp.17.01677

Huang, C.-H., Zhang, C., Liu, M., Hu, Y., Gao, T., Qi, J., & Ma, H. (2016). Multiple Polyploidization Events across Asteraceae with Two Nested Events in the Early History Revealed by Nuclear Phylogenomics. *Molecular Biology and Evolution*, *33*(11), 2820–2835. https://doi.org/10.1093/molbev/msw157

Huang, J., Ma, Q., Cai, Z., Xia, Q., Li, S., Jia, J., Chu, L., Lian, T., Nian, H., & Cheng, Y. (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), 6448–6460. https://doi.org/10.1021/acs.jafc.0c01271

Huang, L., Jiang, H., Ren, X., Chen, Y., Xiao, Y., Zhao, X., Tang, M., Huang, J., Upadhyaya, H. D., & Liao, B. (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, B., Marques, A., Soriano, A., Marquès, L., Divol, F., Doumas, P., Sallet, E., Mancinotti, D., Carrere, S., Marande, W., Arribat, S., Keller, J., Huneau, C., Blein, T., Aimé, D., Laguerre, M., Taylor, J., Schubert, V., Nelson, M., … Péret, B. (2020). High-quality genome sequence of white lupin provides insight into soil exploration and seed quality. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-019-14197-9

James, A. T., & Nichols, B. W. (1966). Lipids of Photosynthetic Systems. *Nature*, *210*(5034), 372–375. https://doi.org/10.1038/210372a0

Jessen, D., Roth, C., Wiermer, M., & Fulda, M. (2015). Two Activities of Long-Chain Acyl-Coenzyme A Synthetase Are Involved in Lipid Trafficking between the Endoplasmic Reticulum and the Plastid in Arabidopsis. *Plant Physiology*, *167*(2), 351–366. https://doi.org/10.1104/pp.114.250365

Jiao, Y., Leebens-Mack, J., Ayyampalayam, S., Bowers, J. E., McKain, M. R., McNeal, J., Rolf, M., Ruzicka, D. R., Wafula, E., Wickett, N. J., Wu, X., Zhang, Y., Wang, J., Zhang, Y., Carpenter, E. J., Deyholos, M. K., Kutchan, T. M., Chanderbali, A. S., Soltis, P. S., … dePamphilis, C. W. (2012). A genome triplication associated with early diversification of the core eudicots. *Genome Biology*, *13*(1), R3. https://doi.org/10.1186/gb-2012-13-1-r3

Jiao, Y., Li, J., Tang, H., & Paterson, A. H. (2014). Integrated Syntenic and Phylogenomic Analyses Reveal an Ancient Genome Duplication in Monocots. *The Plant Cell*, *26*(7), 2792–2802. https://doi.org/10.1105/tpc.114.127597

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., Tomsho, L. P., Hu, Y., Liang, H., Soltis, P. S., Soltis, D. E., Clifton, S. W., Schlarbaum, S. E., Schuster, S. C., Ma, H., Leebens-Mack, J., & dePamphilis, C. W. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature*, *473*(7345), 97–100. https://doi.org/10.1038/nature09916

Jing, G., Tang, D., Yao, Y., Su, Y., Shen, Y., Bai, Y., Jing, W., Zhang, Q., Lin, F., Guo, D., & Zhang, W. (2021). Seed specifically over-expressing *DGAT2A* enhances oil and linoleic acid contents in soybean seeds. *Biochemical and Biophysical Research Communications*, *568*, 143–150. https://doi.org/10.1016/j.bbrc.2021.06.087

Jusoh, M., Loh, S. H., Chuah, T. S., Aziz, A., & Cha, T. S. (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*, 65–71. https://doi.org/10.1016/j.phytochem.2014.12.022

Kagale, S., Robinson, S. J., Nixon, J., Xiao, R., Huebert, T., Condie, J., Kessler, D., Clarke, W. E., Edger, P. P., Links, M. G., Sharpe, A. G., & Parkin, I. A. P. (2014). Polyploid Evolution of the Brassicaceae during the Cenozoic Era. *The Plant Cell*, *26*(7), 2777–2791. https://doi.org/10.1105/tpc.114.126391

Kanai, M., Yamada, T., Hayashi, M., Mano, S., & Nishimura, M. (2019). Soybean (Glycine max L.) triacylglycerol lipase GmSDP1 regulates the quality and quantity of seed oil. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-45331-8

Kang, S.-H., Pandey, R. P., Lee, C.-M., Sim, J.-S., Jeong, J.-T., Choi, B.-S., Jung, M., Ginzburg, D., Zhao, K., Won, S. Y., Oh, T.-J., Yu, Y., Kim, N.-H., Lee, O. R., Lee, T.-H., Bashyal, P., Kim, T.-S., Lee, W.-H., Hawkins, C., … Sohng, J. K. (2020). Genome-enabled discovery of anthraquinone biosynthesis in Senna tora. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-19681-1

Kang, Y. J., Kim, S. K., Kim, M. Y., Lestari, P., Kim, K. H., Ha, B.-K., Jun, T. H., Hwang, W. J., Lee, T., Lee, J., Shim, S., Yoon, M. Y., Jang, Y. E., Han, K. S., Taeprayoon, P., Yoon, N., Somta, P., Tanya, P., Kim, K. S., … Lee, S.-H. (2014). Genome sequence of mungbean and insights into evolution within Vigna species. *Nature Communications*, *5*(1), Article 1. https://doi.org/10.1038/ncomms6443

Kim, K. D., El Baidouri, M., Abernathy, B., Iwata-Otsubo, A., Chavarro, C., Gonzales, M., Libault, M., Grimwood, J., & Jackson, S. A. (2015). A Comparative Epigenomic Analysis of Polyploidy-Derived Genes in Soybean and Common Bean. *Plant Physiology*, *168*(4), 1433–1447. https://doi.org/10.1104/pp.15.00408

Kim, S., Yamaoka, Y., Ono, H., Kim, H., Shim, D., Maeshima, M., Martinoia, E., Cahoon, E. B., Nishida, I., & Lee, Y. (2013). AtABCA9 transporter supplies fatty acids for lipid synthesis to the endoplasmic reticulum. *Proceedings of the National Academy of Sciences*, *110*(2), 773–778. https://doi.org/10.1073/pnas.1214159110

Kim, S.-C., & Wang, X. (2020). Phosphatidic acid: An emerging versatile class of cellular mediators. *Essays in Biochemistry*, *64*(3), 533–546. https://doi.org/10.1042/EBC20190089

Kreplak, J., Madoui, M.-A., Cápal, P., Novák, P., Labadie, K., Aubert, G., Bayer, P. E., Gali, K. K., Syme, R. A., Main, D., Klein, A., Bérard, A., Vrbová, I., Fournier, C., d'Agata, L., Belser, C., Berrabah, W., Toegelová, H., Milec, Z., … Burstin, J. (2019). A reference genome for pea provides insight into legume genome evolution. *Nature Genetics*, *51*(9), Article 9. https://doi.org/10.1038/s41588-019-0480-1

Kryuchkova-Mostacci, N., & Robinson-Rechavi, M. (2017). A benchmark of gene expression tissue-specificity metrics. *Briefings in Bioinformatics*, *18*(2), 205–214.

https://doi.org/10.1093/bib/bbw008

Kumar, S., Suleski, M., Craig, J. M., Kasprowicz, A. E., Sanderford, M., Li, M., Stecher, G., & Hedges, S. B. (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, P., & Horvath, S. (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, M., Herendeen, P. S., & Wojciechowski, M. F. (2005). Evolutionary Rates Analysis of Leguminosae Implicates a Rapid Diversification of Lineages during the Tertiary. *Systematic Biology*, *54*(4), 575–594. https://doi.org/10.1080/10635150590947131

Lee, E.-J., Oh, M., Hwang, J.-U., Li-Beisson, Y., Nishida, I., & Lee, Y. (2017). Seed-Specific Overexpression of the Pyruvate Transporter BASS2 Increases Oil Content in Arabidopsis Seeds. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.00194

Lee, J., Welti, R., Schapaugh, W. T., & Trick, H. N. (2011). Phospholipid and triacylglycerol profiles modified by PLD suppression in soybean seed. *Plant Biotechnology Journal*, *9*(3), 359–372. https://doi.org/10.1111/j.1467-7652.2010.00562.x

Lemus-Conejo, A., Rivero-Pino, F., Montserrat-de la Paz, S., & Millan-Linares, M. C. (2023). Nutritional composition and biological activity of narrow-leafed lupins (*Lupinus angustifolius* L.) hydrolysates and seeds. *Food Chemistry*, *420*, 136104. https://doi.org/10.1016/j.foodchem.2023.136104

Li, J., Zhang, Y., Ma, R., Huang, W., Hou, J., Fang, C., Wang, L., Yuan, Z., Sun, Q., Dong, X., Hou, Y., Wang, Y., Kong, F., & Sun, L. (2022). Identification of ST1 reveals a selection involving hitchhiking of seed morphology and oil content during soybean domestication. *Plant Biotechnology Journal*, *20*(6), 1110–1121. https://doi.org/10.1111/pbi.13791

Li, J., Zhao, J., Li, Y., Gao, Y., Hua, S., Nadeem, M., Sun, G., Zhang, W., Hou, J., Wang, X., & Qiu, L. (2019). Identification of a novel seed size associated locus *SW9-1* in soybean. *The Crop Journal*, *7*(4), 548–559. https://doi.org/10.1016/j.cj.2018.12.010

Li, N., Meng, H., Li, S., Zhang, Z., Zhao, X., Wang, S., Liu, A., Li, Q., Song, Q., Li, X., Guo, L., Li, H., Zuo, J., & Luo, K. (2020). Two Plastid Fatty Acid Exporters Contribute to Seed Oil Accumulation in Arabidopsis1. *Plant Physiology*, *182*(4), 1910–1919. https://doi.org/10.1104/pp.19.01344

Li, N., Xu, C., Li-Beisson, Y., & Philippar, K. (2016). Fatty Acid and Lipid Transport in Plant Cells. *Trends in Plant Science*, *21*(2), 145–158. https://doi.org/10.1016/j.tplants.2015.10.011

Li, Q.-T., Lu, X., Song, Q.-X., Chen, H.-W., Wei, W., Tao, J.-J., Bian, X.-H., Shen, M., Ma, B., Zhang, W.-K., Bi, Y.-D., Li, W., Lai, Y.-C., Lam, S.-M., Shui, G.-H., Chen, S.-Y., & Zhang, J.-S. (2017). Selection for a Zinc-Finger Protein Contributes to Seed Oil Increase during Soybean Domestication. *Plant Physiology*, *173*(4), 2208–2224. https://doi.org/10.1104/pp.16.01610

Li, R., Yu, K., & Hildebrand, D. F. (2010). DGAT1, DGAT2 and PDAT Expression in Seeds and Other Tissues of Epoxy and Hydroxy Fatty Acid Accumulating Plants. *Lipids*, *45*(2), 145–157. https://doi.org/10.1007/s11745-010-3385-4

Li, S.-Y., Zhang, Q., Jin, Y.-H., Zou, J.-X., Zheng, Y.-S., & Li, D.-D. (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), 1505–1516. https://doi.org/10.1007/s00299-020-02579-z

Li, Z., McKibben, M. T. W., Finch, G. S., Blischak, P. D., Sutherland, B. L., & Barker, M. S. (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, Y., Neunzig, J., Lee, Y., & Philippar, K. (2017). Plant membrane-protein mediated intracellular traffic of fatty acids and acyl lipids. *Current Opinion in Plant Biology*, *40*, 138–146. https://doi.org/10.1016/j.pbi.2017.09.006

Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M. X., Arondel, V., Bates, P. D., Baud, S., Bird, D., DeBono, A., & Durrett, T. P. (2013). Acyl-lipid metabolism. *The Arabidopsis Book/American Society of Plant Biologists*, *11*.

Lin, L.-J., Tai, S. S. K., Peng, C.-C., & Tzen, J. T. C. (2002). Steroleosin, a Sterol-Binding Dehydrogenase in Seed Oil Bodies. *Plant Physiology*, *128*(4), 1200–1211. https://doi.org/10.1104/pp.010982

Liu, A., Cheng, S.-S., Yung, W.-S., Li, M.-W., & Lam, H.-M. (2022). Chapter Nine—Genetic regulations of the oil and protein contents in soybean seeds and strategies for improvement. In H.-M. Lam & M.-W. Li (Eds.), *Advances in Botanical Research* (Vol. 102, pp. 259–293). Academic Press. https://doi.org/10.1016/bs.abr.2022.03.002

Liu, C., Chen, H., Yu, Q., Gu, H., Li, Y., Tu, B., Zhang, H., Zhang, Q., & Liu, X. (2023). Identification of quantitative trait loci and candidate genes for seed sucrose and soluble sugar concentrations in soybean. *Crop Science*, *63*(5), 2976–2992. https://doi.org/10.1002/csc2.21080

Liu, H., Wei, L., Zhu, J., Zhang, B., Gan, Y., & Zheng, Y. (2022). Identification of GmGPATs and their effect on glycerolipid biosynthesis through seed-specific expression in soybean. *Molecular Biology Reports*, *49*(10), 9585–9592. https://doi.org/10.1007/s11033-022-07852-w

Liu, J.-Y., Li, P., Zhang, Y.-W., Zuo, J.-F., Li, G., Han, X., Dunwell, J. M., & Zhang, Y.-M. (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), 1103–1124. https://doi.org/10.1111/tpj.14788

Liu, J.-Y., Zhang, Y.-W., Han, X., Zuo, J.-F., Zhang, Z., Shang, H., Song, Q., & Zhang, Y.-M. (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, P., Xie, T., Wu, X., Han, G., Gupta, S. D., Zhang, Z., Yue, J., Dong, F., Gable, K., Niranjanakumari, S., Li, W., Wang, L., Liu, W., Yao, R., Cahoon, E. B., Dunn, T. M., & Gong, X. (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, W., Hildebrand, D. F., & Collins, G. B. (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

Liu, W. X., Liu, H. L., & Qu, L. Q. (2013). Embryo-specific expression of soybean oleosin altered oil body morphogenesis and increased lipid content in transgenic rice seeds. *Theoretical and Applied Genetics*, *126*(9), 2289–2297. https://doi.org/10.1007/s00122-013-2135-4

Liu, Y.-F., Li, Q.-T., Lu, X., Song, Q.-X., Lam, S.-M., Zhang, W.-K., Ma, B., Lin, Q., Man, W.-Q., Du, W.-G., Shui, G.-H., Chen, S.-Y., & Zhang, J.-S. (2014). Soybean GmMYB73 promotes lipid accumulation in transgenic plants. *BMC Plant Biology*, *14*(1), 73. https://doi.org/10.1186/1471-2229-14-73

Lonardi, S., Muñoz-Amatriaín, M., Liang, Q., Shu, S., Wanamaker, S. I., Lo, S., Tanskanen, J., Schulman, A. H., Zhu, T., Luo, M.-C., Alhakami, H., Ounit, R., Hasan, A. Md., Verdier, J., Roberts, P. A., Santos, J. R. P., Ndeve, A., Doležel, J., Vrána, J., … Close, T. J. (2019). The genome of cowpea (Vigna unguiculata [L.] Walp.). *The Plant Journal*, *98*(5), 767–782. https://doi.org/10.1111/tpj.14349

Lu, C., Xin, Z., Ren, Z., Miquel, M., & Browse, J. (2009). An enzyme regulating triacylglycerol composition is encoded by the ROD1 gene of Arabidopsis. *Proceedings of the National Academy of Sciences*, *106*(44), 18837–18842. https://doi.org/10.1073/pnas.0908848106

Lu, L., Wei, W., Li, Q.-T., Bian, X.-H., Lu, X., Hu, Y., Cheng, T., Wang, Z.-Y., Jin, M., Tao, J.-J., Yin, C.-C., He, S.-J., Man, W.-Q., Li, W., Lai, Y.-C., Zhang, W.-K., Chen, S.-Y., & Zhang, J.-S. (2021). A transcriptional regulatory module controls lipid accumulation in soybean. *New Phytologist*, *231*(2), 661–678. https://doi.org/10.1111/nph.17401

Lu, Q., Li, H., Hong, Y., Zhang, G., Wen, S., Li, X., Zhou, G., Li, S., Liu, H., Liu, H., Liu, Z., Varshney, R. K., Chen, X., & Liang, X. (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, S., Fang, C., Abe, J., Kong, F., & Liu, B. (2022). Current overview on the genetic basis of key genes involved in soybean domestication. *ABIOTECH*, *3*(2), 126–139. https://doi.org/10.1007/s42994-022-00074-5

Lu, X., Li, Q.-T., Xiong, Q., Li, W., Bi, Y.-D., Lai, Y.-C., Liu, X.-L., Man, W.-Q., Zhang, W.-K., Ma, B., Chen, S.-Y., & Zhang, J.-S. (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), 530–544. https://doi.org/10.1111/tpj.13181

Lynch, M., & Conery, J. S. (2000). The Evolutionary Fate and Consequences of Duplicate Genes. *Science*, *290*(5494), 1151–1155. https://doi.org/10.1126/science.290.5494.1151

Lynch, M., & Force, A. (2000). The Probability of Duplicate Gene Preservation by Subfunctionalization. *Genetics*, *154*(1), 459–473. https://doi.org/10.1093/genetics/154.1.459

Manan, S., Ahmad, M. Z., Zhang, G., Chen, B., Haq, B. U., Yang, J., & Zhao, J. (2017). Soybean LEC2 Regulates Subsets of Genes Involved in Controlling the Biosynthesis and Catabolism of Seed Storage Substances and Seed Development. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.01604

Manan, S., Chen, B., She, G., Wan, X., & Zhao, J. (2017). Transport and transcriptional regulation of oil production in plants. *Critical Reviews in Biotechnology*, *37*(5), 641–655. https://doi.org/10.1080/07388551.2016.1212185

Marchive, C., Nikovics, K., To, A., Lepiniec, L., & Baud, S. (2014). Transcriptional regulation of fatty acid production in higher plants: Molecular bases and biotechnological outcomes. *European Journal of Lipid Science and Technology*, *116*(10), 1332–1343. https://doi.org/10.1002/ejlt.201400027

Martin, C. A., Visentainer, J. V., Oliveira, A. N. de, Oliveira, C. C. de, Matsushita, M., & Souza, N. E. de. (2008). Fatty acid contents of Brazilian soybean oils with emphasis on trans fatty acids. *Journal of the Brazilian Chemical Society*, *19*, 117–122. https://doi.org/10.1590/S0103-50532008000100017

Mason, A. S., & Pires, J. C. (2015). Unreduced gametes: Meiotic mishap or evolutionary mechanism? *Trends in Genetics*, *31*(1), 5–10. https://doi.org/10.1016/j.tig.2014.09.011

McGlew, K., Shaw, V., Zhang, M., Kim, R. J., Yang, W., Shorrosh, B., Suh, M. C., &

Ohlrogge, J. (2015). An annotated database of Arabidopsis mutants of acyl lipid metabolism. *Plant Cell Reports*, *34*(4), 519–532. https://doi.org/10.1007/s00299-014-1710-8

Meinke, D. W., Chen, J., & Beachy, R. N. (1981). Expression of storage-protein genes during soybean seed development. *Planta*, *153*(2), 130–139. https://doi.org/10.1007/BF00384094

Mendes, F. K., Vanderpool, D., Fulton, B., & Hahn, M. W. (2020). CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics (Oxford, England)*, btaa1022. https://doi.org/10.1093/bioinformatics/btaa1022

Messina, M., & Messina, V. (2010). The Role of Soy in Vegetarian Diets. *Nutrients*, *2*(8), Article 8. https://doi.org/10.3390/nu2080855

Miao, L., Yang, S., Zhang, K., He, J., Wu, C., Ren, Y., Gai, J., & Li, Y. (2020). Natural variation and selection in GmSWEET39 affect soybean seed oil content. *New Phytologist*, *225*(4), 1651–1666. https://doi.org/10.1111/nph.16250

Michaelson, L. V., Napier, J. A., Molino, D., & Faure, J.-D. (2016). Plant sphingolipids: Their importance in cellular organization and adaption. *Biochimica et Biophysica Acta*, *1861*(9Part B), 1329–1335. https://doi.org/10.1016/j.bbalip.2016.04.003

Minka, S. R., & Bruneteau, M. (2000). Partial chemical composition of bambara pea [*Vigna subterranea* (L.) *Verde*]. *Food Chemistry*, *68*(3), 273–276. https://doi.org/10.1016/S0308-8146(99)00186-7

Moghaddam, S. M., Oladzad, A., Koh, C., Ramsay, L., Hart, J. P., Mamidi, S., Hoopes, G., Sreedasyam, A., Wiersma, A., Zhao, D., Grimwood, J., Hamilton, J. P., Jenkins, J., Vaillancourt, B., Wood, J. C., Schmutz, J., Kagale, S., Porch, T., Bett, K. E., … McClean, P. E. (2021). The tepary bean genome provides insight into evolution and domestication under heat stress. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-22858-x

Moharana, K. C., & Venancio, T. M. (2020). Polyploidization events shaped the transcription factor repertoires in legumes (Fabaceae). *The Plant Journal: For Cell and Molecular Biology*, *103*(2), 726–741. https://doi.org/10.1111/tpj.14765

Morley, S. A., Ma, F., Alazem, M., Frankfater, C., Yi, H., Burch-Smith, T., Clemente, T. E., Veena, V., Nguyen, H., & Allen, D. K. (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

Nelson, D. L., & Cox, M. M. (2010). *Princípios de bioquímica de Lehninger* (5th ed.). Artmed.

Nguyen, Q. T., Kisiala, A., Andreas, P., Emery, R. J. N., & Narine, S. (2016). Soybean Seed Development: Fatty Acid and Phytohormone Metabolism and Their Interactions. *Current Genomics*, *17*(3), 241.https://doi.org/10.2174/1389202917666160202220238

Niu, Y., Zhang, G., Wan, F., & Zhang, Y.-M. (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

Ohlrogge, J. B., & Jaworski, J. G. (1997). REGULATION OF FATTY ACID SYNTHESIS. *Annual Review of Plant Biology*, *48*(Volume 48, 1997), 109–136. https://doi.org/10.1146/annurev.arplant.48.1.109

Ohlrogge, J., & Browse, J. (1995). Lipid biosynthesis. *The Plant Cell*, *7*(7), 957–970. https://doi.org/10.1105/tpc.7.7.957

O'Rourke, J. A., Iniguez, L. P., Fu, F., Bucciarelli, B., Miller, S. S., Jackson, S. A., McClean, P. E., Li, J., Dai, X., Zhao, P. X., Hernandez, G., & Vance, C. P. (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

Otto, S. P. (2007). The Evolutionary Consequences of Polyploidy. *Cell*, *131*(3), 452–462. https://doi.org/10.1016/j.cell.2007.10.022

Palupi, H. T., Estiasih, T., Yunianta, & Sutrisno, A. (2022). Physicochemical and protein characterization of lima bean (Phaseolus lunatus L) seed. *Food Research*, *6*(1), 168–177. https://doi.org/10.26656/fr.2017.6(1).107

Pan, X., Peng, F. Y., & Weselake, R. J. (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 Paralogs. *Plant Physiology*, *167*(3), 887–904. https://doi.org/10.1104/pp.114.253658

Panchy, N., Lehti-Shiu, M., & Shiu, S.-H. (2016). Evolution of Gene Duplication in Plants1[OPEN]. *Plant Physiology*, *171*(4), 2294–2316. https://doi.org/10.1104/pp.16.00523

Patil, G., Vuong, T. D., Kale, S., Valliyodan, B., Deshmukh, R., Zhu, C., Wu, X., Bai, Y., Yungbluth, D., Lu, F., Kumpatla, S., Shannon, J. G., Varshney, R. K., & Nguyen, H. T. (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), 1939–1953. https://doi.org/10.1111/pbi.12929

Pattee, H. E., Salunkhe, D. K., Sathe, S. K., Reddy, N. R., & Ory, R. L. (1983). Legume lipids. *C R C Critical Reviews in Food Science and Nutrition*, *17*(2), 97–139. https://doi.org/10.1080/10408398209527345

Pelletier, J. M., Kwong, R. W., Park, S., Le, B. H., Baden, R., Cagliari, A., Hashimoto, M., Munoz, M. D., Fischer, R. L., Goldberg, R. B., & Harada, J. J. (2017). LEC1

sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. *Proceedings of the National Academy of Sciences*, *114*(32), E6710–E6719. https://doi.org/10.1073/pnas.1707957114

Perchuk, I., Shelenga, T., Gurkina, M., Miroshnichenko, E., & Burlyaeva, M. (2020). Composition of Primary and Secondary Metabolite Compounds in Seeds and Pods of Asparagus Bean (Vigna unguiculata (L.) Walp.) from China. *Molecules*, *25*(17), Article 17. https://doi.org/10.3390/molecules25173778

Pham, A.-T., Shannon, J. G., & Bilyeu, K. D. (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), 503–515. https://doi.org/10.1007/s00122-012-1849-z

Pyc, M., Gidda, S. K., Seay, D., Esnay, N., Kretzschmar, F. K., Cai, Y., Doner, N. M., Greer, M. S., Hull, J. J., Coulon, D., Bréhélin, C., Yurchenko, O., de Vries, J., Valerius, O., Braus, G. H., Ischebeck, T., Chapman, K. D., Dyer, J. M., & Mullen, R. T. (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, Z., Guo, C., Li, H., Qiu, H., Li, H., Jong, C., Yu, G., Zhang, Y., Hu, L., Wu, X., Xin, D., Yang, M., Liu, C., Lv, J., Wang, X., Kong, F., & Chen, Q. (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, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., Zhang, S., & Paterson, A. H. (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, P., Wang, T., & Luo, Y. (2022). A review on plant-based proteins from soybean: Health benefits and soy product development. *Journal of Agriculture and Food Research*, *7*, 100265. https://doi.org/10.1016/j.jafr.2021.100265

Qin, S., Wu, L., Wei, K., Liang, Y., Song, Z., Zhou, X., Wang, S., Li, M., Wu, Q., Zhang, K., Hui, Y., Wang, S., Miao, J., & Zhang, Z. (2019). A draft genome for Spatholobus suberectus. *Scientific Data*, *6*(1), Article 1. https://doi.org/10.1038/s41597-019-0110-x

Rackis, J. J., Sessa, D. J., & Honig, D. H. (1979). Flavor problems of vegetable food proteins. *Journal of the American Oil Chemists' Society*, *56*(3), 262–271. https://doi.org/10.1007/BF02671470

Rahman, S., Islam, S., Yu, Z., She, M., Nevo, E., & Ma, W. (2020). Current Progress in Understanding and Recovering the Wheat Genes Lost in Evolution and Domestication. *International Journal of Molecular Sciences*, *21*(16), Article 16. https://doi.org/10.3390/ijms21165836

Ramsey, J., & Ramsey, T. S. (2014). Ecological studies of polyploidy in the 100 years following its discovery. *Philosophical Transactions of the Royal Society B: Biological*

Sciences, *369*(1648), 20130352. https://doi.org/10.1098/rstb.2013.0352

Ren, R., Wang, H., Guo, C., Zhang, N., Zeng, L., Chen, Y., Ma, H., & Qi, J. (2018). Widespread Whole Genome Duplications Contribute to Genome Complexity and Species Diversity in Angiosperms. *Molecular Plant*, *11*(3), 414–428. https://doi.org/10.1016/j.molp.2018.01.002

Renny-Byfield, S., & Wendel, J. F. (2014). Doubling down on genomes: Polyploidy and crop plants. *American Journal of Botany*, *101*(10), 1711–1725. https://doi.org/10.3732/ajb.1400119

Reszczyńska, E., & Hanaka, A. (2020). Lipids Composition in Plant Membranes. *Cell Biochemistry and Biophysics*, *78*(4), 401–414. https://doi.org/10.1007/s12013-020-00947-w

Roscoe, T. T., Guilleminot, J., Bessoule, J.-J., Berger, F., & Devic, M. (2015). Complementation of Seed Maturation Phenotypes by Ectopic Expression of ABSCISIC ACID INSENSITIVE3, FUSCA3 and LEAFY COTYLEDON2 in Arabidopsis. *Plant and Cell Physiology*, *56*(6), 1215–1228. https://doi.org/10.1093/pcp/pcv049

Rotundo, J. L., Marshall, R., McCormick, R., Truong, S. K., Styles, D., Gerde, J. A., Gonzalez-Escobar, E., Carmo-Silva, E., Janes-Bassett, V., Logue, J., Annicchiarico, P., de Visser, C., Dind, A., Dodd, I. C., Dye, L., Long, S. P., Lopes, M. S., Pannecoucque, J., Reckling, M., … Rufino, M. C. (2024). European soybean to benefit people and the environment. *Scientific Reports*, *14*(1), 7612. https://doi.org/10.1038/s41598-024-57522-z

Roy, S., Liu, W., Nandety, R. S., Crook, A., Mysore, K. S., Pislariu, C. I., Frugoli, J., Dickstein, R., & Udvardi, M. K. (2020). Celebrating 20 Years of Genetic Discoveries in Legume Nodulation and Symbiotic Nitrogen Fixation. *The Plant Cell*, *32*(1), 15–41. https://doi.org/10.1105/tpc.19.00279

Saadat, F. (2023). A computational study on the structure–function relationships of plant caleosins. *Scientific Reports*, *13*(1), 72. https://doi.org/10.1038/s41598-022-26936-y

Salie, M. J., & Thelen, J. J. (2016). Regulation and structure of the heteromeric acetyl-CoA carboxylase. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, *1861*(9, Part B), 1207–1213. https://doi.org/10.1016/j.bbalip.2016.04.004

Sandhu, D., Alt, J. L., Scherder, C. W., Fehr, W. R., & Bhattacharyya, M. K. (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, M. E., Diaz-Mendoza, M., Diaz, I., & Martinez, M. (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

Santos-Mendoza, M., Dubreucq, B., Baud, S., Parcy, F., Caboche, M., & Lepiniec, L. (2008). Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. *The Plant Journal*, *54*(4), 608–620. https://doi.org/10.1111/j.1365-313X.2008.03461.x

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., Sasamoto, S., Watanabe, A., Ono, A., Kawashima, K., Fujishiro, T., Katoh, M., Kohara, M., Kishida, Y., Minami, C., Nakayama, S., Nakazaki, N., Shimizu, Y., Shinpo, S., … Tabata, S. (2008). Genome Structure of the Legume, Lotus japonicus. *DNA Research*, *15*(4), 227–239. https://doi.org/10.1093/dnares/dsn008

Schaefer, R. J., Michno, J.-M., & Myers, C. L. (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.bbagrm.2016.07.016

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J. J., Cheng, J., Xu, D., Hellsten, U., May, G. D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M. K., Sandhu, D., Valliyodan, B., … Jackson, S. A. (2010). Genome sequence of the palaeopolyploid soybean. *Nature*, *463*(7278), Article 7278. https://doi.org/10.1038/nature08670

Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., Jenkins, J., Shu, S., Song, Q., Chavarro, C., Torres-Torres, M., Geffroy, V., Moghaddam, S. M., Gao, D., Abernathy, B., Barry, K., Blair, M., Brick, M. A., Chovatia, M., … Jackson, S. A. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics*, *46*(7), Article 7. https://doi.org/10.1038/ng.3008

Seijo, G., Lavia, G. I., Fernández, A., Krapovickas, A., Ducasse, D. A., Bertioli, D. J., & Moscone, E. A. (2007). Genomic relationships between the cultivated peanut (Arachis hypogaea, Leguminosae) and its close relatives revealed by double GISH. *American Journal of Botany*, *94*(12), 1963–1971. https://doi.org/10.3732/ajb.94.12.1963

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (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, S., Nidhi, A., & Preeti, V. (2011). *Pigeon pea (Cajanus cajan L.): A hidden treasure of regime nutrition*. https://www.indianjournals.com/ijor.aspx?target=ijor:jfeb&volume=1&issue=2&article=

002

Shasidhar, Y., Vishwakarma, M. K., Pandey, M. K., Janila, P., Variath, M. T., Manohar, S. S., Nigam, S. N., Guo, B., & Varshney, R. K. (2017). Molecular Mapping of Oil Content and Fatty Acids Using Dense Genetic Maps in Groundnut (Arachis hypogaea L.). *Frontiers in Plant Science*, *8*, 794. https://doi.org/10.3389/fpls.2017.00794

Shen, Y., Hong, S., & Li, Y. (2022). Chapter Three - Pea protein composition, functionality, modification, and food applications: A review. In J. Wu (Ed.), *Advances in Food and Nutrition Research* (Vol. 101, pp. 71–127). Academic Press. https://doi.org/10.1016/bs.afnr.2022.02.002

Shimizu-Inatsugi, R., Terada, A., Hirose, K., Kudoh, H., Sese, J., & Shimizu, K. K. (2017). Plant adaptive radiation mediated by polyploid plasticity in transcriptomes. *Molecular Ecology*, *26*(1), 193–207. https://doi.org/10.1111/mec.13738

Shirasawa, K., Isobe, S., Tabata, S., & Hirakawa, H. (2014). Kazusa Marker DataBase: A database for genomics, genetics, and molecular breeding in plants. *Breeding Science*, *64*(3), 264–271. https://doi.org/10.1270/jsbbs.64.264

Shoemaker, R. C., Schlueter, J., & Doyle, J. J. (2006). Paleopolyploidy and gene duplication in soybean and other legumes. *Current Opinion in Plant Biology*, *9*(2), 104–109. https://doi.org/10.1016/j.pbi.2006.01.007

Shu, Y., Yu, D., Wang, D., Guo, D., & Guo, C. (2013). Genome-wide survey and expression analysis of the MADS-box gene family in soybean. *Molecular Biology Reports*, *40*(6), 3901–3911. https://doi.org/10.1007/s11033-012-2438-6

Shweta, & Katoch, R. (2014). *Comparison of nutritional composition of Vigna spp. Prevalent in the mid-hill region of Himachal Pradesh*. *27*, 202–207.

Siloto, R. M. P., Findlay, K., Lopez-Villalobos, A., Yeung, E. C., Nykiforuk, C. L., & Moloney, M. M. (2006). The Accumulation of Oleosins Determines the Size of Seed Oilbodies in Arabidopsis. *The Plant Cell*, *18*(8), 1961–1974. https://doi.org/10.1105/tpc.106.041269

Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., Sankoff, D., de Pamphilis, C. W., Wall, P. K., & Soltis, P. S. (2009). Polyploidy and angiosperm diversification. *American Journal of Botany*, *96*(1), 336–348. https://doi.org/10.3732/ajb.0800079

Song, H., Taylor, D. C., & Zhang, M. (2023). Bioengineering of Soybean Oil and Its Impact on Agronomic Traits. *International Journal of Molecular Sciences*, *24*(3), Article 3. https://doi.org/10.3390/ijms24032256

Song, H., Wang, P., Li, C., Han, S., Lopez-Baltazar, J., Zhang, X., & Wang, X. (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, Q.-X., Li, Q.-T., Liu, Y.-F., Zhang, F.-X., Ma, B., Zhang, W.-K., Man, W.-Q., Du, W.-G., Wang, G.-D., Chen, S.-Y., & Zhang, J.-S. (2013). Soybean GmbZIP123 gene enhances lipid content in the seeds of transgenic Arabidopsis plants. *Journal of Experimental Botany*, *64*(14), 4329–4341. https://doi.org/10.1093/jxb/ert238

Song, Y., He, L., Wang, X.-D., Smith, N., Wheeler, S., Garg, M. L., & Rose, R. J. (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

Song, Y., Wang, X.-D., & Rose, R. J. (2017). Oil body biogenesis and biotechnology in legume seeds. *Plant Cell Reports*, *36*(10), 1519–1532. https://doi.org/10.1007/s00299-017-2201-5

Stefanović, S., Pfeil, B. E., Palmer, J. D., & Doyle, J. J. (2009). Relationships among Phaseoloid Legumes Based on Sequences from Eight Chloroplast Regions. *Systematic Botany*, *34*(1), 115–128.

Stincone, A., Prigione, A., Cramer, T., Wamelink, M. M. C., Campbell, K., Cheung, E., Olin-Sandoval, V., Grüning, N.-M., Krüger, A., Tauqeer Alam, M., Keller, M. A., Breitenbach, M., Brindle, K. M., Rabinowitz, J. D., & Ralser, M. (2015). The return of metabolism: Biochemistry and physiology of the pentose phosphate pathway. *Biological Reviews*, *90*(3), 927–963. https://doi.org/10.1111/brv.12140

Su, H.-G., Zhang, X.-H., Wang, T.-T., Wei, W.-L., Wang, Y.-X., Chen, J., Zhou, Y.-B., Chen, M., Ma, Y.-Z., Xu, Z.-S., & Min, D.-H. (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, R., Gao, L., Mi, Z., Zheng, Y., & Li, D. (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, N., Moser, B. R., Sharma, B. K., Evangelista, R. L., Cheng, H. N., Lesch, W. C., Tangsrud, R. R., & Biswas, A. (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, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., & Paterson, A. H. (2008). Synteny and Collinearity in Plant Genomes. *Science*, *320*(5875), 486–488. https://doi.org/10.1126/science.1153917

Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., Gentzbittel, L.,

Childs, K. L., Yandell, M., Gundlach, H., Mayer, K. F., Schwartz, D. C., & Town, C. D. (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. M., Brendel, V., Whitham, S. A., & Bhattacharyya, M. K. (2008). Expression and evolution of the phosphoinositide-specific phospholipase C gene family in Arabidopsis thaliana. *Plant Physiology and Biochemistry: PPB*, *46*(7), 627–637. https://doi.org/10.1016/j.plaphy.2008.04.015

Taurino, M., Costantini, S., De Domenico, S., Stefanelli, F., Ruano, G., Delgadillo, M. O., Sánchez-Serrano, J. J., Sanmartín, M., Santino, A., & Rojo, E. (2018). SEIPIN Proteins Mediate Lipid Droplet Biogenesis to Promote Pollen Transmission and Reduce Seed Dormancy. *Plant Physiology*, *176*(2), 1531–1546. https://doi.org/10.1104/pp.17.01430

Theodoulou, F. L., & Eastmond, P. J. (2012). Seed storage oil catabolism: A story of give and take. *Current Opinion in Plant Biology*, *15*(3), 322–328. https://doi.org/10.1016/j.pbi.2012.03.017

Thiam, A. R., & Ikonen, E. (2021). Lipid Droplet Nucleation. *Trends in Cell Biology*, *31*(2), 108–118. https://doi.org/10.1016/j.tcb.2020.11.006

Thorne, R. F. (2002). How many species of seed plants are there? *TAXON*, *51*(3), 511–512. https://doi.org/10.2307/1554864

Tian, Y., Lv, X., Xie, G., Wang, L., Dai, T., Qin, X., Chen, F., & Xu, Y. (2019). FAX2 Mediates Fatty Acid Export from Plastids in Developing Arabidopsis Seeds. *Plant and Cell Physiology*, *60*(10), 2231–2242. https://doi.org/10.1093/pcp/pcz117

Tjellström, H., Yang, Z., Allen, D. K., & Ohlrogge, J. B. (2012). Rapid Kinetic Labeling of Arabidopsis Cell Suspension Cultures: Implications for Models of Lipid Export from Plastids. *Plant Physiology*, *158*(2), 601–611. https://doi.org/10.1104/pp.111.186122

Torabi, S., Sukumaran, A., Dhaubhadel, S., Johnson, S. E., LaFayette, P., Parrott, W. A., Rajcan, I., & Eskandari, M. (2021). Effects of type I Diacylglycerol O-acyltransferase (DGAT1) genes on soybean (Glycine max L.) seed composition. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-82131-5

Turquetti-Moraes, D. K., Moharana, K. C., Almeida-Silva, F., Pedrosa-Silva, F., & Venancio, T. M. (2022). Integrating omics approaches to discover and prioritize candidate genes involved in oil biosynthesis in soybean. *Gene*, *808*, 145976. https://doi.org/10.1016/j.gene.2021.145976

Tzen, JTC., Cao, Yz., Laurent, P., Ratnayake, C., & Huang, AHC. (1993). Lipids, Proteins, and Structure of Seed Oil Bodies from Diverse Species. *Plant Physiology*, *101*(1), 267–276. https://doi.org/10.1104/pp.101.1.267

Valliyodan, B., Cannon, S. B., Bayer, P. E., Shu, S., Brown, A. V., Ren, L., Jenkins, J.,

Chung, C. Y.-L., Chan, T.-F., Daum, C. G., Plott, C., Hastie, A., Baruch, K., Barry, K. W., Huang, W., Patil, G., Varshney, R. K., Hu, H., Batley, J., … Nguyen, H. T. (2019). Construction and comparison of three reference-quality genome assemblies for soybean. *The Plant Journal*, *100*(5), 1066–1082. https://doi.org/10.1111/tpj.14500

Van de Peer, Y., Maere, S., & Meyer, A. (2009). The evolutionary significance of ancient genome duplications. *Nature Reviews Genetics*, *10*(10), 725–732. https://doi.org/10.1038/nrg2600

Van de Peer, Y., Mizrachi, E., & Marchal, K. (2017). The evolutionary significance of polyploidy. *Nature Reviews Genetics*, *18*(7), Article 7. https://doi.org/10.1038/nrg.2017.26

Van Bel, M., Silvestri, F., Weitz, E. M., Kreft, L., Botzki, A., Coppens, F., & Vandepoele, K. (2022). PLAZA 5.0: Extending the scope and power of comparative and functional genomics in plants. *Nucleic Acids Research*, *50*(D1), D1468–D1474. https://doi.org/10.1093/nar/gkab1024

Vanneste, K., Van de Peer, Y., & Maere, S. (2013). Inference of Genome Duplications from Age Distributions Revisited. *Molecular Biology and Evolution*, *30*(1), 177–190. https://doi.org/10.1093/molbev/mss214

Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., Cannon, S., Baek, J., Rosen, B. D., Tar'an, B., Millan, T., Zhang, X., Ramsay, L. D., Iwata, A., Wang, Y., Nelson, W., Farmer, A. D., Gaur, P. M., Soderlund, C., … Cook, D. R. (2013). Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement. *Nature Biotechnology*, *31*(3), 240–246. https://doi.org/10.1038/nbt.2491

Vision, T. (2010). The Dryad Digital Repository: Published evolutionary data as part of the greater data ecosystem. *Nature Precedings*, 1–1. https://doi.org/10.1038/npre.2010.4595.1

Voelker, T., & Kinney, A. J. (2001). VARIATIONS IN THE BIOSYNTHESIS OF SEED-STORAGE LIPIDS. *Annual Review of Plant Biology*, *52*(Volume 52, 2001), 335–361. https://doi.org/10.1146/annurev.arplant.52.1.335

Wakelam, M. J. O. (1998). Diacylglycerol – when is it an intracellular messenger? *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, *1436*(1), 117–126. https://doi.org/10.1016/S0005-2760(98)00123-4

Wang, H.-W., Zhang, B., Hao, Y.-J., Huang, J., Tian, A.-G., Liao, Y., Zhang, J.-S., & Chen, S.-Y. (2007). The soybean Dof-type transcription factor genes, GmDof4 and GmDof11, enhance lipid content in the seeds of transgenic Arabidopsis plants. *The Plant Journal*, *52*(4), 716–729. https://doi.org/10.1111/j.1365-313X.2007.03268.x

Wang, J., Kuang, H., Zhang, Z., Yang, Y., Yan, L., Zhang, M., Song, S., & Guan, Y. (2020).

Generation of seed lipoxygenase-free soybean using CRISPR-Cas9. *The Crop Journal*, *8*(3), 432–439. https://doi.org/10.1016/j.cj.2019.08.008

Wang, L., Jia, G., Jiang, X., Cao, S., Chen, Z. J., & Song, Q. (2021). Altered chromatin architecture and gene expression during polyploidization and domestication of soybean. *The Plant Cell*, *33*(5), 1430–1446. https://doi.org/10.1093/plcell/koab081

Wang, S., Liu, S., Wang, J., Yokosho, K., Zhou, B., Yu, Y.-C., Liu, Z., Frommer, W. B., Ma, J. F., Chen, L.-Q., Guan, Y., Shou, H., & Tian, Z. (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), 1776–1786. https://doi.org/10.1093/nsr/nwaa110

Wang, X., Li, Q., Zhang, Q., Huang, S., Yu, J., Qin, H., Qi, H., Li, Y., Li, Y., Yin, Z., Han, X., Wu, X., Xin, D., Chen, Q., & Qi, Z. (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

Wilson, R. F. (2004). Seed Composition. In *Soybeans: Improvement, Production, and Uses* (pp. 621–677). John Wiley & Sons, Ltd. https://doi.org/10.2134/agronmonogr16.3ed.c13

Wobus, U., & Weber, H. (1999). Seed maturation: Genetic programmes and control signals. *Current Opinion in Plant Biology*, *2*(1), 33–38. https://doi.org/10.1016/S1369-5266(99)80007-7

Xu, Y., Caldo, K. M. P., Pal-Nath, D., Ozga, J., Lemieux, M. J., Weselake, R. J., & Chen, G. (2018a). Properties and Biotechnological Applications of Acyl-CoA:diacylglycerol Acyltransferase and Phospholipid:diacylglycerol Acyltransferase from Terrestrial Plants and Microalgae. *Lipids*, *53*(7), 663–688. https://doi.org/10.1002/lipd.12081

Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E., Lancet, D., & Shmueli, O. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*, *21*(5), 650–659. https://doi.org/10.1093/bioinformatics/bti042

Yang, H., Wang, W., He, Q., Xiang, S., Tian, D., Zhao, T., & Gai, J. (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, K., Tian, Z., Chen, C., Luo, L., Zhao, B., Wang, Z., Yu, L., Li, Y., Sun, Y., Li, W., Chen, Y., Li, Y., Zhang, Y., Ai, D., Zhao, J., Shang, C., Ma, Y., Wu, B., Wang, M., … Wan, P. (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), 13213–13218. https://doi.org/10.1073/pnas.1420949112

Yang, S., Miao, L., He, J., Zhang, K., Li, Y., & Gai, J. (2019). Dynamic Transcriptome Changes Related to Oil Accumulation in Developing Soybean Seeds. *International Journal of Molecular Sciences*, *20*(9), 2202. https://doi.org/10.3390/ijms20092202

Yang, W., Wang, G., Li, J., Bates, P. D., Wang, X., & Allen, D. K. (2017). Phospholipase Dζ Enhances Diacylglycerol Flux into Triacylglycerol. *Plant Physiology*, *174*(1), 110–123. https://doi.org/10.1104/pp.17.00026

Yant, L., & Bomblies, K. (2017). Genomic studies of adaptive evolution in outcrossing Arabidopsis species. *Current Opinion in Plant Biology*, *36*, 9–14.

Yao, Y., You, Q., Duan, G., Ren, J., Chu, S., Zhao, J., Li, X., Zhou, X., & Jiao, Y. (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, W.-C., Lee, F.-C., Shabari Shan, D. K., Musa, H., Appleton, D. R., & Kulaveerasingam, H. (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

Yeom, W. W., Kim, H. J., Lee, K.-R., Cho, H. S., Kim, J.-Y., Jung, H. W., Oh, S.-W., Jun, S. E., Kim, H. U., & Chung, Y.-S. (2020). Increased Production of α-Linolenic Acid in Soybean Seeds by Overexpression of Lesquerella FAD3-1. *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/fpls.2019.01812

Zaaboul, F., Zhao, Q., Xu, Y., & Liu, Y. (2022). Soybean oil bodies: A review on composition, properties, food applications, and future research aspects. *Food Hydrocolloids*, *124*, 107296. https://doi.org/10.1016/j.foodhyd.2021.107296

Zeng, X., Liu, H., Du, H., Wang, S., Yang, W., Chi, Y., Wang, J., Huang, F., & Yu, D. (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, B., & Horvath, S. (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, D., Sun, L., Li, S., Wang, W., Ding, Y., Swarm, S. A., Li, L., Wang, X., Tang, X., Zhang, Z., Tian, Z., Brown, P. J., Cai, C., Nelson, R. L., & Ma, J. (2018). Elevation of soybean seed oil content through selection for seed coat shininess. *Nature Plants*, *4*(1), 30–35. https://doi.org/10.1038/s41477-017-0084-7

Zhang, M., Liu, S., Wang, Z., Yuan, Y., Zhang, Z., Liang, Q., Yang, X., Duan, Z., Liu, Y., Kong, F., Liu, B., Ren, B., & Tian, Z. (2022). Progress in soybean functional genomics

over the past decade. *Plant Biotechnology Journal*, *20*(2), 256–282. https://doi.org/10.1111/pbi.13682

Zhang, Y.-Q., Lu, X., Zhao, F.-Y., Li, Q.-T., Niu, S.-L., Wei, W., Zhang, W.-K., Ma, B., Chen, S.-Y., & Zhang, J.-S. (2016). Soybean GmDREBL Increases Lipid Content in Seeds of Transgenic Arabidopsis. *Scientific Reports*, *6*(1), 34307. https://doi.org/10.1038/srep34307

Zhao, J., Bi, R., Li, S., Zhou, D., Bai, Y., Jing, G., Zhang, K., & Zhang, W. (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*, 153019. https://doi.org/10.1016/j.jplph.2019.153019

Zhao, S., Yan, F., Liu, Y., Sun, M., Wang, Y., Li, J., Zhang, X., Yang, X., & Wang, Q. (2023). Genome-wide identification and expression analysis of diacylglycerol acyltransferase genes in soybean (Glycine max). *PeerJ*, *11*, e14941. https://doi.org/10.7717/peerj.14941

Zhao, Y., Dong, Q., Geng, Y., Ma, C., & Shao, Q. (2023). Dynamic Regulation of Lipid Droplet Biogenesis in Plant Cells and Proteins Involved in the Process. *International Journal of Molecular Sciences*, *24*(8), Article 8. https://doi.org/10.3390/ijms24087476

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., Yu, Y., Shu, L., Zhao, Y., Ma, Y., Fang, C., Shen, Y., Liu, T., Li, C., Li, Q., Wu, M., Wang, M., Wu, Y., Dong, Y., … Tian, Z. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology*, *33*(4), 408–414. https://doi.org/10.1038/nbt.3096

Zhou, Z., Lakhssassi, N., Cullen, M. A., El Baz, A., Vuong, T. D., Nguyen, H. T., & Meksem, K. (2019). Assessment of Phenotypic Variations and Correlation among Seed Composition Traits in Mutagenized Soybean Populations. *Genes*, *10*(12), Article 12. https://doi.org/10.3390/genes10120975

Zhou, Z., Lakhssassi, N., Knizia, D., Cullen, M. A., El Baz, A., Embaby, M. G., Liu, S., Badad, O., Vuong, T. D., AbuGhazaleh, A., Nguyen, H. T., & Meksem, K. (2021). Genome-wide identification and analysis of soybean acyl-ACP thioesterase gene family reveals the role of GmFAT to improve fatty acid composition in soybean seed. *Theoretical and Applied Genetics*, *134*(11), 3611–3623. https://doi.org/10.1007/s00122-021-03917-9

Zhuang, Y., Wang, X., Li, X., Hu, J., Fan, L., Landis, J. B., Cannon, S. B., Grimwood, J., Schmutz, J., Jackson, S. A., Doyle, J. J., Zhang, X. S., Zhang, D., & Ma, J. (2022). Phylogenomics of the genus Glycine sheds light on polyploid evolution and life-strategy transition. *Nature Plants*, *8*(3), 233–244.

https://doi.org/10.1038/s41477-022-01102-4

Zia-Ul-Haq, M., Ahmad, M., & Iqbal, S. (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), 851–856. https://doi.org/10.1007/s11746-008-1269-z

Zia-Ul-Haq, M., Ahmad, M., Iqbal, S., Ahmad, S., & Ali, H. (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), 1143–1148. https://doi.org/10.1007/s11746-007-1136-3

6 - APÊNDICE

6.1 Apêndice Capítulo 1

Apêndice 1. Lista dos principais genes codificadores de fatores de transcrição e outros genes associados ao acúmulo de óleo em sementes de soja.

Gene	Gene_ID	Genoma	Publicação
GmWRI1a	Glyma.15G221600	Wm82.v2	(L. Chen et al., 2018; Guo et al., 2020)
GmWRI1b	Glyma.08G227700	Wm82.v2	
GmWRI1c	Glyma.15G34770	Wm82.v1	
GmWRI1	Glyma.02G33090	Wm82.v1	
GmWRI1	Glyma.07G02380	Wm82.v1	(Manan, Chen, et al., 2017)
GmLEC1-1	Glyma.07G268100	Wm82.v2	(Pelletier et al., 2017)
GmLEC1	Glyma.07G39820	Wm82.v1	
ybeaGmLEC2a	Glyma.20G035800	Wm82.v2	
GmLEC2b	Glyma.20G035700	Wm82.v2	(Manan, Ahmad, et al., 2017)
GmFUS3-like	Glyma.19G27336	Wm82.v1	
GmABI3-like	Glyma.08G47240	Wm82.v1	
GmZIP123	Glyma.06G010200	Wm82.v2	(Q.-X. Song et al., 2013)
GmNFYA	Glyma.02G303800	Wm82.v2	(X. Lu et al., 2016)
GmZF351	Glyma.06G290100	Wm82.v2	(Q.-T. Li et al., 2017)
GmZF392	Glyma.12G205700	Wm82.v2	(L. Lu et al., 2021)
GmB1	Glyma.13G241700	Wm82.v2	(D. Zhang et al., 2018)
GmaDof4	Glyma.17G081800	Wm82.v2	(H.-W. Wang et al., 2007)
GmDof11	Glyma.13G329000	Wm82.v2	
GmMYB73	Glyma.06G303100	Wm82.v2	(Y.-F. Liu et al., 2014)

GmaDREBL	Glyma.12G103100	Wm82.v2	(Y.-Q. Zhang et al., 2016)
GmST1	Glyma.08G109100	Wm82.v2	(J. Li et al., 2022)
GmSWEET10a	Glyma.15G049200	Wm82.v2	(Miao et al., 2020; S. Wang et al., 2020)
GmSWEET10b	Glyma.08G183500	Wm82.v2	(S. Wang et al., 2020)
GmST05	Glyma.05G244100	Wm82.v2	(Duan et al., 2022)

Apêndice 2. Lista dos genes codificadores de enzimas e subunidades envolvidas na síntese de ácidos graxos em sementes de soja.

Gene	Gene_ID	Descrição
KAS I	Glyma.05G129600	*3-cetoacyl-ACP synthase I*
KAS I	Glyma.05G218600	*3-cetoacyl-ACP synthase I*
KAS I	Glyma.08G024700	*3-cetoacyl-ACP synthase I*
KAS I	Glyma.08G084300	*3-cetoacyl-ACP synthase I*
KAS II	Glyma.13G112700	*3-cetoacyl-ACP synthase II*
KAS II	Glyma.15G181500	*3-cetoacyl-ACP synthase II*
KAS II	Glyma.17G047000	*3-cetoacyl-ACP synthase II*
KAS III	Glyma.09G277400	*3-cetoacyl-ACP synthase III*
KAS III	Glyma.15G003100	*3-cetoacyl-ACP synthase III*
KAS III	Glyma.18G211400	*3-cetoacyl-ACP synthase III*
FAE	Glyma.06G058500	*fatty acid elongase*
FAE	Glyma.02G001500	*fatty acid elongase*
FAE	Glyma.03G260300	*fatty acid elongase*
FAE	Glyma.04G057800	*fatty acid elongase*
FAE	Glyma.04G149300	*fatty acid elongase*
FAE	Glyma.05G011100	*fatty acid elongase*
FAE	Glyma.05G083000	*fatty acid elongase*
FAE	Glyma.06G012500	*fatty acid elongase*
FAE	Glyma.06G214800	*fatty acid elongase*
FAE	Glyma.08G261100	*fatty acid elongase*
FAE	Glyma.10G001800	*fatty acid elongase*
FAE	Glyma.10G179400	*fatty acid elongase*
FAE	Glyma.10G241700	*fatty acid elongase*

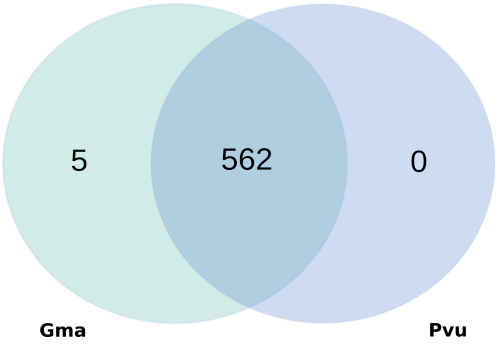
FAE	Glyma.10G274400	*fatty acid elongase*
FAE	Glyma.10G291700	*fatty acid elongase*
FAE	Glyma.11G144809	*fatty acid elongase*
FAE	Glyma.11G245600	*fatty acid elongase*
FAE	Glyma.11G245700	*fatty acid elongase*
FAE	Glyma.12G075100	*fatty acid elongase*
FAE	Glyma.13G238600	*fatty acid elongase*
FAE	Glyma.13G331600	*fatty acid elongase*
FAE	Glyma.14G074300	*fatty acid elongase*
FAE	Glyma.15G042500	*fatty acid elongase*
FAE	Glyma.15G046300	*fatty acid elongase*
FAE	Glyma.15G074700	*fatty acid elongase*
FAE	Glyma.15G149400	*fatty acid elongase*
FAE	Glyma.17G118700	*fatty acid elongase*
FAE	Glyma.17G183700	*fatty acid elongase*
FAE	Glyma.17G251000	*fatty acid elongase*
FAE	Glyma.18G011500	*fatty acid elongase*
FAE	Glyma.18G011600	*fatty acid elongase*
FAE	Glyma.20G115500	*fatty acid elongase*
FAE	Glyma.20G152500	*fatty acid elongase*
FAE	Glyma.20G210900	*fatty acid elongase*
FAE	Glyma.20G240900	*fatty acid elongase*
ACP	Glyma.03G242600	*acyl carrier protein*
ACP	Glyma.05G201300	*acyl carrier protein*
ACP	Glyma.07G042900	*acyl carrier protein*
ACP	Glyma.08G008800	*acyl carrier protein*
ACP	Glyma.10G158500	*acyl carrier protein*
ACP	Glyma.13G214600	*acyl carrier protein*
ACP	Glyma.15G098500	*acyl carrier protein*
ACP	Glyma.19G240100	*acyl carrier protein*
HD	Glyma.05G118800	*beta-hidroxiacyl-ACP dehidratase*
HD	Glyma.05G118900	*beta-hidroxiacyl-ACP dehidratase*
HD	Glyma.08G073900	*beta-hidroxiacyl-ACP dehidratase*
HD	Glyma.08G179900	*beta-hidroxiacyl-ACP dehidratase*
HD	Glyma.15G052500	*beta-hidroxiacyl-ACP dehidratase*

SAD	Glyma.02G138100	*stearoyl-ACP desaturase*
SAD	Glyma.07G207200	*stearoyl-ACP desaturase*
SAD	Glyma.13G038600	*stearoyl-ACP desaturase*
SAD	Glyma.14G121400	*stearoyl-ACP desaturase*
FATA	Glyma.08G349200	*palmitoyl-ACP thioesterase*
FATA	Glyma.18G167300	*palmitoyl-ACP thioesterase*
FATB	Glyma.04G151600	*palmitoyl-ACP thioesterase*
FATB	Glyma.04G197400	*palmitoyl-ACP thioesterase*
FATB	Glyma.05G012300	*palmitoyl-ACP thioesterase*
FATB	Glyma.06G168100	*palmitoyl-ACP thioesterase*
FATB	Glyma.06G211300	*palmitoyl-ACP thioesterase*
FATB	Glyma.17G120400	*palmitoyl-ACP thioesterase*
LACS9	Glyma.06G112900	*long-chain acyl-CoA synthetase*
LACS9	Glyma.13G010100	*long-chain acyl-CoA synthetase*
LACS9	Glyma.13G079900	*long-chain acyl-CoA synthetase*
LACS9	Glyma.14G149700	*long-chain acyl-CoA synthetase*
LACS9	Glyma.20G060100	*long-chain acyl-CoA synthetase*
LACS9	Glyma.20G060300	*long-chain acyl-CoA synthetase*
FAX	Glyma.03G133200	*fatty acid export protein*
FAX	Glyma.07G108500	*fatty acid export protein*
FAX	Glyma.09G274500	*fatty acid export protein*
FAX	Glyma.14G024900	*fatty acid export protein*
FAX	Glyma.18G213500	*fatty acid export protein*
FAX	Glyma.19G135000	*fatty acid export protein*
BC	Glyma.05G221100	*biotin carboxylase*
BC	Glyma.08G027600	*biotin carboxylase*
BCCP	Glyma.09G248900	*biotin carboxyl carrier*
BCCP	Glyma.13G057400	*biotin carboxyl carrier*
BCCP	Glyma.18G243500	*biotin carboxyl carrier*
BCCP	Glyma.18G265300	*biotin carboxyl carrier*
BCCP	Glyma.19G028800	*biotin carboxyl carrier*
CT	Glyma.18G195700	*carboxyltransferase*
CT	Glyma.18G195900	*carboxyltransferase*
CT	Glyma.18G196000	*carboxyltransferase*

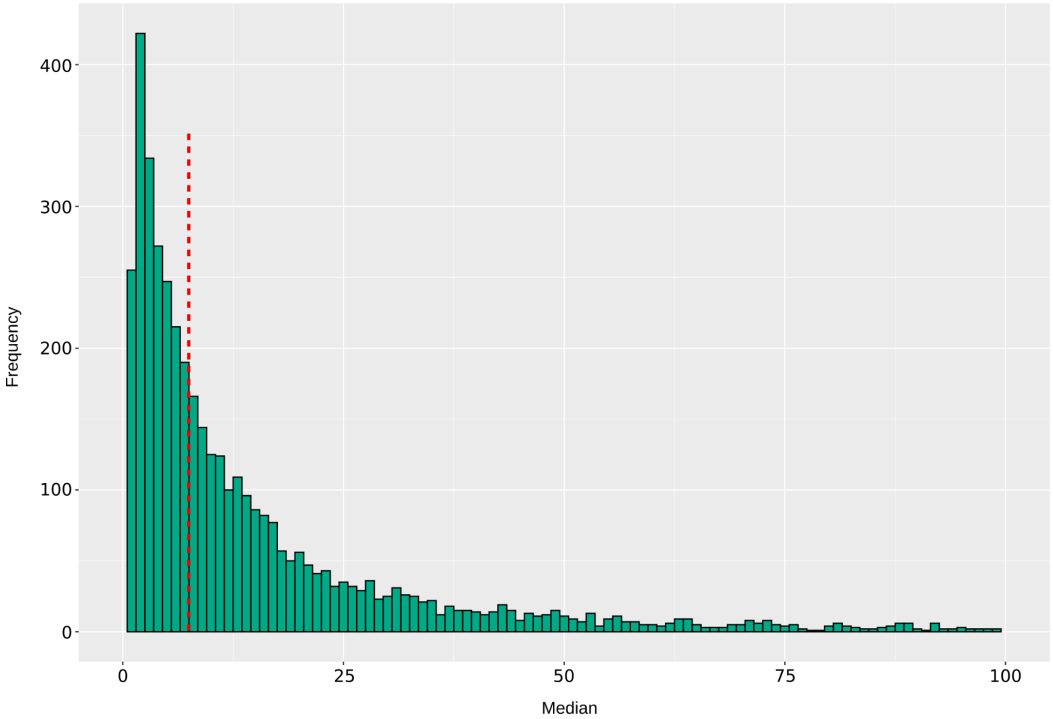
FAD	Glyma.01G071800	*fatty acid desaturase*
FAD	Glyma.01G120400	*fatty acid desaturase*
FAD	Glyma.02G203300	*fatty acid desaturase*
FAD	Glyma.02G227200	*fatty acid desaturase*
FAD	Glyma.03G056700	*fatty acid desaturase*
FAD	Glyma.03G144500	*fatty acid desaturase*
FAD	Glyma.07G151300	*fatty acid desaturase*
FAD	Glyma.09G111900	*fatty acid desaturase*
FAD	Glyma.10G278000	*fatty acid desaturase*
FAD	Glyma.11G174100	*fatty acid desaturase*
FAD	Glyma.14G194300	*fatty acid desaturase*
FAD	Glyma.15G195201	*fatty acid desaturase*
FAD	Glyma.17G074400	*fatty acid desaturase*
FAD	Glyma.18G062000	*fatty acid desaturase*
FAD	Glyma.18G202600	*fatty acid desaturase*
FAD	Glyma.19G147300	*fatty acid desaturase*
FAD	Glyma.19G147400	*fatty acid desaturase*
FAD	Glyma.20G111000	*fatty acid desaturase*
ER	Glyma.08G345900	*enoyl-ACP reductase*
ER	Glyma.11G101400	*enoyl-ACP reductase*
ER	Glyma.12G027300	*enoyl-ACP reductase*
ER	Glyma.18G156100	*enoyl-ACP reductase*
ABCA	Glyma.04G173000	*fatty acid transporter*
ABCA	Glyma.06G191300	*fatty acid transporter*

Fonte: Mapman - disponível em: https://bioinformatics.psb.ugent.be/plaza

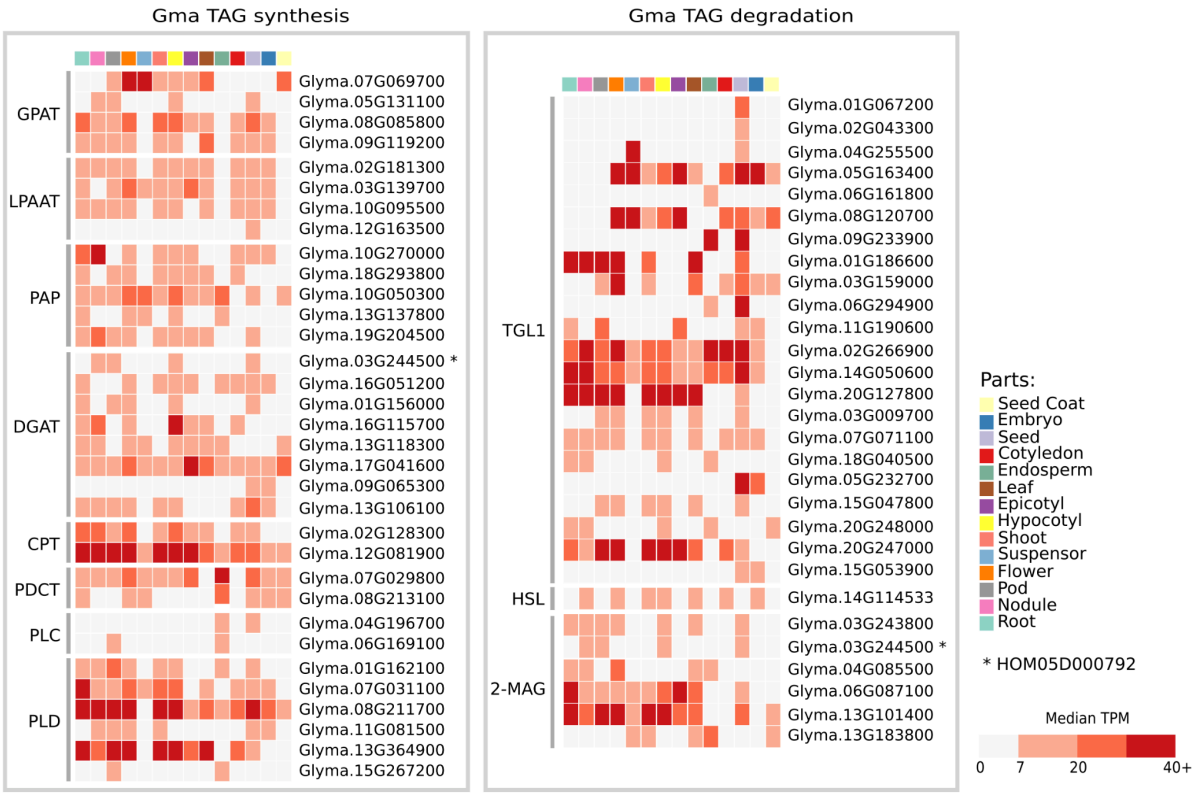
6.2 Apêndice Capítulo 2



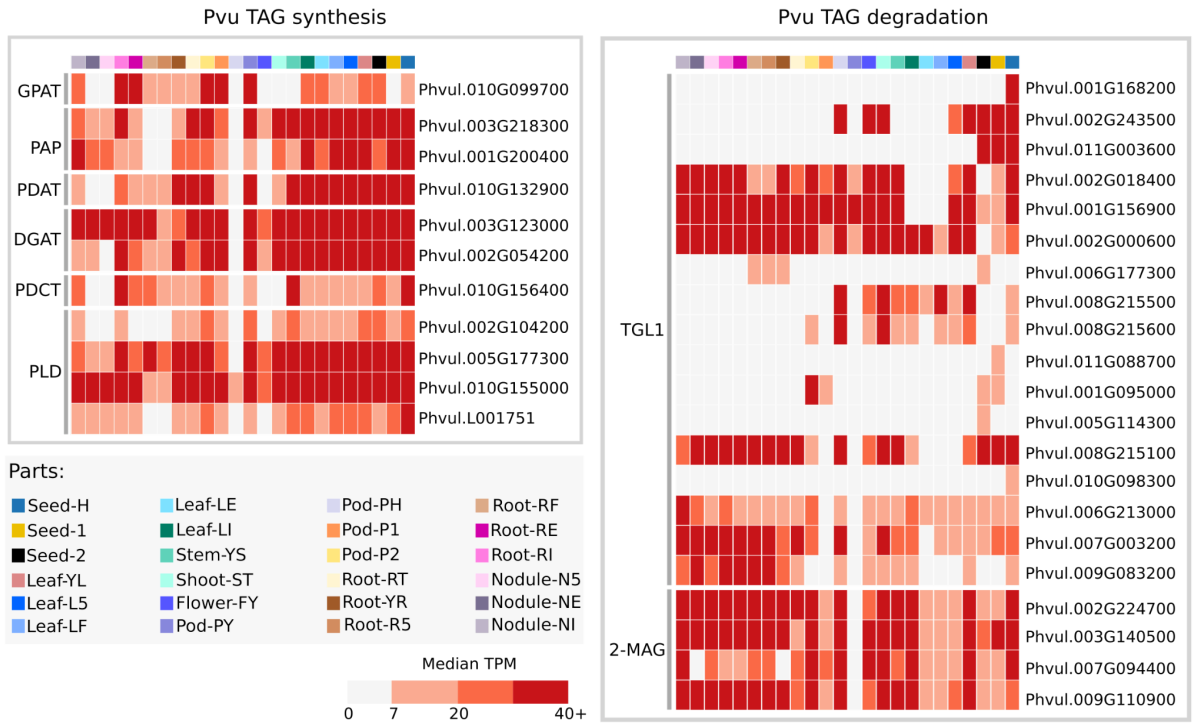
Supplementary Figure 1. Candidate oil homologous families for soybean (Gma) and common bean (Pvu). In total, we found 567 families of which 562 contain at least one common bean homolog, except HOM05D015518 (Glyma.13G001800; Glyma.20G068000), HOM05D006604 (Glyma.01G103450; Glyma.16G133700), HOM05D031525 (Glyma.07G184950; Glyma.08G064400), HOM05D130031 (Glyma.02G006100) and HOM05D039847 (Glyma.02G281400; Glyma.02G281500; Glyma.14G033100). Venn diagram was generated using Venn diagram (https://bioinformatics.psb.ugent.be/webtools/Venn/).



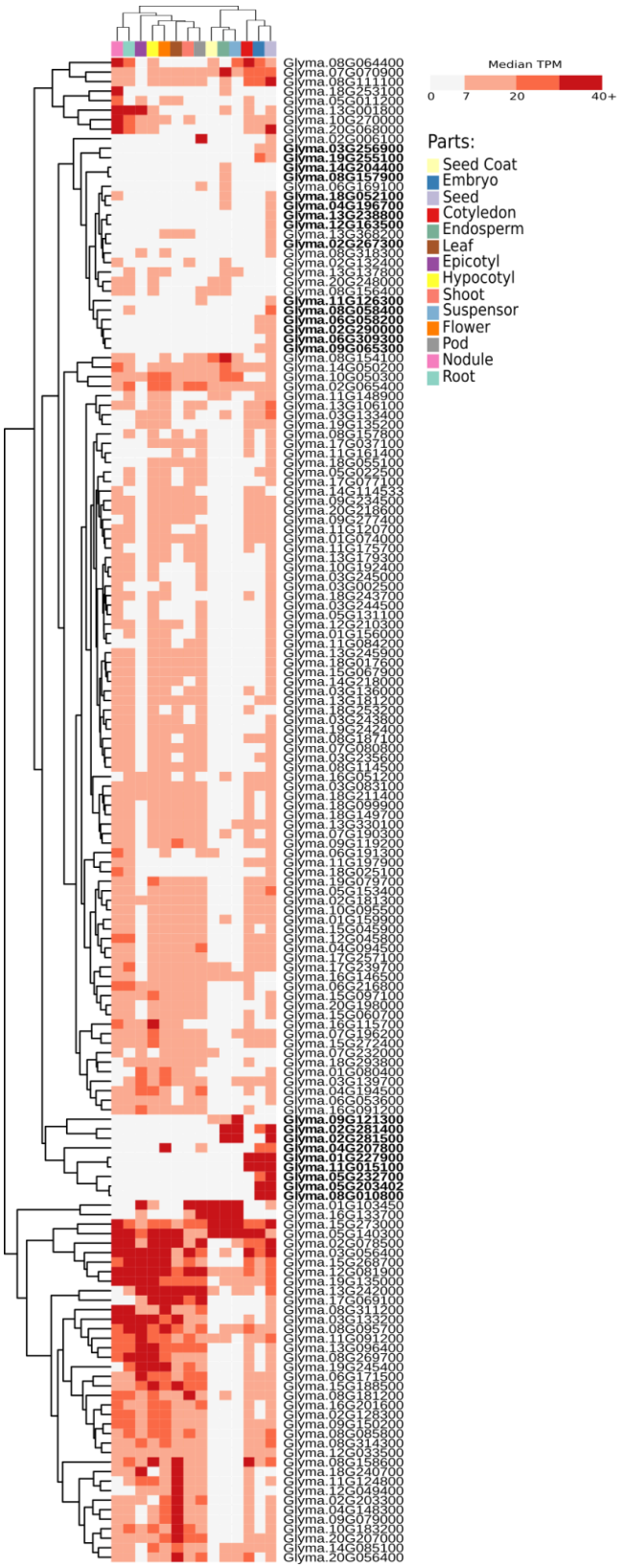
Supplementary Figure 2. Frequency distribution of maximum median TPM values for genes expressed in soybean seeds and seed subregions. A dashed red line indicates the threshold of 8 TPM.



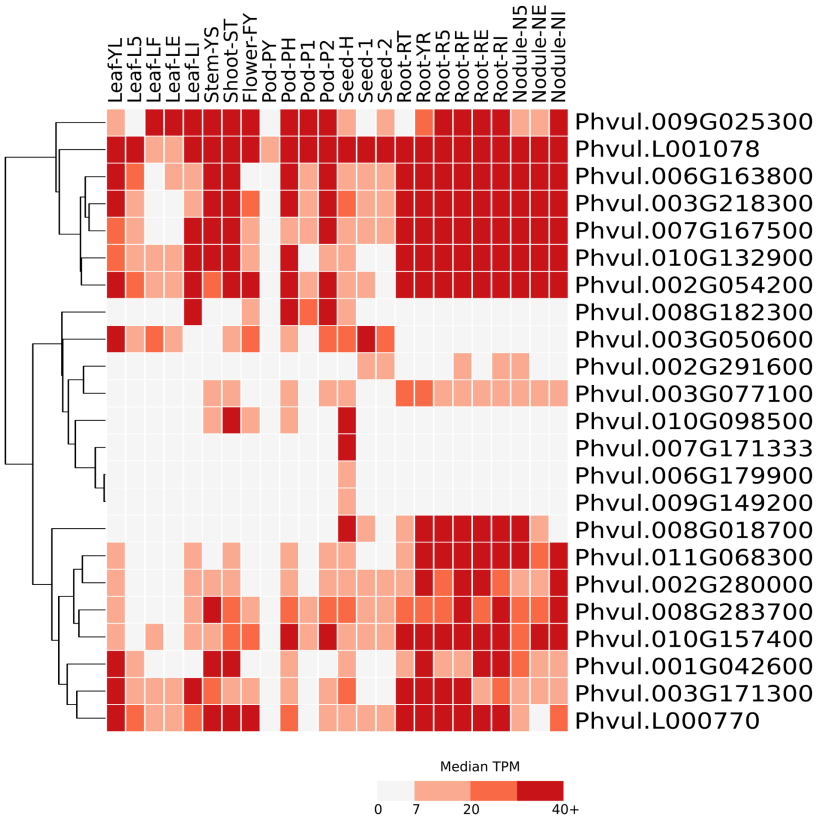
Supplementary Figure 3. Expression of soybean TAG pathway genes with at least 8 TPM in seeds or seed parts. **TPM:** transcripts per million. **GPAT**: acyl-CoA glycerol-3-phosphate acyltransferase; **LPAAT**: acyl-CoA lysophosphatidic acid acyltransferase; **PAP**: phosphatidic acid phosphatase; **DGAT**: acyl-CoA diacylglycerol acyltransferase; **PLC**: phospholipase C; **PLD**: phospholipase D; **CPT**: cytidine diphosphate-choline diacylglycerol cholinephosphotransferase; **PDCT**: phosphatidylcholine diacylglycerol cholinephosphotransferase; **TGL1**: triacylglycerol lipase; **HSL**: hormone-sensitive lipase; **2-MAG**: 2-monoacylglycerol acylhydrolase.



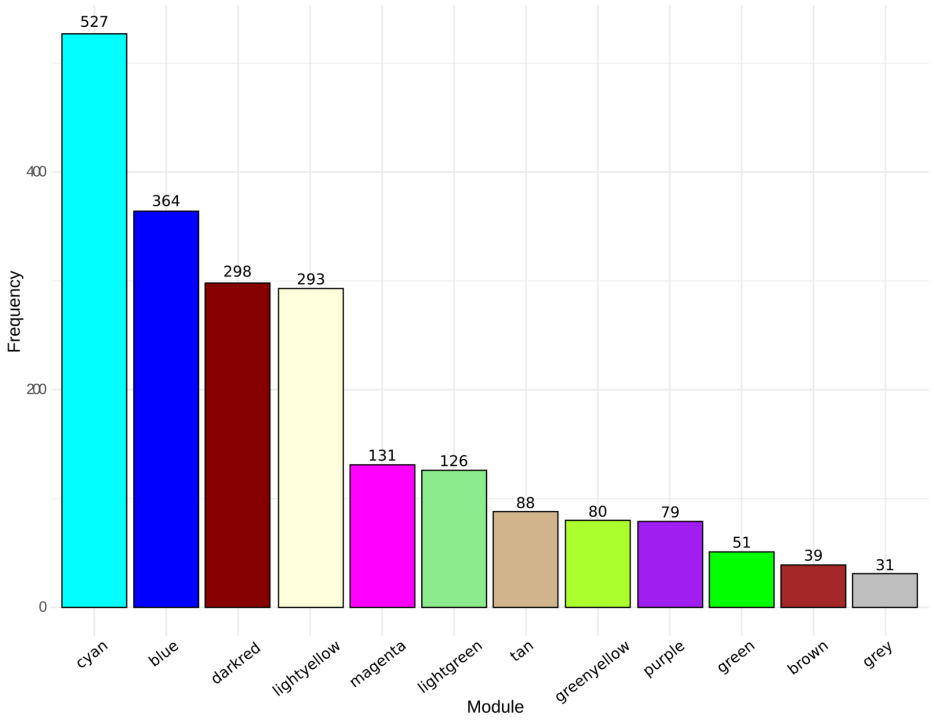
Supplementary Figure 4. Expression of common bean TAG pathway genes (at least 8 TPM in seeds). **TPM**: transcripts per million. **Seed-H**: seed heart stage; **Seed-1**: seeds with ~50mg; **Seed-2**: seeds with ~150mg; **Leaf-YL**: fully expanded second trifoliate leaf; **Leaf-L5**: leaf tissue collected 5 days after plants were inoculated with effective rhizobium; **Leaf-LF**: leaf tissue from fertilized plants; **Leaf-LE**: leaf tissue collected 21 days after plants were inoculated with effective rhizobium; **Leaf-LI**: leaf tissue collected 21 days after plants were inoculated with ineffective rhizobium; **Stem-YS**: all stem internodes above the cotyledon collected at the second trifoliate stage; **Shoot-ST**: shoot tip; **Flower-FY**: young flowers; **Pod-PY**: young pods; **Pod-PH**: pods 9cm long; **Pod-P1**: pods between 10 and 11 cm long; **Pod-P2**: pods between 12 and 13 cm long; **Root-RT**: root tips; **Root-YR:** whole roots at the second trifoliate stage of development; **Root-RS**: whole roots separated from 5 days old pre-fixing nodules; **Root-RF**: whole roots from fertilized plants; **Root-RE:** whole roots separated from fix+ nodules collected 21 days after inoculation; **Root-RI**: whole roots separated from fix- nodules collected 21 days after inoculation; **Nodule-NS**: pre-fixing effective nodules collected 5 days after inoculation; **Nodule-NE**: effectively fixing nodules collected 21 days after inoculation; **Nodule-NI**: ineffectively fixing nodules collected 21 days after inoculation.



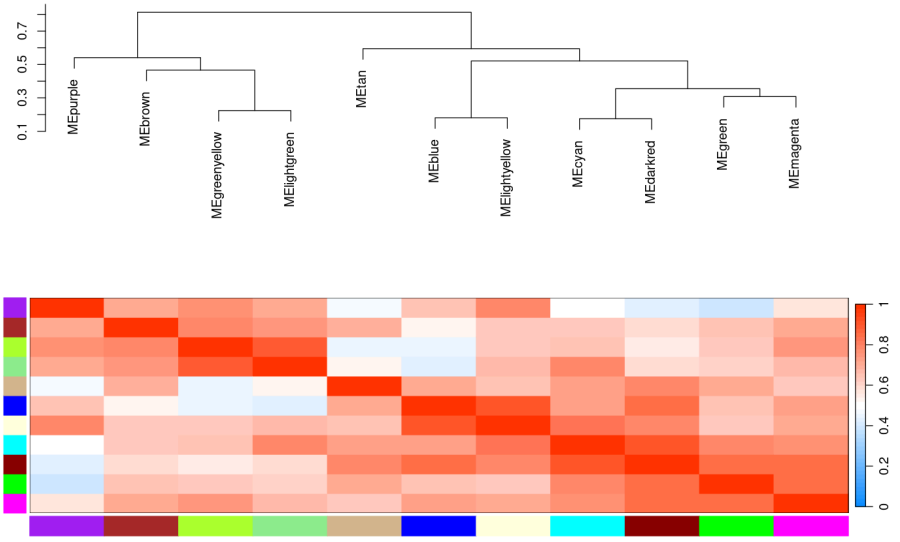
Supplementary Figure 5. Genes from 90 homologous families with expression in soybean but not in common bean seeds. Genes more expressed in seed or seed subregions than in other parts are marked in bold.



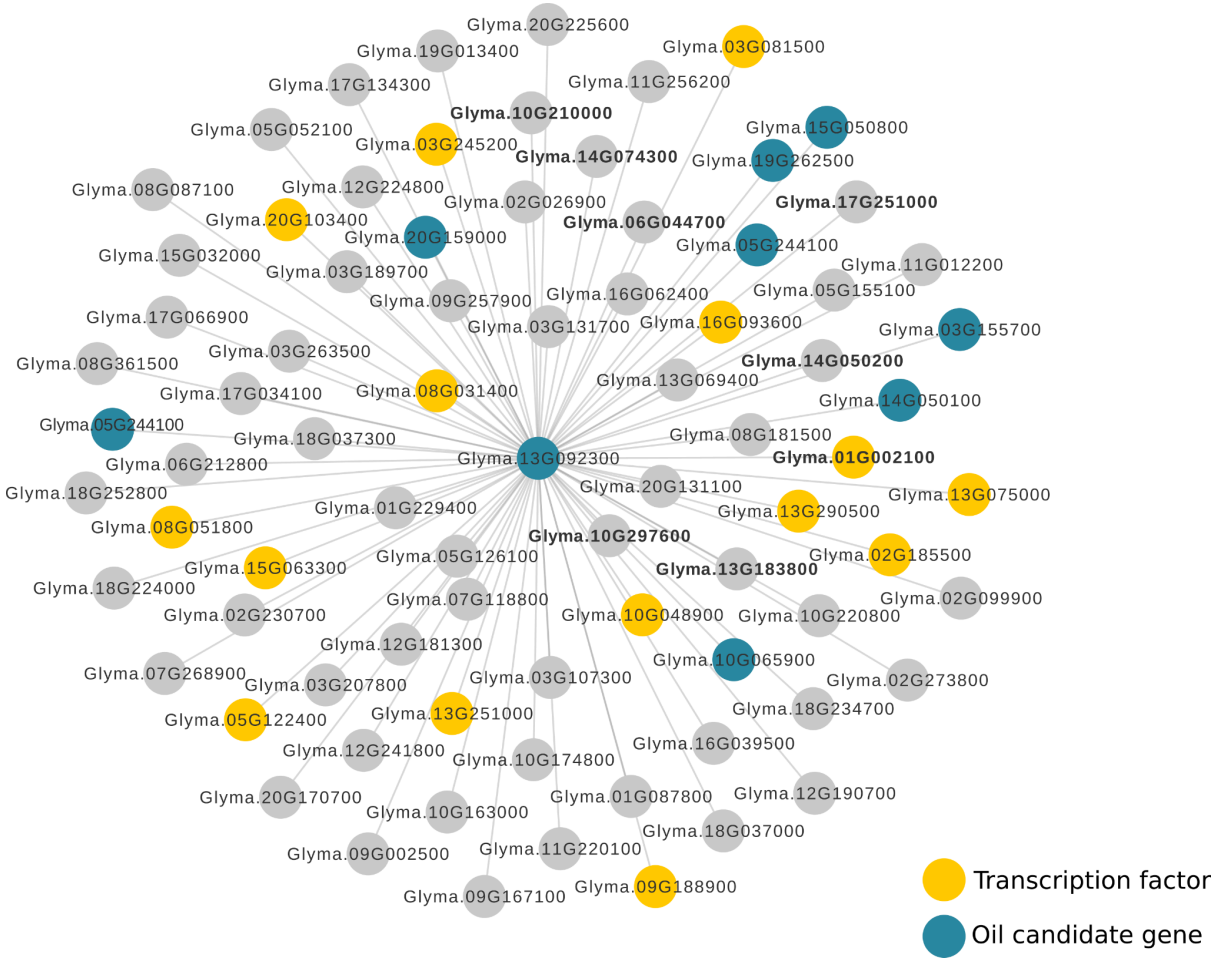
Supplementary Figure 6. Expression of common bean genes from the third category of genes with high expression in seeds. **TPM**: transcripts per million. **Seed-H**: seed heart stage; **Seed-1**: seeds with ~50mg; **Seed-2**: seeds with ~150mg; **Leaf-YL**: fully expanded second trifoliate leaf; **Leaf-L5**: leaf tissue collected 5 days after plants were inoculated with effective rhizobium; **Leaf-LF**: leaf tissue from fertilized plants; **Leaf-LE**: leaf tissue collected 21 days after plants were inoculated with effective rhizobium; **Leaf-LI**: leaf tissue collected 21 days after plants were inoculated with ineffective rhizobium; **Stem-YS**: all stem internodes above the cotyledon collected at the second trifoliate stage; **Shoot-ST**: shoot tip; **Flower-FY**: young flowers; **Pod-PY**: young pods; **Pod-PH**: pods 9cm long; **Pod-P1**: pods between 10 and 11 cm long; **Pod-P2**: pods between 12 and 13 cm long; **Root-RT**: root tips; **Root-YR:** whole roots at the second trifoliate stage of development; **Root-RS**: whole roots separated from 5 days old pre-fixing nodules; **Root-RF**: whole roots from fertilized plants; **Root-RE:** whole roots separated from fix+ nodules collected 21 days after inoculation; **Root-RI**: whole roots separated from fix- nodules collected 21 days after inoculation; **Nodule-NS**: pre-fixing effective nodules collected 5 days after inoculation; **Nodule-NE**: effectively fixing nodules collected 21 days after inoculation; **Nodule-NI**: ineffectively fixing nodules collected 21 days after inoculation.



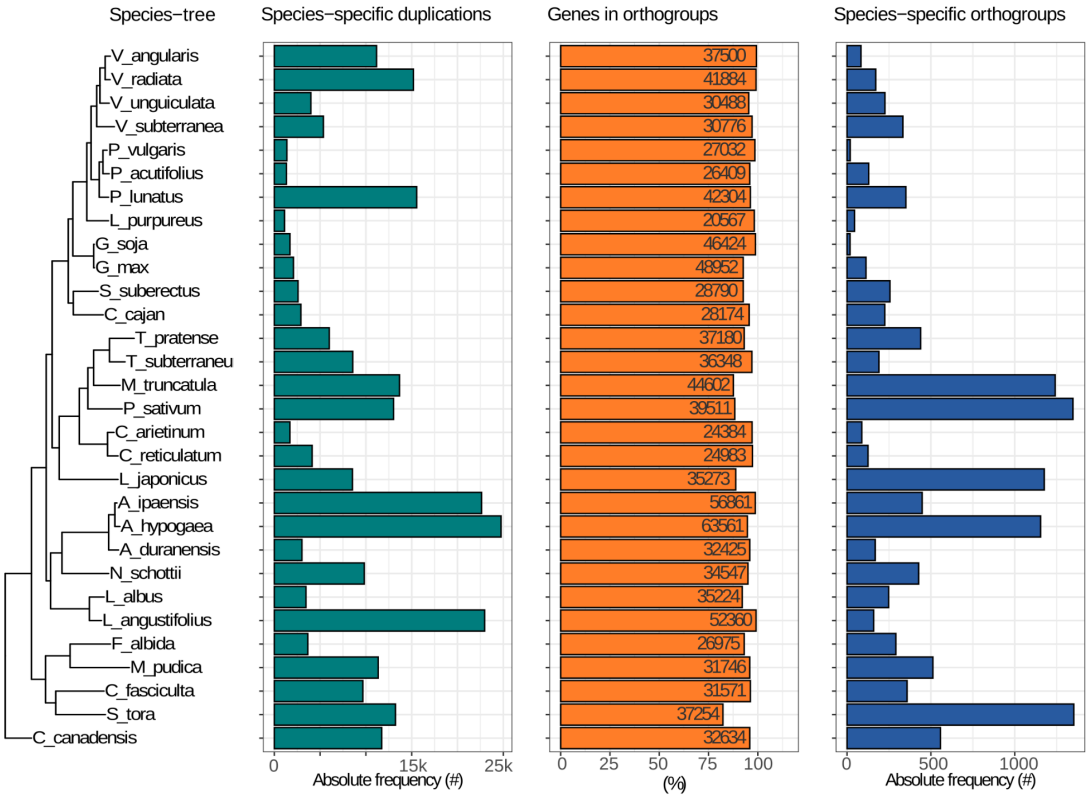
Supplementary Figure 7. Frequency of soybean oil genes in twelve co-expression modules.



Supplementary Figure 8. Modules eigengene indicating the proximity (clusters) of the expression profile among eleven modules. Module eigengene is the first principal component of a principal component analysis, which summarizes the expression of the module.



Supplementary Figure 9. Connections of the hub Glyma.13G092300. Transcription factors and previously reported oil candidate genes (Niu et al., 2020) are highlighted in yellow and green, respectively. The connections of these three hubs with relevant genes to lipid metabolism and TFs support their importance in oil-related traits. Genes reported as associated with lipid metabolism in Aralip or Mapman are marked in bold. Network was generated using Cytoscape (Shannon et al., 2003, available at https://doi.org/10.1101/gr.1239303).



Supplementary Figure 10. Summary of the orthogroup analysis. From left to right, the panels contain the species-tree, absolute frequency of the species-specific duplications, percentage of the genes in orthogroups and absolute frequency of the species-specific orthogroups. The image was generated using Cogeqc R/Bioconductor package (Almeida-Silva and Van de Peer, 2023).