UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE BIOCIÊNCIAS

DEPARTAMENTO DE GENÉTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

História evolutiva do fator de transcrição NF-Y (*Nuclear Factor of Y Box*) em plantas e seu papel na regulação do desenvolvimento e acúmulo de lipídeos em oleaginosas

Alexandro Cagliari

Orientadora: Dra. Márcia Maria Pinheiro Margis

Co-orientador: Dr. Rogério Margis

Porto Alegre, agosto de 2013.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE BIOCIÊNCIAS

DEPARTAMENTO DE GENÉTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

História evolutiva do fator de transcrição NF-Y (*Nuclear Factor of Y Box*) em plantas e seu papel na regulação do desenvolvimento e acúmulo de lipídeos em oleaginosas

Alexandro Cagliari

Tese de doutorado submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul (UFRGS) como requisito parcial para a obtenção do grau de Doutor em Ciências (Genética e Biologia Molecular)

Orientadora: Dra. Márcia Maria Pinheiro Margis

Co-orientador: Dr. Rogério Margis

Porto Alegre, agosto de 2013.

Este trabalho foi desenvolvido no Núcleo de Genômica Funcional de Plantas do Departamento de Genética da Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brasil. O trabalho faz parte do “Projeto Estruturante de Agroenergia do Estado do Rio Grande do Sul” sendo subvencionado pela FINEP (Financiadora de Estudos e Projetos) e pela FAPERGS (Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul). O doutorando obteve bolsa de estudos no Brasil e bolsa de doutorado sanduíche realizado pelo período de um ano na Universidade da California – Davis – EUA do Centro Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Dedico esta tese a minha mãe, Iraci, meu maior exemplo e minha maior inspiração

Ando Devagar

Renato Teixeira

Ando devagar por que já tive pressa
E levo esse sorriso por que já chorei demais
Hoje me sinto mais forte, mais feliz quem sabe,
Só levo a certeza de que muito pouco eu sei
Nada sei.

Conhecer as manhas e as manhãs,
O sabor das massas e das maçãs,
É preciso amor pra poder pulsar,
É preciso paz pra poder sorrir,
É preciso a chuva para florir

Penso que cumprir a vida seja simplesmente
Compreender a marcha e ir tocando em frente
Como um velho boiadeiro levando a boiada
Eu vou tocando dias pela longa estrada eu vou
Estrada eu sou.

Todo mundo ama um dia todo mundo chora,
Um dia a gente chega, no outro vai embora
Cada um de nós compõe a sua história
Cada ser em si carrega o dom de ser capaz
E ser feliz.

Ando devagar porque já tive pressa
E levo esse sorriso porque já chorei demais
Cada um de nós compõe a sua história,
Cada ser em si carrega o dom de ser capaz
E ser feliz.

AGRADECIMENTOS

Aos meus orientadores, Profª. Drª. Márcia Margis e Prof. Dr. Rogério Margis, pelo voto de confiança, paciência, estímulo e amizade aos longos desses anos de trabalho. Pelo exemplo como profissionais e como seres humanos.

Ao Dr. Harada pelos ensinamentos e por possibilitar importante experiência junto à Universidade da Califórnia-Davis.

Às professoras Maria Helena Bodaneze Zanettini, Luciane Passaglia, Eliane Kaltchuk dos Santos e Fernanda Bered pelo apoio científico e contribuições para o bom funcionamento do laboratório.

Aos colaboradores e co-autores Prof. Dr. Felipe Dos Santos Maraschin, Dra. Andreia Turchetto e Dra. Ana Paula Korbes, pelo auxílio e co-orientação ao longo do trabalho.

Ao coodenador administrativo do PPGBM Elmo J. Antunes Cardoso pela atenção e disponibilidade.

Ao Centro Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro.

Às professoras da UnC, Drª Vânia Techio, Drª Elisete Barp e Msc. Celí Araldi pelo incentivo em meus primeiros passos na biologia.

À minha namorada Carol, pela amizade, dedicação e cumplicidade. Por sempre estar ao meu lado em todas as situações e por ser minha maior incentivadora. Te amo!

Aos amigos e co-orientadores esporádicos, Rafael Arenhart, Rafael Oliveira e Lauro Bucker, pelas dicas, auxílio e ensinamentos.

Aos colegas do laboratório de genética vegetal, em especial aos amigos Rafael Rauber, João Bactéria, Ronei, Betinho, Júlio Lima.

Aos grandes amigos e irmãos do Daikatana hakushin-gumi himitsu shakai, Gustavo, Emili, Ivanei, Adenilso, Yuka, Marcelo, Marcela, Marcos, Jaime, Eleandro, Adilso, Carlos, Cleudemar, Giovani, Rudi, Ezequiel e Thiago. Namaste!

Aos amigos de infância Jhonatan, Adriano, Cledson, Vanderlei.

A todos, que de alguma forma contribuíram e conviveram comigo, meus sinceros agradecimentos.

SUMÁRIO

RESUMO

Em sementes, os lipídeos de reserva são compostos, principalmente, por triacilglicerídeos (TAG), os quais se acumulam durante a fase de maturação do embrião ou endosperma e, durante a germinação da semente são catabolizados, fornecendo energia para o estabelecimento e desenvolvimento da plântula. Na presente tese buscamos, primeiramente, contribuir para um melhor entendimento do processo de síntese de ácidos graxos e TAGs em plantas e algas, por meio de um estudo comparativo das diferentes rotas envolvidas na produção e acúmulo de lipídeos de reserva nestes grupos de organismos. As principais rotas de biossíntese de ácidos graxos e TAGs em algas são análogas às apresentadas em plantas e, ao contrário de plantas, as algas acumulam quantidades significativas de lipídeos na forma de TAGs apenas sob condições desfavoráveis de crescimento ou de estresse. Posteriormente, passamos às análises acerca do fator de transcrição eucariótico NF-Y (*Nuclear Factor of Y box*), o qual participa da regulação de processos como maturação, síntese e acúmulo de lipídeos em sementes. NF-Y é um complexo hetero-oligomérico formado por três subunidades (NF-YA, NF-YB e NF-YC). Análises filogenéticas mostraram que a diversificação em todas as subunidades do fator de transcrição NF-Y é resultado de diversos eventos de duplicação ao longo da evolução e diversificação das plantas, muitos dos quais ocorridos após a divergência entre mono e dicotiledôneas. O estudo detalhado dos genes que codificam a subunidade NF-YB, sugere um modelo no qual uma subclasse de genes denominados tipo LEC1 (LEAFY COTYLEDON1) teria se originado em plantas vasculares através de processos de neofuncionalização e/ou subfuncionalização. Análises de expressão e de interação proteína-proteína demonstraram a existência de diferentes parceiros NF-YC para LEC1 e L1L, indicando que os genes do tipo LEC1 em mamona não são redundantes e podem ter evoluído para desempenhar diferentes papéis durante o desenvolvimento da semente de mamona. Por fim, realizamos a identificação dos genes alvo diretos de LEC1 nos estágios cotiledonar e de início da maturação da semente de soja (*Glycine max*). Nossos resultados demonstraram que LEC1 apresenta uma grande quantidade de genes alvo comuns às duas fases de desenvolvimento da semente analisadas, mas também

apresenta enriquecimento em diferentes processos metabólicos nos estágios cotiledonar e de início da maturação. Além disso, LEC1 se liga não somente a promotores alvo enriquecidos na sequência canônica CCAAT como também a outras sequências, com destaque para motivos de ligação de fatores de transcrição do tipo bZIP, o que pode indicar que tais fatores de transcrição podem estar relacionados com a regulação do desenvolvimento e maturação da semente exercida por LEC1.

ABSTRACT

In seeds, lipids of storage are composed mainly of triacilglicerídeos (TAG), which accumulate during the maturation phase of embryo or endosperm, and during seed germination are catabolized, providing energy for the establishment and seedling development . In this thesis, first of all, we focused on the understanding of fatty acid and TAG biosynthesis in plants and algae, through a comparative study of the different pathways involved with lipid production and accumulation in these organisms. The main pathways for fatty acids and TAG biosynthesis in algae are analogous to those demonstrated in higher plants and, in contrast with plants, algae accumulate high levels of TAGs only under unfavorable growth or stress conditions. We also analyzed the NF-Y (Nuclear Factor of Y box) eukaryotic transcription factor, which is involved in the regulation of maturation, lipid biosynthesis and accumulation in seeds. NF-Y is an oligotrimer complex composed by three subunits (NF-YA, NF-YB and NF-YC). Phylogenetic analysis showed that the gene diversification observed for all NF-Y subunits likely resulted from several duplication events along evolution and diversification of plants, some of them occurred after the divergence of monocots and eudicots. The study of the NF-YB subunit encoding genes suggested a model in which LEC1 (LEAFY COTYLEDON1) type gene family originated in vascular plants through a process of neofunctionalization and/or subfunctionalization. Expression analysis and protein interaction experiments demonstrated that are different NF-YC partners for LEC1 and L1L, indicating that LEC1-type genes in castor bean are not redundant and could have evolved to play different roles during castor seed development. Finally, we identified the direct targets of LEC1 in the cotyledon and early maturation stages during soybean seed development (*Glycine max*). We observed that LEC1 has a high number of common direct targets to both stages of soybean seed development, but also presents enrichment in different metabolic process in cotyledon and early maturation stages. Besides that, LEC1 bind not only the promoter regions enriched in CCAAT binding sequences but also other sequences, especially bZIP binding sites, which can indicate that bZIP transcription factor could be related with the regulation of seed development and maturation performed by LEC1.

1 INTRODUÇÃO

1.1 Óleos vegetais

Óleos vegetais desempenham um importante papel em muitos aspectos da vida dos seres humanos, compreendendo desde o fornecimento de energia até sua utilização como matéria-prima para aplicação industrial (Napier and Graham, 2010; Snyder *et al.*, 2009). Segundo o Departamento Americano de Agricultura (USDA, 2013), os óleos vegetais são produzidos, principalmente, a partir de quatro culturas: palma, soja, canola e girassol (Tabela 1), que juntas respondem por aproximadamente 85% do total de óleo vegetal produzido no mundo. Ainda segundo o USDA, o Brasil atualmente é o segundo maior produtor (84,51 milhões de toneladas na safra 2012/2013) e o maior exportador mundial de oleaginosas (37,99 milhões de toneladas na safra 2012/2013).

O óleo vegetal é o bioproduto quimicamente mais similar ao óleo fóssil, apresenta um alto conteúdo energético (Rogalski and Carrer, 2011) e, por sua natureza renovável, apresenta grande potencial para substituir o petróleo na indústria química (Abbadi *et al.*, 2004). Tais características, associadas à busca por fontes renováveis de energia que possam substituir o atual consumo e a dependência por combustíveis fósseis têm levado muitos países a explorar fontes alternativas de combustíveis e matéria-prima industrial, dentre as quais se destaca a busca por combustíveis derivados de óleos vegetais (Durrett *et al.*, 2008; Dyer and Mullen, 2008; Halim *et al.*, 2012; Wilhelm and Jakob, 2011; Xie *et al.*, 2012).

Em sementes, os lipídeos de reserva são compostos, principalmente, por três cadeias de ácidos graxos esterificadas em uma molécula de glicerol, estrutura que passa então a ser denominada triacilglicerídeo (TAG) (Durrett *et al.*, 2008). TAGs são acumulados durante a fase de maturação do embrião ou endosperma e, durante a germinação da semente, são catabolizados, fornecendo energia para o estabelecimento e desenvolvimento da plântula (Voelker and Kinney, 2001).

Tabela 1. Produção mundial de óleo vegetal. Produção mundial de óleo vegetal em milhões de toneladas para cada uma das principais oleaginosas cultivadas.

Produção	2009/10	2010/11	2011/12	2012/13
Palma	45.91	48.68	51.88	55.29
Soja	38.79	41.29	42.35	43
Canola	22.56	23.52	24.3	24.14
Girassol	12.28	12.42	15.34	14.06
Algodão	4.6	4.97	5.26	5.28
Amendoim	4.72	5.08	5.06	5.26
Côco	3.63	3.81	3.5	3.75
Oliva	3.08	3.25	3.24	2.67
Total	141.14	148.69	156.96	159.86

Fonte: Departamento Americano de Agricultura (USDA, 2013).

Muitas espécies de plantas ditas “não-domesticadas” sintetizam e acumulam TAGs enriquecidos em ácidos graxos estruturalmente diferentes, na composição de seus óleos. A maioria destes óleos é majoritariamente composta por um único tipo de ácido graxo, “não usual”, o que simplifica seu processamento e, em muitos casos, inclusive, permite o uso direto do óleo em aplicações industriais (Dyer and Mullen, 2008).

Ácidos graxos “não usuais” podem ser definidos como aqueles que possuem estruturas químicas que diferem das apresentadas pelos principais ácidos graxos constituintes de óleos armazenados em sementes da maioria das culturas convencionais: ácidos palmítico (16:0), esteárico (18:0), oleico (18:1), linoleico (18:2) e linolênico (18:3). Muitos ácidos graxos “não usuais” apresentam propriedades que são desejáveis do ponto de vista industrial, incluindo comprimento da cadeia carbônica, posição e configuração de duplas ligações e/ou presença de grupos funcionais, como grupos hidroxil, no caso do ácido ricinoleico de mamona (Jaworski and Cahoon, 2003). As espécies de plantas que produzem estes tipos de óleos industrialmente importantes, geralmente, apresentam características agronômicas limitantes, tais como sementes pequenas, baixa produtividade, toxicidade ou área de distribuição geográfica

limitada, dificultando a obtenção suficiente de óleo, de uma maneira competitiva e financeiramente viável, para seu uso na indústria. Devido a isto, um dos principais objetivos da pesquisa nesta área tem sido estudar estas espécies selvagens e identificar genes da rota de biossíntese de ácidos graxos, e então transferi-los, com o auxílio de ferramentas biotecnológicas, para espécies cultivadas e altamente produtivas (Cahoon *et al.*, 2007; Jaworski and Cahoon, 2003; Napier, 2007). Neste contexto, o melhor entendimento das enzimas e genes envolvidos na biossíntese de óleos em sementes é essencial para ultrapassar a limitação na produção de ácidos graxos em plantas agronomicamente importantes (Ruiz-Lopez *et al.*, 2012).

1.2 Desenvolvimento de sementes

O desenvolvimento de sementes é um processo crucial no ciclo de vida das plantas, possibilitando a ligação entre duas distintas gerações esporofíticas e, portanto, promovendo a perpetuação da espécie (Gutierrez *et al.*, 2007). O desenvolvimento e a maturação da semente representam uma vantagem evolutiva que permite a muitas plantas resistir a condições ambientais desfavoráveis pela interrupção de seu ciclo de vida, reassumindo seu crescimento quando surgem condições favoráveis a seu reestabelecimento (Bewley, 1997). Além disso, as sementes representam fonte de energia para sustentar a plântula durante a germinação, antes do estabelecimento do processo de fotossíntese (Le *et al.*, 2007).

O desenvolvimento da semente é iniciado com a fusão do óvulo e das células centrais do gametófito feminino com as duas células espermáticas de origem masculina. Essa dupla fertilização, exclusiva de plantas com flores, resulta na formação da região do embrião e do endosperma da semente, respectivamente (Belmonte *et al.*, 2013).

O desenvolvimento de sementes pode ser dividido em dois estágios: desenvolvimento do embrião e endosperma (ou morfogênese) e maturação da semente (West and Harada, 1993).

Durante o estágio embriogênico, o embrião maduro desenvolve-se a partir de uma única célula fertilizada por uma série de processos, juntos denominados de morfogênese (Angelovici *et al.*, 2010; Braybrook and Harada, 2008). Tal processo é geneticamente programado e inclui uma grande gama de mudanças em nível metabólico. A diferenciação celular ocorre sucessivamente ao longo do processo, iniciando com modificações no tecido materno e posterior formação de órgãos filiais, que se tornam tecidos de reserva altamente especializados (Weber *et al.*, 2005). Nos estágios iniciais da morfogênese, o crescimento do embrião é controlado pela restrição física imposta pelo tecido de revestimento da semente. Esse estágio é caracterizado por uma fase de intensa divisão celular e morfogênese, que originaram as principais estruturas morfológicas encontradas no embrião maduro (Wobus and Weber, 1999).

Posteriormente, inicia-se a fase de maturação da semente, a qual inclui o acúmulo de substâncias de reserva, a supressão da germinação, a aquisição de tolerância à dessecação, perda de água e, geralmente, por um processo de indução de dormência. As proteínas, bem como lipídeos, amidos e açúcares acumulados durante a fase de maturação têm como finalidade a proteção do embrião contra a dessecação bem como o fornecimento dos nutrientes necessários para o estabelecimento da plântula após a germinação (Braybrook and Harada, 2008). Uma complexa rede regulatória desencadeia o início da maturação e o acúmulo de substâncias de reserva. Esta inclui uma reprogramação transcricional e fisiológica por meio de rotas responsivas a açúcares e hormônios (Gibson, 2004).

O início do processo de maturação interrompe o desenvolvimento da semente. Este processo de interrupção caracteriza-se como uma fase de transição durante o qual ocorre aumento a mudança de um crescimento mitótico para um crescimento promovido por expansão celular, ganho de atividade

fotossintética e drástica indução de genes relacionados à degradação de produtos de reserva (Figura 1) (Weber *et al.*, 2005).

Após o período de acúmulo de reserva, a semente entra no último período de maturação, o qual é denominado dessecação. Nesta etapa, acentua-se a perda de água, tornando a semente extremamente seca, em preparação para um período de quiescência, que permite ao embrião sobreviver a drásticas condições ambientais até que seja iniciado o processo de germinação (Angelovici *et al.*, 2010; Gutierrez *et al.*, 2007).

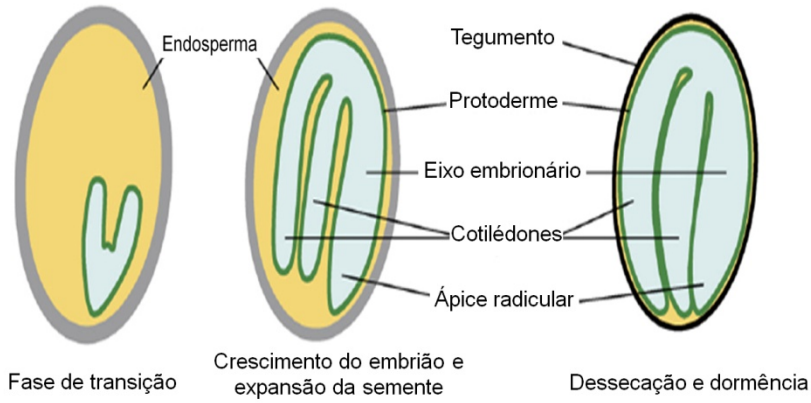


Figura 1. Processo de maturação da semente. Após a fase de transição (de crescimento mitótico para expansão celular), a semente passa por um processo de crescimento do embrião durante o qual se acumulam produtos de reserva. Em seguida, a semente cessa seu crescimento e o tecido materno que forma a testa diferencia-se dando origem ao tegumento protetor que reveste a semente. O processo de maturação termina com a dessecação da semente e, em geral, na indução de dormência (Gutierrez *et al.*, 2007).

1.2.1 Regulação do desenvolvimento da semente

A maturação da semente é metabolicamente controlada, dentre outros fatores, através da disponibilidade de nitrogênio, uma vez que a síntese de

proteínas de reserva é limitada, entre outras coisas, pelo conteúdo de nitrogênio disponível na semente (Rolletschek *et al.*, 2005). Estudos em soja (*Glycine Max*) sugerem que o conteúdo de asparagina em cotilédones em desenvolvimento também pode ser um fator determinante na biossíntese de proteínas de reserva (Hernandez-Sebastia *et al.*, 2005).

Além disso, uma segunda mudança metabólica foi descrita como operando durante a transição para a fase de dessecação (Fait *et al.*, 2006). Nessa etapa, o metabolismo da semente muda de uma etapa de acúmulo de óleo e proteínas de reserva para outra na qual há um aumento no conteúdo de aminoácidos livres, açucares e ácidos graxos degradados. Durante esta fase de desenvolvimento, o acúmulo de sacarose parece ser um pré-requisito para a aquisição de tolerância à dessecação (Buitink *et al.*, 2006).

Os mecanismos latentes que regulam o acúmulo dos metabólitos durante a fase de maturação do embrião permanecem desconhecidos (Gutierrez *et al.*, 2007). Entretanto, em sementes de leguminosas tem sido descrito que a sinalização metabólica é regulada através do processo de fosforilação protéica (Weber *et al.*, 2005). Por exemplo, em estudo proteômico em larga escala realizado em canola (*Brassica napus*), demonstrou-se que um significativo número de fosfoproteínas funcionalmente diversas é expresso durante a fase de expansão da semente (Agrawal and Thelen, 2006).

1.2.2 Sinalização hormonal mediada por ABA

Ácido abscísico (ABA) regula muitos processos importantes do desenvolvimento de uma planta, incluindo a síntese de proteínas e lipídeos de reserva em sementes, a aquisição de tolerância à dessecação e dormência e a inibição da fase de transição de crescimento embriogênico para germinativo (Finkelstein *et al.*, 2002; Nakashima and Yamaguchi-Shinozaki, 2013).

Na semente, ABA é primeiramente produzido nos tecidos maternos e posteriormente no embrião, onde seu pico de acúmulo ocorre durante a maturação (Frey *et al.*, 2004). A transição de uma fase de divisão celular para uma de crescimento celular e acúmulo de reservas, que ocorre durante a maturação, está correlacionada com o aumento do conteúdo de ABA na semente, consistente com o fato de que ABA pode induzir a expressão de inibidores de cinases dependentes de ciclinas (ICK1) (Wang *et al.*, 1998) que podem levar ao retardo na transição da fase G1 para a S do ciclo celular (Finkelstein *et al.*, 2002).

ABA esta relacionado com a síntese de protetores osmóticos e proteínas LEA (*late embryogenesis abundant*), sob condições de estresse osmótico, através da ativação de fatores de transcrição do tipo bZIP, principalmente ABI5, bZIP67, AREB3 e EEL (Bensmihen *et al.*, 2005; Bensmihen *et al.*, 2002; Nakashima *et al.*, 2009).

ABA é um hormônio chave requerido para o estabelecimento do processo de maturação da semente. A importância de ABA nesse processo tem sido demonstrada em inúmeros experimentos de modulação nos quais reduzidas quantidades de ABA resultaram na mudança de programação de maturação da semente para um programa de preparação para a germinação (Finch-Savage and Leubner-Metzger, 2006; Nambara and Marion-Poll, 2003). O nível de ABA aumenta drasticamente com o início da fase de maturação, permanece alto durante a maior parte desta fase, sofre declínio no estádio final da maturação e apresenta-se muito baixo durante o desenvolvimento da plântula (Hays *et al.*, 1999).

Durante as últimas fases da maturação, a indução de dormência é dependente de uma forte sinalização promovida por ABA (alta proporção de ABA em relação à giberelinas) que é principalmente regulada pelo balanço dinâmico entre síntese e degradação de ABA (Cadman *et al.*, 2006). Estudos do padrão de expressão de genes envolvidos na síntese (incluindo zeaxantina epoxidase e 9-cis-epoxicarotenóide dioxigenase) e catabolismo de ABA (incluindo ácido abcísico hidroxilase) têm demonstrado seus papéis chave no acúmulo temporal e espacial

de ABA em sementes (Nambara and Marion-Poll, 2005). Diferentes etapas de síntese e sinalização mediada por ABA ocorrem em diferentes tecidos de uma semente durante sua maturação (Gutierrez *et al.*, 2007). Em *Arabidopsis thaliana*, por exemplo, genes CYP707As, que codificam ácido abscísico hidroxilases, são conhecidos como sendo os principais reguladores da degradação de ABA. CYP707A1 e CYP707A2 desempenham importante papel na redução do conteúdo de ABA na fase intermediária da maturação do embrião, tanto no embrião quanto no endosperma durante os estágios finais da maturação (Okamoto *et al.*, 2006). Outros pesquisadores também demonstraram que a expressão de genes CYP707As é o principal mecanismo de regulação do catabolismo de ABA em sementes de feijão (*Phaseolus vulgaris*) e cevada (*Hordeum vulgare*) (Millar *et al.*, 2006; Yang and Zeevaart, 2006).

Durante a maturação da semente, a sinalização mediada por ABA está intimamente relacionada com a expressão de quatro importantes fatores de transcrição: *Abscisic Acid Insensitive* (ABI3), ABI4, ABI5, FUSCA3 (FUS3), *Leafy Cotyledon 1* (LEC1) and LEC2 (Finkelstein *et al.*, 2002).

Ensaios de germinação com adição exógena de ABA mostraram que o duplo mutante *abi/lec1* apresentou sensibilidade cerca de 30 vezes menor a ABA do que o mutante monogênico *abi.* Embora, neste mesmo estudo, o mutante *lec1* tenha apresentado pouca alteração quanto à sensibilidade a ABA, LEC1 parece potencializar a resposta a ABA através da interação genética com ABI3, ABI4 e ABI5. No caso específico de ABI3, LEC1 (assim como FUS3) regula positivamente a abundância da proteína ABI3 em sementes (Parcy *et al.*, 1997).

A observação de que mutantes *lec1* e *fus3* apresentam viviparidade ocasional levanta a possibilidade de que tais genes afetem a resposta a ABA (Raz *et al.*, 2001). Consistente com essa descoberta, a expressão ectópica de FUS3 resultou em elevados níveis de ABA (Gazzarrini *et al.*, 2004), uma interessante observação dado que o acúmulo de RNAs codificantes de proteínas de reserva mediado por LEC1 e FUS3 é aumentado por ABA. A ativação do gene FUS3 por LEC1 ocorre antes da ativação dos genes de proteínas de estocagem da semente

e a expressão de FUS3 é aumentada por ABA exógeno (Kagaya *et al.*, 2005a; Kagaya *et al.*, 2005b). Essas evidências sugerem que ABA e FUS3 regulam positivamente um ao outro, embora o mecanismo de regulação ainda seja desconhecido (Braybrook and Harada, 2008).

O atual modelo para a ação de ABA propõe que, na ausência de ABA, a fosfatase PP2C regula negativamente proteínas cinases da subclasse III (SnRK2) através de interação direta e defosforilação. Durante o desenvolvimento da semente, os níveis elevados de ABA endógeno ativam a ligação de RCAR (*Regulatory component of ABA receptor 1*) à ABA e sua interação com PP2C, resultando na inibição de sua atividade fosfatase. Como consequência, SnRK2 torna-se livre da repressão dependente de PP2C, fosforila fatores de transcrição bZIPs que, por sua vez, ligam-se às regiões promotoras de seus genes alvos, especialmente proteínas LEA, ativando sua expressão (Nakashima and Yamaguchi-Shinozaki, 2013). As proteínas LEA são altamente hidrofílicas e têm como função a manutenção do conteúdo de água e a proteção de macromoléculas em células desidratadas (Chakrabortee *et al.*, 2007).

1.2.3 Sinalização mediada por açúcares

Açúcares, além de desempenharem seus papéis como fonte de energia e carbono, também atuam como moléculas sinalizadoras para regular a expressão gênica (Hartig and Beck, 2006; Rolland *et al.*, 2002; Wang and Ruan, 2012; Wang and Ruan, 2013). Gradientes de concentração de metabólitos fornecem sinais para o início da fase de maturação da semente bem como para sua regulação. Muitos estudos têm demonstrado que a mudança da fase de expansão para a fase de acúmulo de substâncias é fortemente afetada por sinalização mediada por açucares (Wang and Ruan, 2013), especialmente por um aumento na proporção de sacarose: hexose (glicose) no embrião (Weber *et al.*, 2005).

A proliferação sequencial do endosperma e do embrião são fortemente controladas por diferentes mecanismos regulatórios, incluindo sinalização por

açucares (Wang and Ruan, 2013). As divisões nucleares no endosperma ocorrem mais cedo e mais rapidamente do que a proliferação celular do embrião (Bate *et al.*, 2004; Nowack *et al.*, 2007) e são resultantes de um gradiente de concentração de glicose maior na região do endosperma do que na região do embrião, favorecendo assim a divisão nuclear sobre a proliferação celular do embrião durante o estabelecimento da semente (Wang and Ruan, 2012). Consistente com esses dados, um inibidor da enzima invertase (responsável pela conversão de sacarose em glicose ou frutose) foi identificado na fronteira entre endosperma e embrião em sementes de milho (Bate *et al.*, 2004), auxiliando na diminuição do fluxo de hexose para o embrião, favorecendo assim a divisão nuclear no endosperma e um estado quiescente no embrião durante esse estágio. Em termos gerais, a sinalização mediada por alterações na proporção sacarose:glicose, que é conhecida por apresentar potencial de ação regulatória em nível transcricional e pós-transcricional, controla os processos de estocagem e diferenciação através da regulação de enzimas metabólicas, expressão e atividade gênica (Gibson, 2005; Rolland *et al.*, 2006).

1.3 Fatores de transcrição envolvidos na maturação da semente e síntese de ácidos graxos

Nos vegetais superiores, a biossíntese da maioria dos ácidos graxos é fisiologicamente acoplada com o desenvolvimento da semente. Vários reguladores mestres, que controlam a maturação da semente foram identificados em *A. thaliana*, incluindo LEC1, LEC2, ABI3 e FUS3. Eles agem através da interação direta com elementos regulatórios presentes na região promotora de genes relacionados à maturação (Gutierrez *et al.*, 2007). Mutações em qualquer um desses quatro fatores de transcrição afetam a maturação, causando a transição de uma fase de desenvolvimento embriogênico para o processo de germinação (Vicente-Carbajosa and Carbonero, 2005). Embora cada um desses fatores de transcrição ative um grupo específico de genes alvos durante a fase de maturação, em geral, estes genes são controlados de uma maneira sobreposta, ao invés de hierárquica (Gutierrez *et al.*, 2007).

Em termos de alterações fenotípicas, FUS3, LEC1 e LEC2 pertencem ao tipo *Leafy Cotyledon (*Genes LEC*)* (Meinke *et al.*, 1994), uma vez que a perda de função de qualquer um deles resulta no aparecimento de características de folhas primitivas nos cotilédones (formação de tricomas e alterações no conteúdo de antocianinas, por exemplo), além de crescimento precoce dos meristemas radicular e apical, acúmulo de substâncias de reserva e redução no estabelecimento da tolerância à dessecação e dormência. A função desses fatores de transcrição é aparentemente redundante, atuando de forma sobreposta na regulação da maturação de sementes (Vicente-Carbajosa and Carbonero, 2005).

Os três genes LEC (LEC1, LEC2 e FUS3) apresentam papel central na embriogênese como indicado por experimentos mostrando fenótipos de ganho e perda de função. A expressão ectópica dos três genes LEC ocasionou o aparecimento de características intrínsecas da fase de maturação do embrião em tecidos vegetativos e reprodutivos. Além disso, embriões com mutações de perda de função nestes genes são intolerantes à dessecação e apresentam defeitos no acúmulo de reservas (Kagaya *et al.*, 2005b; Lotan *et al.*, 1998; Stone *et al.*, 2001). Além disso, a expressão ectópica de LEC1 e LEC2 desencadeou o processo de embriogênese somática em células vegetativas (Lotan *et al.*, 1998; Stone *et al.*, 2001). Genes LEC também são necessários para o progresso normal para a fase de morfogênese durante o processo de embriogênese zigótica. Tais trabalhos demonstram, portanto, que fatores de transcrição do tipo genes LEC são suficientes para induzir embriogênese somática, indicando sua importância para a totipotência celular (Braybrook *et al.*, 2006). Uma indicação de como os genes LEC podem exercer papel na indução da embriogênese somática, mesmo na ausência de auxina exógena (hormônios normalmente utilizados para essa finalidade), vem de um estudo mostrando que LEC2 pode rapidamente ativar os genes YUC2 e YUC4, enzimas chaves na biossíntese de auxinas (Stone *et al.*, 2008).

Fatores de transcrição do tipo LEC exercem controle também sobre a principal fase de maturação da semente: a síntese de macromoléculas de

reserva. Mutantes LEC apresentaram defeitos no acúmulo de proteínas de reserva e lipídeos. Por outro lado, plantas expressando ectopicamente genes LEC acumularam proteínas e lipídeos característicos de sementes em tecidos reprodutivos e vegetativos (Baud *et al.*, 2007; Santos Mendoza *et al.*, 2005; Stone *et al.*, 2008; Wang *et al.*, 2007). Esses resultados indicam que fatores de transcrição do tipo LEC são necessários e suficientes para a indução da fase de maturação durante a embriogênese somática (Braybrook and Harada, 2008).

Embora estruturalmente relacionado à FUS3 e LEC2, ABI3 não é considerado como pertencente ao grupo de fatores de transcrição que produz alterações fenotípicas do tipo LEC. Perda de função de ABI3 resulta em falha na degradação de clorofila, alteração no acúmulo de substâncias de reserva, perda de tolerância à dessecação e dormência e insensibilidade a ABA após a germinação (Vicente-Carbajosa and Carbonero, 2005).

Análises de duplos mutantes com *lec1*, *lec2* e *fus3* sugerem que LEC1 pode atuar a montante de LEC2 e FUS3 (Meinke *et al.*, 1994). Além disso, outro estudo utilizando análises genéticas da regulação da expressão gênica em mutantes simples e múltiplos (To *et al.*, 2006) demonstrou que a expressão tanto de FUS3 quanto ABI3 é controlada por uma complexa e redundante rede de regulação transcricional envolvendo LEC1 e LEC2 (Figura 2).

Existem evidências de que esses *loci* podem também estar envolvidos com a regulação do metabolismo de ácidos graxos (Mu *et al.*, 2008). Em *A. thaliana*, por exemplo, observou-se que níveis reduzidos de mRNA de FUS3 coincidiram com a redução no conteúdo de óleo em sementes transgênicas e que os níveis de transcritos de FUS3 (regulados positivamente por LEC1 e LEC2) aumentam concomitantemente com os níveis de genes envolvidos nas etapas iniciais da síntese de lipídeos, sugerindo o envolvimento desses fatores de transcrição (especialmente FUS3) na biossíntese de lipídeos (Wang *et al.*, 2007).

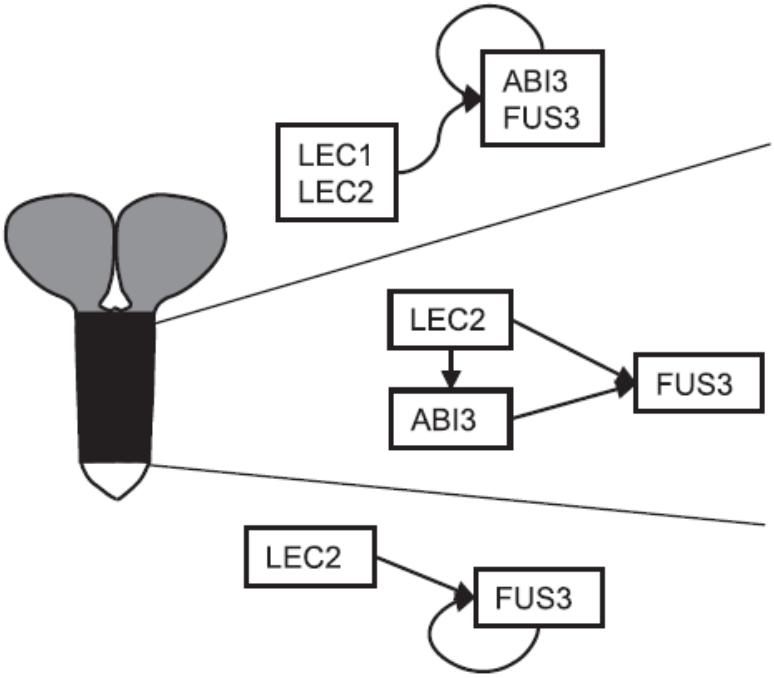


Figura 2. Complexa rede de regulação local e redundante. Representação esquemática do controle genético em vários tecidos embriogênicos (branco, ápice da raiz; preto, eixo embrionário; cinza, cotilédones) em *A. thaliana*. Figura adaptada de (Santos-Mendoza *et al.*, 2008).

1.3.1 O fator de transcrição NF-Y e os genes do tipo LEC1 em plantas vasculares

O fator de transcrição eucariótico NF-Y/CBF (*Nuclear Factor of the Y box*/CCAAT *Binding Factor)* é um complexo hetero-oligomérico formado por pelo menos três subunidades, NF-YA (também conhecida como HAP2 ou CBF-B), NF-YB (HAP3 ou CBF-A) e NF-YC (HAP5 ou CBF-C) (Kwong *et al.*, 2003). Cada subunidade é codificada por um único gene em animais e fungos e por famílias gênicas em plantas e contém um domínio evolutivamente conservado, que é responsável pela ligação ao DNA e interação proteína-proteína (Thirumurugan *et al.*, 2008).

Em mamíferos, onde a montagem do complexo NF-Y é melhor entendida, as subunidades NF-YB e NF-YC formam um heterodímero que direciona-se ao

núcleo, onde a terceira subunidade (NF-YA) é recrutada para compor o complexo NF-Y completo (Sinha *et al.*, 1996; Sinha *et al.*, 1995), que, por sua vez, pode atuar regulando a transcrição tanto positiva quanto negativamente (Ceribelli *et al.*, 2008; Peng and Jahroudi, 2002; Peng and Jahroudi, 2003; Siefers *et al.*, 2009). Em plantas, onde é encontrado um grande número de genes codificando para cada uma das três subunidades, informações quanto à dinâmica de formação do complexo *in vivo* são escassas.

Membros das subunidades que compõem o fator de transcrição NF-Y estão envolvidos na regulação da expressão de genes que participam de uma grande gama de diferentes processos metabólicos, incluindo regulação da fase G1 para S durante o ciclo celular (Matuoka and Yu Chen, 1999), resposta a danos celulares e a estresses (Huang *et al.*, 1999; Matuoka and Chen, 2002), crescimento reprodutivo em flores e sílicas, embriogênese e maturação da semente (Kwong *et al.*, 2003; Lee *et al.*, 2003; Lotan *et al.*, 1998; Watanabe and Yamamoto, 2009; Yu *et al.*, 2011), dentre outras.

Os genes que codificam a subunidade NF-YB, por sua vez, podem ser divididos em genes do tipo LEC1 e genes do tipo não LEC1 baseado na presença de resíduos de aminoácidos característicos no domínio conservado da proteína (Holdsworth *et al.*, 2008; Kwong *et al.*, 2003) e com base no padrão de expressão, uma vez que genes do tipo LEC1 apresentam expressão restrita a sementes e sílicas em desenvolvimento, enquanto genes do tipo não LEC1 apresentam expressão ubíqua (Cao *et al.*, 2011; Kwong *et al.*, 2003; Lotan *et al.*, 1998). Em Arabidopsis, os genes do tipo LEC1 são representados por LEC1 e seu homólogo LEC1-like (L1L), sendo que LEC1, por ter sido o primeiro a ser descrito, é bem melhor caracterizado do que L1L.

A expressão de LEC1 em *A. thaliana* é restrita à embriogênese e é reprimida em tecidos vegetais pós-germinativos, em parte pelo controle exercido por Pickle (PKL), um fator de remodelamento da cromatina (Lotan *et al.*, 1998; Ogas *et al.*, 1999). A atividade de LEC1 deve ser reprimida após o desenvolvimento do embrião de forma a permitir que o desenvolvimento

vegetativo prossiga (Braybrook and Harada, 2008). Diversos trabalhos sugerem que tal repressão de LEC1 bem como dos demais genes LEC ocorre através de mudanças na estrutura da cromatina (Steimer *et al.*, 2004; Suzuki *et al.*, 2007; Tanaka *et al.*, 2008; Tsukagoshi *et al.*, 2007).

Análises da expressão ectópica de LEC1 demonstraram que a expressão em sementes de genes relacionados a proteínas de estocagem de substâncias de reserva é controlada por LEC1 através da modulação da expressão de ABI3 e FUS3. Estes trabalhos também demonstraram que a expressão de FUS3 e ABI3 foi reprimida em sílicas em desenvolvimento de Arabidopsis mutadas para LEC1 (Kagaya *et al.*, 2005b).

Em *A. thaliana* já foi demonstrado que LEC1 desempenha papel crítico para o desenvolvimento normal do embrião tanto durante as fases iniciais quanto finais da embriogênese e sua superexpressão foi suficiente para conferir competência embriogênica a células vegetativas (Kwong *et al.*, 2003). A expressão ectópica de LEC1 também resultou na ativação de rotas de maturação de semente em tecidos vegetativos (Casson and Lindsey, 2006).

Além das funções conhecidas dos genes do tipo LEC1 durante a embriogênese e maturação da semente (Kwong *et al.*, 2003; Lee *et al.*, 2003; Lotan *et al.*, 1998; Watanabe and Yamamoto, 2009; Yu *et al.*, 2011), existem evidências de que LEC1 e L1L participam direta ou indiretamente na ativação de genes de estocagem em sementes, resultando no acúmulo de macromoléculas de reserva (Braybrook and Harada, 2008). LEC1 e L1L também apresentam funções como reguladores chave da biossíntese de ácidos graxos: a superexpressão induzida de LEC1 em Arabidopsis resultou na indução das rotas de biossíntese de TAGs (Mu *et al.*, 2008) e a superexpressão dos ortólogos de LEC1 e L1L em *Brassica napus* e milho (*Zea mays*) foi suficiente para aumentar o conteúdo de ácidos graxos nestas espécies (Shen *et al.*, 2010; Tan *et al.*, 2011).

1.4 A mamona

A mamona (*Ricinus communis L.*) é uma planta xerófila e heliófila da família Euphorbiaceae. De origem africana, apresenta grande variabilidade, com seis subespécies e 25 variedades botânicas, além de milhares de cultivares comerciais híbridas em todo o mundo (EMBRAPA, 2004). A mamoneira é uma planta de hábito arbustivo, de portes diversos, com variadas colorações de caule, folhas e racemos (cachos), podendo ou não possuir cera no caule e pecíolo. As sementes apresentam tamanho e formato variados, além de grande variabilidade de coloração (Chiaradia, 2005).

Os antigos egípcios já faziam uso da mamona em celebrações religiosas e na medicina popular. No Brasil, sua introdução ocorreu durante a colonização portuguesa com a finalidade de utilizar seu óleo para iluminação e para lubrificação dos eixos das carroças. Seu cultivo, historicamente, foi uma das opções para o semiárido Nordestino, sobretudo entre as décadas de 1940 e 1960. Contudo, a partir do início da década de 1980, ocorreu um contínuo declínio da área cultivada, resultando na desativação de indústrias beneficiadoras e na redução da importância da atividade na economia regional. Posteriormente, o cultivo da mamona desenvolveu-se nas regiões Sudeste e Sul e revitalizou-se no Nordeste do Brasil, graças ao desenvolvimento de técnicas que facilitaram a mecanização da colheita e o desenvolvimento de variedades mais rentáveis, de baixo porte e indeiscentes, cujos frutos têm maturação sincronizada, o que permite colheita única (Vital, 2005).

O Brasil ocupa atualmente a terceira posição como maior produtor mundial de mamona atrás somente de Índia e China (FAO, 2013). Segundo dados fornecidos pela Companhia Nacional de Abastecimento (CONAB, 2013), a safra 2010/2011 de mamona foi de 132.7 mil toneladas distribuídas numa área de 195.1 mil hectares e com uma produção média de 681 Kg/ha. Quanto à safra 2011/2012, devido a problemas climáticos, a produção de mamona está estimada em cerca de 105 mil toneladas. A maior região brasileira produtora de mamona é a região Nordeste (122,6 mil toneladas na safra 2010/2011), especialmente o estado da Bahia (100,1 mil toneladas na safra 2010/2011) (CONAB, 2013).

Embora o potencial teórico seja de 10.000 kg/ha de bagas, a produção na Bahia alcança apenas 10% deste valor, reflexo da utilização cultivares locais, resultantes da miscigenação de variedades com frutos deiscentes, pouca ou nenhuma resistência a pragas e doenças, e baixa produtividade (Coelho, 1979; Kouri, 2004). Isso faz com que estudos mais aprofundados se tornem essenciais para o correto delineamento de estratégias efetivas de melhoria da cultura de mamona no Brasil (Santos-Filho, 2004).

No caso da mamona, o ácido ricinoleico (ácido 12-hidroxi-octadeccis-9-enoico: 18:1-OH), o qual possui um grupo funcional hidroxila (OH) no carbono 12 de sua cadeia, é um ácido graxo não usual de interesse econômico (Dyer and Mullen, 2008), a partir do qual se obtém grande variedade de subprodutos, derivados das modificações de suas cadeias carbônicas, que são utilizados para a fabricação de desinfetantes, óleos lubrificantes, tintas e corantes; manufatura de fungicidas, e como base para inseticidas (Vital, 2005). O hidroxiácido graxo é formado pela hidroxilação do ácido oléico (18:1) esterificado na posição Sn-2 da fosfatidilcolina. A reação é catalisada por uma oleato-12-hidroxilase (FAH12) do retículo endoplasmático e os ácidos graxos resultantes são armazenados na semente na forma de triacilglicerídeos (TAG) (Bafor *et al.*, 1991).

1.5 A soja

A soja [*Glycine max* (L.) Merrill] é uma espécie anual, de autofecundação, pertencente à família das leguminosas e representa uma cultura de grande importância no cenário mundial, sendo utilizada para prover alimentos para humanos e animais, e, mais recentemente, como matéria-prima na produção de biodiesel (www.biodiesel.gov.br). Além disso, a soja apresenta propriedades valiosas para a saúde, sendo também utilizada na obtenção de produtos farmacêuticos (Duranti, 2006). O óleo de soja é o líder mundial dos óleos vegetais e é responsável por 20 a 24% de todas as gorduras e óleos consumidos no mundo (Gepts *et al.*, 2005).

A produção brasileira de soja na safra 2012/2013 foi estimada 81.28 milhões de toneladas. Este volume é 22,4% superior à produção obtida na safra 2011/2012, quando foram colhidas 66,38 milhões de toneladas (CONAB, 2013). Previsões indicam que a produção agrícola brasileira vai crescer mais durante a próxima década do que qualquer outra no mundo, aumentando em mais de 40% até o ano de 2019 (Tollefson, 2010).

Recentemente, foi realizado o sequenciamento do genoma da cultivar Willians 82 de soja, compreendendo 950 Mb, o que representa 85% do genoma total da espécie (Schmutz *et al.*, 2010). Contendo 20 cromossomos, foram preditos 46.430 genes que codificam proteínas, ou seja, 70% a mais do que é predito para *A. thaliana*. Dois eventos de duplicação do genoma parecem ter ocorrido e estes datam de 59 e 13 milhões de anos, resultando em um genoma altamente duplicado, com 75% dos genes apresentando-se em múltiplas cópias. Os dois eventos de duplicação do genoma foram seguidos por inúmeros rearranjos cromossômicos, perdas e diversificação gênica. Estes resultados vêm possibilitando estudos visando uma maior compreensão das respostas da planta de soja às inúmeras alterações ambientais, bem como um melhor entendimento da biologia do organismo como um todo.

2 Justificativa

A habilidade das plantas em produzir sementes tem conferido significativa vantagem seletiva, sendo responsável em grande parte pelo sucesso adaptativo das angiospermas. Neste contexto, o processo de maturação da semente e aquisição de tolerância à dessecação conferiram vantagens adaptativas frente a ambientes adversos, fazendo com que esta fase fosse integrada ao ciclo de vida durante a evolução das angiospermas. Programas regulatórios precisos que iniciam e finalizam o processo de maturação evoluíram de forma a incorporar esta fase à embriogênese (Vicente-Carbajosa and Carbonero, 2005).

O processo de maturação que ocorre em vários tecidos da semente (endosperma, embrião, dentre outros) contribui para a qualidade da semente, permitindo uma dispersão eficiente e o estabelecimento da plântula. A qualidade da semente depende, entretanto, do fino controle da morfogênese do embrião, maturação e germinação. Neste cenário, ainda é evidente a carência de informação quanto à natureza e origem dos mecanismos moleculares que controlam tanto o início da maturação quanto aqueles que previnem a divisão e o crescimento celular (Santos-Mendoza *et al.*, 2008).

Excetuando-se os genes que codificam enzimas de estocagem, cuja estrutura dos *cis* elementos regulatórios é bem documentada, muito pouco se sabe sobre as interações entre proteínas mestres reguladoras e seus prováveis genes alvo. Além disso, parte do processo de maturação da semente é indiretamente regulado por estes genes mestres via fatores de transcrição secundários, que ativam seu próprio programa transcricional requerido para a maturação da semente. Estes fatores de transcrição secundários permanecem, em geral, desconhecidos e a ligação entre alguns processos regulatórios relacionados à maturação e ao controle exercido por genes que codificam fatores de transcrição mestres ainda não é totalmente compreendida (Gutierrez *et al.*, 2007). Portanto, o conhecimento das interconexões entre diferentes redes de regulação é um dos principais desafios no estudo do processo de maturação da semente (Kagaya *et al.*, 2005b). Além disso, a identificação dos genes alvos

desses reguladores chave pode contribuir para o melhor entendimento da cascata de regulação envolvida durante os processos de desenvolvimento e maturação da semente.

A existência de múltiplos genes para cada uma das subunidades da família NF-Y no genoma das plantas levanta a questão se a especificidade da interação das subunidades é determinada pela especificidade da interação proteína-proteína, pela expressão tecido específica de cada gene ou se é uma combinação de ambas, especificidade de interação proteína-proteína e expressão gênica. Para entender a variação e diversidade funcional dos complexos NF-Y em plantas, as especificidades da interação entre as proteínas NF-YA, NF-YB e NF-YC precisam ser elucidadas (Thirumurugan *et al.*, 2008).

As principais rotas metabólicas necessárias para o acúmulo de óleo, amido e proteínas já são bem caracterizadas. Entretanto, os mecanismos regulatórios que interconectam estas várias rotas permanecem obscuros. Tal conhecimento pode fornecer novas ferramentas moleculares que visem a melhoria da qualidade das plantas cultivadas (Focks and Benning, 1998). Neste contexto, o estudo dos fatores de transcrição do tipo LEC1 pode fornecer informações importantes quanto a estes eventos regulatórios, possibilitando inclusive a identificação dos genes alvos desses fatores de transcrição (Braybrook and Harada, 2008). Além disso, fatores de transcrição semente-específicos representam alvos promissores visando o aumento na produção de óleo em plantas cultivadas, tais como soja e mamona, devido à relação destes com a síntese de lipídeos (Mu *et al.*, 2008). A produção de óleos industrialmente importantes em plantas cultivadas é um dos maiores objetivos da biotecnologia de plantas e representa um importante passo na transição de uma sociedade baseada na utilização de óleos fósseis para uma economia mais sustentável, baseada na utilização de fontes mais limpas e renováveis de energia (Dyer and Mullen, 2008).

3 OBJETIVOS

3.1 Objetivo geral

Esta tese tem como objetivo o estudo da história evolutiva do fator de transcrição eucariótico NF-Y e seu papel na regulação do desenvolvimento e acúmulo de macromoléculas de reserva, especialmente lipídeos, em sementes de oleaginosas.

3.2 Objetivos específicos

1) Descrever as principais rotas metabólicas envolvidas na síntese e acúmulo de TAGs em plantas e algas;

2) Identificar e analisar filogeneticamente os genes que codificam as subunidades NF-YA, NF-YB e NF-YC que compõem o complexo transcricional NF-Y em eucariotos;

3) Identificar e analisar filogeneticamente os genes que codificam a subunidade NF-YB com ênfase nos processos evolutivos relacionados à origem dos genes do tipo LEC1 a partir de genes do tipo não LEC1;

4) Identificar os genes que codificam o fator de transcrição do tipo LEC1 (LEC1 e L1L) e a subunidade NF-YC em sementes de mamona;

5) Estudar o padrão de expressão dos genes LEC1, L1L e a subunidade NF-YC durante o desenvolvimento da semente;

6) Identificar as proteínas dos tipos NF-YC que interagem com LEC1 e L1L na formação do fator de transcrição NF-Y em mamona;

7) Identificar os genes alvos regulados por LEC1 durante as fases iniciais do desenvolvimento de semente em soja;

4 RESULTADOS

Os resultados serão apresentados em cinco capítulos. O primeiro capítulo apresenta uma revisão sobre a biossíntese de triacilglicerídeos (TAGs) em plantas vasculares e algas. As duas primeiras seções desse capítulo comparam os perfis de ácidos graxos e as rotas para acúmulo de TAGs nestes organismos. A terceira seção sumariza algumas aplicações econômicas de TAGs e os esforços biotecnológicos objetivando a manipulação genética de ácidos graxos e TAGs para a produção de óleos importantes economicamente em plantas cultiváveis e algas.

O segundo capítulo apresenta os resultados sobre a história evolutiva do fator de transcrição eucariótico NF-Y (CBF –CCAAT Binding Factor), discutindo causas e consequências do processo de duplicação gênica que originaram as famílias gênicas que codificam cada uma de suas três subunidades em plantas, algumas das quais diretamente envolvidas na síntese e no acúmulo de lipídeos em sementes.

O terceiro capítulo apresenta os resultados dos estudos sobre os processos evolutivos ocorridos na subunidade NF-YB do fator de transcrição CBF e que acabaram resultando no surgimento dos genes do tipo LEC1 em plantas vasculares, a partir de genes do tipo não-LEC1.

No quarto capítulo são apresentados os resultados referentes à identificação dos genes do tipo LEC1 e suas interações com membros da subunidades NF-YC ao longo do desenvolvimento da semente de mamona (*Ricinus communis L*.).

O quinto capítulo apresenta parte dos resultados obtidos durante doutorado sanduíche realizado na Universidade da California–Davis sob orientação do Dr. John Harada, durante o qual realizamos a identificação dos genes alvos de LEC1 nos estágios iniciais (cotiledonar e início da fase de maturação) do desenvolvimento da semente de soja (*Glycine max*).

4.1 Capítulo 1

O Presente capítulo foi publicado no periódico ***International Journal of Plant Biology*** (2011).

Biosynthesis of Triacylglycerols (TAGs) in plants and *algae*

Alexandro Cagliari,[1] Rogerio Margis,[2,3] Felipe dos Santos Maraschin,[4] Andreia Carina Turchetto-Zolet,[1,3] Guilherme Loss,[3] Marcia Margis-Pinheiro[1,3]

[1]Departamento de Genética, [2]Departamento de Biofísica, [3]Centro de Biotecnologia, [4]Departamento de Botânica, Universidade Federal do Rio Grande do Sul, Brazil

Abstract

Triacylglycerols (TAGs), which consist of three fatty acids bound to a glycerol backbone, are major storage lipids that accumulate in developing seeds, flower petals, pollen grains, and fruits of innumerous plant species. These storage lipids are of great nutritional and nutraceutical value and, thus, are a common source of edible oils for human consumption and industrial purposes. Two metabolic pathways for the production of TAGs have been clarified: an acyl CoA-dependent pathway and an acyl-CoA-independent pathway. Lipid metabolism, specially the pathways to fatty acids and TAG biosynthesis, is relatively well understood in plants, but poorly known in algae. It is generally accepted that the basic pathways of fatty acid and TAG biosynthesis in algae are analogous to those of higher plants. However, unlike higher plants where individual classes of lipids may be synthesized and localized in a specific cell, tissue or organ, the complete pathway, from carbon dioxide fixation to TAG synthesis and sequestration, takes place within a single algal cell. Another distinguishing feature of some algae is the large amounts of very long-chain polyunsaturated fatty acids (VLC-PUFAs) as major fatty acid components. Nowadays, the focus of attention in biotechnology is the isolation of novel fatty acid metabolizing genes, especially elongases and desaturases that are responsible for PUFAs synthesis, from different species of algae, and its transfer to plants. The aim is to boost the seed oil content and to generate desirable fatty acids in oilseed crops through genetic engineering approaches. This paper presents the current knowledge of the neutral storage lipids in plants and algae from fatty acid biosynthesis to TAG accumulation.

Introduction

The present review describes the current understanding of the fatty acid and triacylglycerol (TAG) biosynthesis in vascular plants and algae. The two initial sections compare fatty acid profiles and the pathways for TAG accumulation present in these organisms, emphasizing the peculiarities of each group. The third section summarizes some economical applications of TAG molecules. In addition, it discusses some biotechnological efforts aimed at the genetic manipulation of fatty acids and TAG content for the production of nutritionally and industrially desirable oils in crop plants and algae. Other classes of lipids, such as carotenoids that also belong to the lipid class of compounds, are not discussed here. For other lipid classes, readers can report to recent reviews on algae[1-3] and on vascular plants.[4-8]

Fatty acid biosynthesis in plants

Triacylglycerols (TAGs), as a highly reduced form of carbon, are an important energy reserve in plant seeds, providing nutrients for subsequent germination and seedling development. These storage lipids are of great nutritional and nutraceutical value, and a common source of edible oils for human consumption and industrial purposes.[9]

TAGs of most seeds usually contain the same acyl groups that are found in membrane lipids: palmitic acid (16:0), stearic acid (18:0), oleic acid (18:1Δ^9), linoleic acid (18:2$\Delta^{9,12}$), and α-linolenic (18:3$\Delta^{9,12,15}$).[10] These fatty acids are often referred to as common fatty acids.[11] The biosynthesis of these five major fatty acids occurs primarily in two subcellular compartments: the *de novo* synthesis (*de novo* synthesis is defined as the synthesis of complex molecules from simple molecules, as opposed to being recycled after partial degradation; in such cases, synthesis of fatty acids from acetyl-CoA, a non-fatty acid precursor) of palmitic and stearic acids and the desaturation of stearic acid to oleic acid occur in plastids, while the conversion of oleic acid to linoleic and then α-linolenic occurs in endoplasmic reticulum (ER).[12,13]

In plastids, fatty acids are synthesized from acetyl-Coenzyme A (acetyl-CoA) in a three-step process: i) irreversible carboxylation of acetyl-CoA by the action of acetyl-CoA carboxylase to form malonyl-CoA. Subsequently, the malonyl group is transferred to acyl carrier protein (ACP) giving rise to malonyl ACP, the primary substrate of the fatty acid synthase complex (Figure 1A).[14] The formation of malonyl CoA, catalyzed by the highly regulated plastidic acetyl CoA carboxylase complex, is the committed step in fatty acid synthesis.[15]; ii) repeated condensation of malonyl-CoA with a growing ACP-bound acyl chain by action of the fatty acid synthase complex, with the consecutive addition of two carbon units for each elongation cycle to form 16:0-ACP (Figure 1A).[14] For each cycle, four separated reactions are necessary. The first step corresponds to the formation of 3-ketobutyl-ACP through the condensation of acetyl-CoA with malonyl-ACP by ketoacyl synthase III (KAS III), followed by reduction to 3-hydroxylacyl-ACP, dehydration to an enoyl-ACP and a second reduction to form the elongated 4:0-ACP. Subsequent rounds of condensation reactions of 4:0-ACP with malonyl-ACP giving rise 14:0-ACP and 16:0-ACP are catalyzed by KAS I enzyme.[16]; the elongation of 16:0-ACP to form 18:0-ACP, catalyzed by KAS II and the first desaturation step (that occurs in the plastid) where a Δ^9-desaturase is the enzyme responsible for the conversion of 18:0-ACP to 18:1-ACP (Figure 1A). These three fatty acids (16:0-ACP, 18:0-ACP and 18:1-ACP) are then exported to the cytosol into the acyl-CoA and acyl-lipid pools.[9,17] In some organisms, the

Correspondence: Márcia Pinheiro Margis, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Prédio 43312, 91501-970, Porto Alegre, Brasil.
Tel. +55.51.3308.9814.
E-mail: marcia.margis@ufrgs.br

Key words: fatty acid biosynthesis, TAG accumulation, lipid metabolism.

Acknowledgments: this project was supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), FAPERGS (Fundação de Amparo a Pesquisa do Estado do Rio Grande do Sul), FINEP (Financiadora de Projetos) and MCT (Ministério de Ciência e Tecnologia).

Contributions: AC is principal author in the conception, design, analysis and interpretation of data, and drafted the article. RM participated in the conception and design, and critically revised the article. FM, AT-Z and GL revised the article and contributed important intellectual content. MM-P participated in conception and design and final approval of the version to be published.

Received for publication: 28 March 2011.
Revision received: 14 July 2011.
Accepted for publication: 11 October 2011

International Journal of Plant Biology 2011; 2:e10
doi:10.4081/pb.2011.e10

elongation process can be extended and fatty acid chains containing up to 18 carbon atoms can be synthesized.[14]

The termination of fatty acid elongation is catalyzed by acyl ACP thioesterases (acyl-ACP hydrolases) enzymes. Thioesterases catalyzes the hydrolysis of acyl-ACP to produce free fatty acids, which are able to cross the plastidial envelope to be reactivated as acyl-CoAs on the outside of the organelle (Figure 1A).[18] In plants, two main thioesterase types were described: the FatA class, which preferentially removes oleate from ACP, and FatB thioesterases that are active with saturated and unsaturated acyl ACPs, and, in some species, with shorter-chain-length acyl ACPs.[19-21] The interplay between the fatty acid synthase complex, Δ^9 desaturase, and the two thioesterases determines the ratio of acyl chains produced in the plastid for TAG formation.[10]

After it has been exported from the plastids, oleic acid enters the cytosolic pool (Figure 1A) and is imported into the ER in association with CoA. Oleic acid is then available for conversion to linoleic and α-linolenic acid by the sequential action of substrate-specific desaturases.[13,14,17,22,23] Alternatively, other fatty acid chain modifications can occur in ER. For example, in Castor bean (*Ricinus communis* L.), oleic acid undergoes a hydroxylation process, through the activity of the oleate hydroxylase (FAH12) enzyme, yielding ricinoleic acid (C18:OH), an unusual hydroxylated (OH) fatty acid.[22] Finally, part of the fatty acids produced in plastids and modified in the ER is retained as structural components of cellular membranes (phospholipids of the ER and galactolipids of plastids, for instance) and the rest is transferred to TAG and accumulate as an energy source.[17]

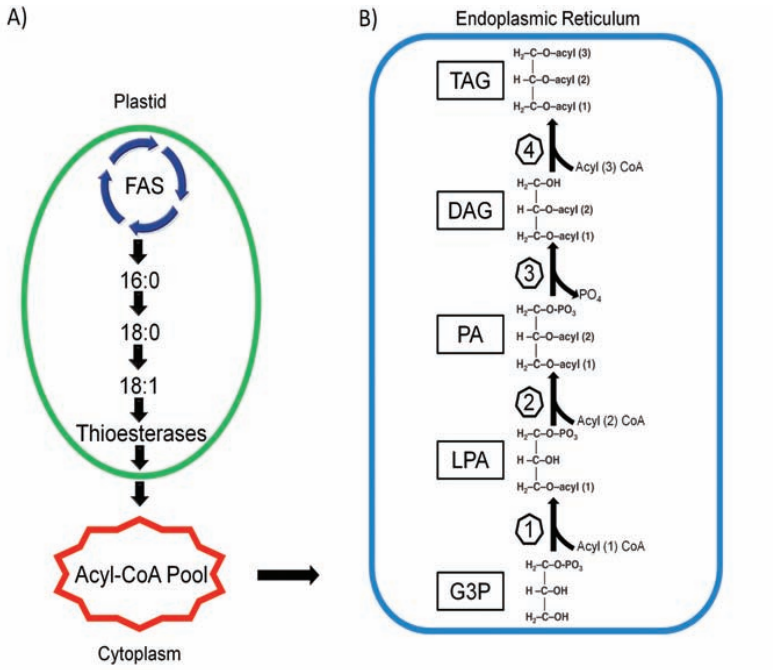


Figure 1. Fatty acid biosynthesis and Kennedy pathway for triacylglycerol biosynthesis in plants. (A) Fatty acid synthesis occurs in plastids through repeated steps of condensation by the action of several enzymes that compound the fatty acid synthase complex (FAS), which promotes the consecutive addition of two carbon units for each elongation cycle to form 16:0. Then, the elongation of 16:0 occurs to form 18:0 and the first desaturation step where 18:0 originates 18:1. The termination of fatty acid elongation is catalyzed by thioesterases enzymes and these three fatty acids are then exported to the cytosol into the acyl-CoA pools and are able to be accumulated as TAG molecules. (B) Kennedy pathway starts with acylation of glycerol-3-phosphate (G3P) to form lysophosphatidic acid (LPA) through the action of *sn*-glycerol-3-phosphate acyltransferase (1). The second acyl-CoA dependent acylation is catalyzed through the catalytic action of lysophosphatidic acid acyltransferase (2), leading to the formation of phosphatidic acid (PA). Phosphatidic acid phosphatase (3) catalyzes the release of phosphate from PA to produce DAG. The final acylation is driven by diacylglycerol acyltransferase (4) that converts DAG into TAG.

Unusual fatty acids

Differently from the conservative fatty acid composition observed in the plant membrane lipids, many evolutionarily divergent angiosperm species accumulate substantial amounts of *unusual* acyl chains in their seed-storage lipids. About 300 naturally occurring different fatty acids have been described in seed oils. However, it has been estimated that thousands more could be present throughout the plant kingdom.[10,11]

Unusual fatty acids correspond to those presenting chemical structures that deviate from the common fatty acids found in the majority of plant species. The structures of the unusual fatty acid can vary in chain length from 8 to 24 carbons. Alternatively, they can present modifications along acyl chain composition, like double bonds in unusual positions or novel functional groups (such as hydroxy, epoxy, cyclic, halogen or an acetylenic group).[11]

In several plant species, unusual fatty acids are the predominant fatty acids represented in the seed oil composition. The reason for such diversity observed in seed oil constituents is unknown, but plants seem to be able to tolerate high levels of unusual fatty acids in storage lipids because they are rapidly sequestered into oil bodies and, therefore, have no structural roles. The physical and chemical properties of many unusual fatty acids might explain, at least in part, why they are excluded from the membrane lipids of seeds, and are absent from other parts of the plant. It is hypothesized that they would disturb the structural integrity of the membrane bilayer and have deleterious effects on the cell. Consequently, storage and membrane lipids have different fatty acid compositions.[11]

Nature contains a wide variety of unusual fatty acids, some of which are important for industry and human health. Producing these unusual fatty acids in agronomical suitable plants has been a long standing goal for companies and researchers involved in the field of oilseed engineering.[24-27]

Saturated acyl chain fatty acids

Species of Araceae, Lauraceae, Lythraceae and Ulmaceae often contain saturated acyl chain lengths ranging from C8 to C14. Coconut oil, for example, possesses more than 90% of its acyl chains composed by saturated fatty acids, generally with fatty acids chain lengths ranging from C8 to C16, predominantly lauric acid (12:0). Another example includes Brassicaceae, which shows a preferential production of very long chain saturated fatty acids with carbon lengths of C20-C24.[10]

Medium-chain fatty acids

Some species of oleaginous plants are able to accumulate high amounts of medium-chain fatty acids, generally showing fatty acid chains with less than 16 carbons (Figure 2A).[17]

For common fatty acid formation, the grow-

ing acyl chain is terminated when it reaches a length of 16 or 18 carbons, by the action of chain-length-specific thioesterases,[28-30] which cleaves the acyl group from the ACP to produce a free fatty acid, thus terminating their elongation. Plants that synthesize medium-chain fatty acids have an additional acyl-ACP thioesterase enzyme, which is responsible for the premature cleavage of the acyl-chain from ACP, redirecting fatty acid synthesis from long (C16–C18) to medium (C8–C14) chains.[11]

Examples of the most commercially important medium-chain fatty acid oils include the palm kernel and coconut oils, which contain predominantly lauric acid (12:0) (Figure 2A) in their fatty acid compositions.[17] In addition, several species of the genus *Cuphea* are also known to accumulate high amounts of medium-chain fatty acids, ranging from C8 to C14, usually with only one chain length dominating in each species.[31]

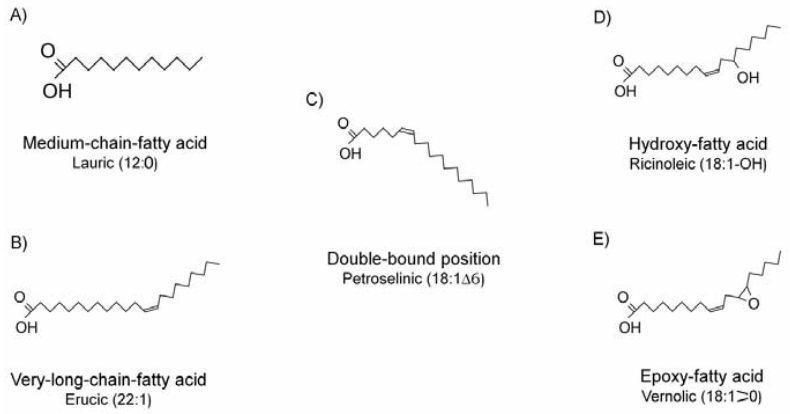


Figure 2. Examples of unusual fatty acids produced by plants. (A) Medium-chain-fatty acid. (B) Very-long-chain-fatty acid. (C) Unusual double-bound position fatty acid. (D) Hydroxy-fatty acid. (E) Epoxy-fatty acid.

Very long chain poly unsaturated fatty acids

Plant membrane sphingolipids contain significant amounts of very long chain FAs (VLC-FA) that possess fatty acid chain lengths ranging from 20 to 26 carbons (Figure 2B). Unsaturated VLC-FAs are also found in the storage oils of some plants (Cruciferae species) and in epicuticular and storage wax esters. They are synthesized outside of plastids by successive rounds of elongation of a C18 fatty acyl precursor by two carbons originating from malonyl-CoA[11] by a membrane-bound elongase complex.[32] In comparison, very long chain polyunsaturated fatty acids (VLC-PUFAs), defined as fatty acids containing 20 or more carbon atoms and three or more double bonds, are almost completely absent in higher plants.[33,34] The biosynthesis of VLC-PUFAs consists of cycles of alternated desaturation and chain elongation, in which the desaturation reactions occur on acyl-PC substrates and the elongation reactions occur on acyl-CoA.[17]

The primary biosynthesizing sources of VLC-PUFAs are marine microorganisms such as algae, which represent the base of an aquatic food network that results in the accumulation of VLC-PUFAs in fish oils.[35,36] However, some fungi and lower plants can synthesize VLC-PUFAs, and animals can convert dietary fatty acids such as linoleic and α-linolenic acids to these more complex forms.[26,34,37,38] From a nutritional point of view, the most important VLC-PUFAs are arachidonic (ARA; ω6-20:4$\Delta^{5,8,11,14}$), eicosapentaenoic (EPA; ω3-20:5$\Delta^{5,8,11,14,17}$), and docosahexaenoic acid (DHA; ω3 22:6$\Delta^{4,7,10,13,16,19}$). VLC-PUFAs are not only required as components of membrane phospholipids in specific tissues or as precursors for the synthesis of the different groups of eicosanoid effectors, but also contribute via a multiplicity of beneficial roles to the maintenance of good health, particularly by reducing the incidence of cardiovascular diseases.[39,40]

Novel monounsaturated fatty acids

The synthesis of common monounsaturated fatty acids is catalyzed by a soluble plastidial desaturase enzyme, which normally introduces a double bond between carbons 9 and 10 of a C18 acyl-ACP.[11] However, some plants that are able to synthesize unusual monounsaturated fatty acids present an additional desaturase enzyme, which is closely related to the Δ^9-desaturase, but introduces a double bond in positions other than the ninth carbon from the carboxyl group.[10]

Umbelliferae species, such as carrot and coriander, are known to contain oils rich in petroselenic acid (Δ^6 18:1) (Figure 2C). This unusual monounsaturated fatty acid is the result of the activity of a plastidial Δ^4 desaturase that introduces a double bond between carbons 4 and 5 of a C16 acyl-ACP, converting palmitoyl-ACP to Δ^4 hexadecanoyl-ACP which is then elongated to petroselinoyl-ACP and cleaved from ACP to produce the free fatty acid.[41]

Another curious example of this unusual desaturation process is observed in Meadowfoam (*Limnanthes alba*) oil. This specie accumulates oil with approximately 65% of 20:1 acid in its fatty acid composition. The 20:1 acid possesses a double bond at C5 carbon.[42]

Hydroxy, epoxy and acetylenic fatty acids

Fatty acids with additional functional groups in the acyl chain represent excellent feedstocks for industry and have been used to produce innumerous bio-based products due to their physical and chemical properties.[17]

The synthesis of the fatty acids containing functional groups such as hydroxyl, epoxy and acetylenic functional groups is achieved by the action of a family of related enzymes. Structurally, these enzymes present similarities to extraplastidial membrane-bound Δ^{12}-desaturases (FAD2), and only four amino acid substitutions are sufficient to convert an 18:1desaturase enzyme into an 18:1 hydroxylase enzyme.[43]

The synthesis of functional groups possessing fatty acids is thought to take place in the ER and uses as a substrate fatty acids esterified to the major membrane lipid phosphatidylcholine (PC).[11]

Castor oil, for example, is rich in hydroxylated fatty acid (OH) (Figure 2D), whereas some plants have epoxidated (Figure 2E) or methylated acyl chains in their TAGs.[10] Examples of plants that accumulate these kinds of fatty acids include *Vernonia galamensis, Euphorbia lagascae* and *Stokesia laevis* that accumulate 60-80% of an epoxy fatty acid known as vernolic acid (cis-12-epoxyoctadeca-*cis*-9-enoic acid).[43,44]

Triacylglycerols biosynthesis in plants

TAGs are present in all eukaryotes, including animals, plants, fungi and protists, and also in some prokaryotes such Actinomycetes,[45-47] *Streptomyces* species[48] and *mycobacteria* species.[49] Due to its importance as an inert storage component, the biosynthesis of TAGs is a common metabolism pathway that appears to be conserved from bacteria to humans.[50]

TAGs are quantitatively the most important seed storage reserve in many plant species including oil crops such as sunflower

(*Helianthus annuus*), oilseed rape (*Brassica napus*), soybean (*Glycine max*), and maize (*Zea mays*).[43,51] Besides their importance as energy source, TAGs also represent an important source for building blocks for membrane lipid biosynthesis.[50] TAGs from plants are important sources for human nutrition, also providing precursors for chemical industry products.[14,43] TAGs are becoming increasingly important raw materials for the production of paints, detergents, lubricants, biofuels and nylon precursors, and can also serve as renewable biofuels as an alternative to crude oil.[9]

TAGs accumulated during seed maturation are stored in the seed oil bodies until germination[51-53] and can either be used for energy production through β-oxidation or serve as substrates for acylation reactions, such as phospholipid biosynthesis, for example.[50]

Pathways for TAG biosynthesis

In seed plants, glycerolipids can be synthesized via two similar pathways that are known as prokaryotic and eukaryotic systems on the basis of their evolutionary origins. These two pathways occur in distinct subcellular compartments and are similar in the two-step enzymatic conversion of *sn*-glycerol-3-phosphate (G3P) to phosphatidic acid (PA), but differ in the subsequent conversion of PA into structural, storage and signaling lipids.[54] Although the two enzymatic pathways are similar, the enzymes that catalyze the acylations are unique to their respective system.[55] Plastidic enzymes are responsible for the prokaryotic system, while cytosolic enzymes are responsible for the eukaryotic system. In the prokaryotic pathway, fatty acids are directly transferred from ACP to G3P, while in the eukaryotic pathway fatty acids are cleaved from the ACP by an acyl-ACP-thioesterase to form free fatty acids that are exported to the cytoplasm, esterified to CoA and then join the acyl-CoA pool. The acyl groups are then used by the acyltransferases of the eukaryotic G3P pathway in the ER to produce membrane and storage lipids (TAGs).[11] Moreover, these enzymes can present structural differences depending on their subcellular localizations, forming independent clusters in phylogenetic studies.[56]

Production of storage lipids in plants involves *de novo* fatty acid synthesis in the stroma of plastids and subsequent incorporation of the fatty acid into glycerol backbone leading to TAG in the ER.[9]

Two metabolic pathways for the production of TAGs have been clarified: an acyl CoA-dependent pathway (Figure 1B) and an acyl-CoA-independent pathway (Figure 3).

Acyl CoA-dependent pathway

In the acyl-CoA dependent pathway, commonly known as the Kennedy pathway (Figure 1B), acyl-CoA is the substrate for successive acylation reactions of the glycerol backbone, with the terminal step being the acylation of *sn*-1,2- diacylglycerol (DAG) by DAG acyltransferases (DGATs).[57] The glycerol backbone for TAG assembly is derived from G3P which is produced via the catalytic action of *sn*-glycerol-3-phosphate dehydrogenase (G3PDH) from dihydroxyacetone phosphate (DHAP), derived from glycolysis.[18]

The Kennedy pathway starts with acylation of G3P to form lysophosphatidic acid (LPA) through the action of *sn*-glycerol-3-phosphate acyltransferase (G3PAT). The second acyl-CoA dependent acylation is catalyzed through the catalytic action of lysophosphatidic acid acyltransferase (LPAAT), leading to the formation of phosphatidic acid (PA). Phosphatidic acid phosphatase (PAP) catalyzes the release of phosphate from PA to produce DAG.[9] The final acylation is driven by DGAT, using acyl CoA as an acyl donor, converting DAG to TAG (Figure 1B).[14]

Glycerol-3-phosphate acyltransferase

Membrane-bound Glycerol-3-phosphate acyltransferase (G3PAT) is a soluble enzyme that initiates the fatty acid incorporation process by transferring fatty acids from either acyl-ACPs or acyl-CoA molecules to the sn-1 position of G3P, forming LPA.[10]

In several plants, two G3PAT isoforms have been found in the plastidial and cytoplasmic cellular compartments. G3PATs can either be selective, preferentially using oleic acid as the acyl donor, or non-selective, using either oleic or the saturated palmitic acid at comparable rates. This differential substrate specificity for saturated *versus* unsaturated fatty acids has been implicated in the sensitivity of plants to chilling temperatures.[58]

Lysophosphatidic acid acyltransferase

Lysophosphatidic acid acyltransferase (LPAAT) catalyzes the transfer of the acyl-chain from the CoA ester to *sn*-2 of LPA, leading to the formation of PA. There are two sequence-diverged gene subfamilies encoding LPAAT.[10]

LPAAT activity is associated, in plants, with multiple membrane systems, including chloroplasts, ER and the outer membrane of mitochondria, which suggests their presence on several different isoforms.[59,60] In plants, this enzyme shows a preference for unsaturated acyl chains,[10] but it is also able to discriminate acyl groups having longer or shorter fatty acid chain lengths. In developing seeds of certain plants, special LPAATs can incorporate unusual acyl groups, such as in Castor bean, which accumulates oils esterified with unusual fatty acids.[61]

Phosphatidic acid phosphatase

The cytoplasmic enzyme Phosphatidic acid phosphatase (PAP) dephosphorylates PA to yield DAG.[62] DAG produced from the hydrolysis of PA is not only a direct precursor of TAG, but also a substrate for the synthesis of membrane phospholipids.[63] In plants, two distinct PAP types involved in glycerolipid synthesis appear to exist in soluble and in membrane-associated fractions (microsomes). The distribution of these forms seems to be affected by the cellular metabolic status.[64]

PAP is also involved in general phospholipid degradation and turnover. It was demonstrated that PA accumulates in plants in a transient manner in response to various forms of stress. The dephosphorylation activity of PAP results in the attenuation of the signaling function of PA through its conversion to DAG, which can

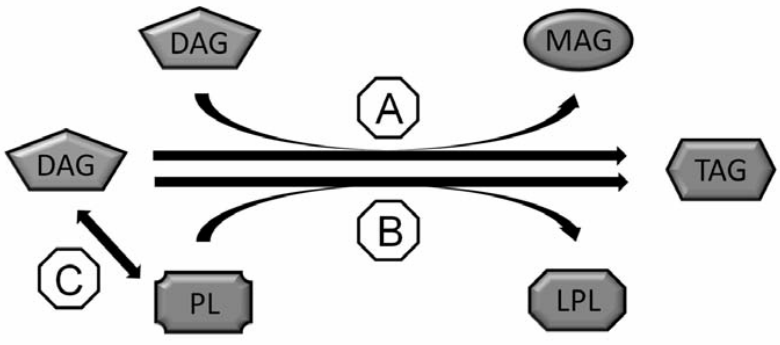


Figure 3. Acyl-CoA independent pathways for triacylglycerol (TAG) biosynthesis. Diacylglycerol transacylase (A) catalyzes the transfer of an acyl moiety between two diacylglicerol (DAG) molecules to form TAG and monoacylglycerol (MAG) as a co-product. Phospholipid: diacylglycrol acyltransferase (B) catalyzes the transfer of fatty acids from the sn-2 position of phospholipids (PL) to DAG to form TAG and lyso-phospholipids (LPL) as a co-product. Choline phosphotransferase (C) can contribute to TAG biosynthesis through a reversible conversion of PL into DAG. Then, the resulting DAG can be converted into TAG by the DGAT or PDAT enzymes.

be important in the stress responses, especially in the remodeling of membrane lipid composition.[64,65]

Diacylglycerol acyltransferase

Kennedy pathway's final enzyme, diacylglycerol acyltransferase (DGAT), catalyzes the third acyl-CoA-dependent acylation reaction that leads to the production of TAG from DAG. DGAT is an integral ER protein and has also been shown to be present in oil bodies and plastids.[66] Although DGAT1 and DGAT2 are the main types of DGAT enzymes, other different, structurally unrelated enzymes with DGAT activity have been described in plants.[67,68] A soluble DGAT from peanuts,[69] a wax ester synthase/acyl-coenzyme A: Diacylglycerol acyltransferase (WSD1) from Arabidopsis[43] and a distinct DGAT (DAcT) from *Euonymus alatus*.[70] The substrate selectivity of DGAT depends on several factors, such as the acyl composition of the DAG pool, acyl-CoA concentration and temperature.[9]

DGAT1 was initially cloned from mouse based on its homology with mammalian acyl-CoA: cholesterol acyltransferase genes.[71] Several homologs of DGAT1 have been cloned and characterized in animals and plants, and their functions have been verified by both overexpression and deletion approaches.[72-75]

A second family of DGAT genes (DGAT2), which have no sequence similarity with DGAT1, were first identified in the oleaginous fungus *Morteriella ramanniana*.[76,77] Several DGAT2 genes have been cloned and characterized from animals, fungi and plants.[77-79] In animals, DGAT2 has different physiological functions *in vivo* and presents a different temporal-spatial expression profile compared to DGAT1.[80-82] In yeast, *Sacharomyces cerevisae*, the DGAT2 enzyme is dominant in the stationary growth phase when the yeast is storing significant amounts of TAG.[83,84] In addition, some experiments suggest a more important role of DGAT2 expression in the accumulation of conjugated and hydroxy fatty acid in seed oils.[43,85-87]

The roles of DGAT1 and DGAT2 in the oil production are apparently species-dependent. In plants, DGAT1 appears to be a major enzyme gene for seed oil accumulation, while DGAT2 appears to play significant roles in the selective accumulation of unusual fatty acids, such as epoxy and hydroxy fatty acid, into seed storage oils.[43,86,88]

A third enzyme is a soluble DGAT (DGAT3), which was only recently identified in peanut and other plant species.[67,86] However, there is little information about this new DGAT isoform. An acyl-CoA-dependent acyltransferase, namely wax ester synthase/diacylglycerol acyltransferase (WS/DGAT), was identified and purified from the bacterium *Acinetobacter* sp. strain ADP1, which can utilize both fatty alcohols and DAG as acyl acceptors to synthesize wax esters and TAGs, respectively.[88-90] A large number of genes with homology to this *Acinetobacter* gene[79,91] were identified in Arabidopsis.

DGATs may be one of the rate-limiting steps in plant storage lipid accumulation,[9,88,92] and thus appear to be crucial for mediating quantitative and qualitative aspects of seed oil synthesis in transgenic plants.[67] In this scenario DGAT seems to be a potential target for the genetic modification of plant lipid biosynthesis in oilseeds for economic benefit.[88] Studies in this field demonstrated that overexpression of DGAT1 resulted in increase,[92] whereas suppression of DGAT activity resulted in a decrease in oil content in Arabidopsis seeds.[75,93] In addition, DGAT activity seems to be also important for the correct channeling of unusual fatty acids into seed storage oils.[67]

Acyl CoA-independent pathway

As an alternative to the Kennedy pathway, nascent fatty acids may be first incorporated into membrane lipids at the plastid envelope and/or in the ER and afterwards accumulated as TAG molecules.[14]

Newly synthesized fatty acids can be incorporated directly into PC via an acyl editing mechanism, rather than through PA and DAG intermediates.[94] Acyl chains from PC can be incorporated into TAG, either through conversion back to DAG or by the action of a phospholipid: diacylglycerol acyltransferase (Figure 3) (PDAT).[14] This enzyme is a member of the lecithin: cholesterol acyltransferase gene family[91] that catalyzes the formation of TAG by an acyl transfer from the *sn*-2 position of phospholipids to DAG (Figure 3), with phosphatidylethanolamine (PE) as the preferred acyl donor in both yeast and plants.[95-97]

PDAT activity has been reported in yeast microsomes and certain oilseeds.[91,96,98] It was demonstrated that Arabidopsis PDAT is able to utilize different phospholipids as acyl donors and accept acyl groups of chain lengths ranging from C10 to C22.[43,97] In yeast, PDAT1 is a major contributor to TAG accumulation during the exponential growth phase.[76,83]

The activity of PDAT enzymes (specially in oilseed plants) may play a critical role in the removal of unusual fatty acids from membrane phospholipids and transfer it into TAG molecules.[9] It was suggested that the balance of the expression of PDAT genes may represent an important mechanism either for the maintenance of membrane lipid homeostasis, contributing to membrane lipid turnover or for the removal of DAG, an effector molecule of the phosphatidylinositol signaling pathway that possesses a critical role for plant stress responses.[91]

In another case of acyl-CoA-independent transacylation, it has been postulated that a DAG transacylase (DGTA) catalyzes the transfer of an acyl moiety between two DAG molecules to form TAG, and monoacylglycerol (MAG) as a co-product (Figure 3),[85,99] and that the reverse reaction participates in the remodeling of TAGs.[88,100] However, no gene encoding such as transacylase has been identified so far.[85]

There is growing evidence for the presence of an alternative acyl-independent pathway for TAG formation in plants, involving an enzyme that is normally related with membrane biosynthesis, named choline phosphotransferase (CPT) or amino-alcohol phosphotransferase (AAPT).[10] CPT can contribute to TAG biosynthesis through a reversible conversion of PC to DAG (Figure 3).[12] Afterwards, the resulting DAG can be converted into TAG by DGAT or PDAT enzymes. It has been hypothesized that the continuous reversible transfer of DAG into PC may control, in part, the PUFA content of the seed oil,[101] making CPT the key enzyme regulating the route by which PUFAs are available for incorporation into TAG molecules.[22,101,102]

In addition, a recent publication demonstrated that a class of phospholipase enzyme (phospholipase D) plays an important role in the conversion of PC into TAG.[103] The attenuation by RNA interference of a soybean phospholipase D was responsible for higher levels of di18: 2 (dilinoleoyl)-PC and PE in seeds compared to the wild type lines. The increased polyunsaturation was at the expense of PC and PE species containing monounsaturated or saturated fatty acids. By contrast, a decrease in the unsaturation of the TAG fraction of the soybean seeds was observed, suggesting that phospholipase D suppression slows the conversion of PC into TAG.[103]

Regulation of oil biosynthesis in plants

The accumulation at high levels of compounds for nutrient storage is a characteristic event of seed development which is regulated by a common, at least in part, genetic program that takes place during the seed maturation phase.[104-108] The regulation of oil synthesis occurs at multiple levels.[109] Several of the enzymes involved in the synthesis, accumulation and degradation of neutral lipids have been identified and a great redundancy for most of the neutral lipid metabolic enzymes was observed. The proteins involved in neutral lipid metabolism are well conserved across species, exhibiting remarkable homology to each other.[50]

Synthesis and accumulation of storage com-

pounds are regulated by numerous transcription factors (TFs) in an intricate network involving genetic programs, and hormonal and metabolic signals.[110-112] Some of these TFs are known as master regulators based on their apparent capacity to regulate the action of other TFs.[18] The most important master regulators of seed maturation and storage accumulation include the LEAFY COTYLEDON genes (LEC1 and LEC2), FUSCA3 (FUS3), ABSCISIC ACID INSENTIVE3 (ABI3) and WRINKLED1 (WRI1).[18,111]

Among the known transcriptional factors that regulate genes related to storage compounds, it was demonstrated that WRI1 activity was responsible for the regulation of oil accumulation by promoting the carbon flux through the glycolysis pathway.[113-116]

WRI1 encodes a transcription factor of the APETALA2-ethylene responsive element-binding protein (AP2-EREBP) family.[114] This TF is responsible for specifying the regulatory action of other TFs, such as LEC2 and possibly LEC1, during the fatty acid biosynthetic network.[110,117]

Studies in Arabidopsis demonstrated that WRI1 up-regulates glycolytic gene expression required for the conversion of sucrose to TAG biosynthesis precursors.[112,114] The corresponding *wri1* mutant is deficient in oil biosynthesis and shows a reduction of 80% in seed oil content.[14,115] The carbohydrate metabolism regulation in *wri1* appears to be affected because the activities of a number of key glycolytic enzymes were highly reduced in that mutant.[18]

Other transcription factors are also involved with the regulation of oil metabolism in developing oil seeds.[14] It was shown that seed-specific overexpression of Arabidopsis LEC1 in developing seedlings of Arabidopsis, using an estradiol-inducible vector, up-regulates the transcription level of known enzyme-coding genes of fatty acid biosynthesis by about 60%.[118] The same study showed that the amount of fatty acids was 4.7 times higher in transgenic than wild-type seedlings and that the function of LEC1 in fatty acid biosynthesis regulation was partially dependent on ABI3, FUS3 and WRI1.[18,118]

FUS3 also seems to play an important role in oil accumulation in Arabidopsis. Transcriptomic analysis has revealed that the abundance of FUS3 transcript increases together with transcripts involved with fatty acid pathway.[113] In addition, the inducible expression of Arabidopsis FUS3 resulted in rapid induction of gene expression associated with FA biosynthesis in seedlings. The same results were also observed in Arabidopsis protoplasts transiently expressing FUS3.[18]

Soybean DNA binding proteins with one zinc-finger motif, or Dof-type TFs, have also been shown to have effects on oil accumulation. Overexpression of GmDof4 and GmDof11 leads to an increased expression level of the genes encoding the β-subunit of acetyl CoA carboxylase (ACCase) and long-chain acyl-CoA synthetase, respectively, both encoding enzymes involved with fatty acid biosynthesis. In addition, the transgenic lines with the highest levels of GmDof4 or GmDof11 expression also present a lipid content increase ranging from 11 to 24%.[18,119]

However, besides the fatty acid biosynthesis regulation promoted by TFs, additional levels of control certainly involve allosteric enzyme regulation, for example, at the level of ACCase[15,120] or plastid pyruvate kinase,[14,121,122] important precursors of TAG biosynthesis.

Oil bodies

Several organisms store lipids in subcellular particles as food reserves, which will be used through a period of active metabolism. These lipid particles can be found in seeds, pollens, flowers, roots, stems of flowering plants, spores and vegetative organs of non-flowering plants and algae. These structures are also represented in some animal cells, besides fungi and *Euglena*. However, seed subcellular storage lipid particles, have been studied more extensively.[123]

Seeds of most plant species store TAGs as food reserves for germination and post-germinative growth.[123] The intracellular storage of neutral lipids occurs in specialized compartments called lipid particles, lipid droplets or oil bodies.[50]

Oil bodies are relatively simple spherical organelles with approximately 1 µm in diameter that arise from the ER, the site of TAG synthesis, and are surrounded by a phospholipid monolayer membrane embedded with proteins called oleosins.[51,124,125] Oleosins in the seeds of diverse species are small proteins of about 15 to 26 Kilodaltons[123,126] that are usually present as two or more highly conserved isoforms.[51,127]

According to the best accepted oil body biogenesis model, proteins involved in neutral lipid metabolism accumulate in certain regions of the ER. Enzymes involved in TAGs formation are found among these polypeptides.[50] TAGs are synthesized in the ER and are sequestered, due to their hydrophobicity, between the two layers of the ER membrane.[123] Because newly formed neutral lipids are unable to integrate into bilayer membranes they cluster and accumulate in the hydrophobic region between the two leaflets of the ER membrane (Figure 4A and B). During ongoing TAG synthesis the droplet grows and forms a bud (Figure 4C). After reaching a certain size, the budding particle, which has a TAG matrix surrounded by a layer of phospholipids and oleosins, is released into the cytosol as a mature oil body (Figure 4D).[50,123]

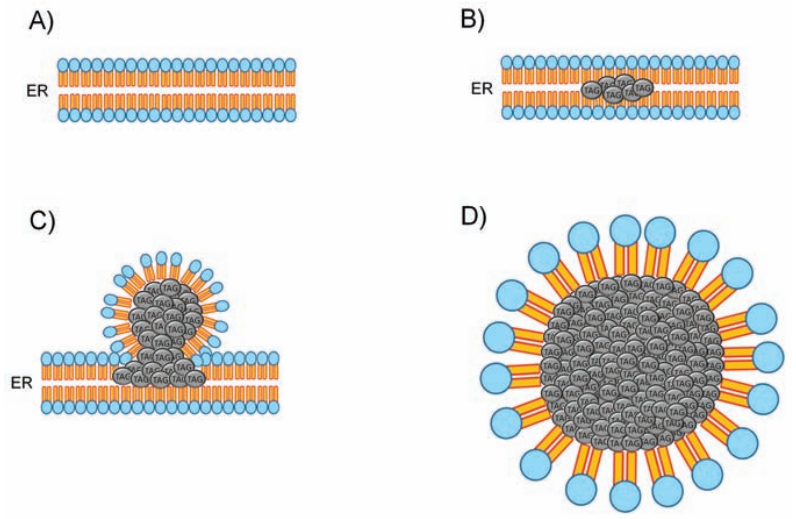


Figure 4. Oil body biosynthesis. (A-B) Accumulation of triacylglycerols (TAGs) occurs between the two layers of endoplasmic reticulum membrane (ER). (C) During ongoing synthesis of TAGs the droplet grows and forms a bud. (D) After reaching a certain size, the budding particle is released into the cytosol as a mature oil body. Adapted from Athenstaedt and Daum.[50]

Lipid biosynthesis in algae

Eukaryotic algae represent a very diverse organism group that is considered a key component for the most diverse ecosystems. They account for over half of the primary production at the base of food chains.[128]

Their variable morphology and habitats mean that eukaryotic algae contain a diverse composition of acyl lipids and unusual fatty acids that are not found in other phyla.[128] It is widely accepted that the excellent capability of algae to adapt to very different environmental conditions results from the fact that they can synthesize a number of unusual compounds, responsible for their unusual pattern of cellular lipids, as well as their ability to modify efficiently the lipid metabolism in response to environmental conditional changes.[129,130] Indeed, some authors suggested that the capacity of some algae to store VLC-PUFAs in TAG could be a reserve, which would allow the organisms to adapt to any further rapid changes in their environment, such as high light intensity, UV radiation and low temperature.[131]

Based on their lipid diversity, several thousands of algae and cyanobacterial species have been screened for high lipid content, resulting in the isolation and characterization of several hundred oleaginous species. These species can be found among diverse taxonomic groups, and the total lipid content may vary noticeably among individual species or strains, within and between taxonomic groups.[132]

Comparison of lipid metabolism in algae and higher plants

Lipid metabolism, specially the pathways to fatty acids and TAG biosynthesis, is less understood in algae than in higher plants. It is generally accepted that the basic pathways of fatty acid and TAG biosynthesis in algae are directly analogous to those demonstrated in higher plants based on the sequence homology and some shared biochemical characteristics of a number of genes and/or enzymes involved in lipid metabolism.[132]

However, there is some evidence of differences in algae lipid metabolism. Unlike higher plants, where individual classes of lipids may be synthesized and localized in a specific cell, tissue or organs (seeds or fruits), the complete pathway from carbon dioxide fixation to TAG synthesis and sequestration takes place within a single algal cell.[132] After being synthesized, the accumulation of algal TAGs occurs in densely packed lipid bodies located in the cytoplasm of the algal cell, although the lipid body formation and accumulation might also occur in the inter-thylakoid space of the chloroplast in some green algae species, such as *Dunaliella bardawil*.[133] Higher plants cannot synthesize significant amounts of VLC-PUFAs above C18, whereas many algae, especially marine species, possess the ability to synthesize and accumulate large quantities of VLC-PUFAs, such as EPA, DHA and ARA.[134] In addition, it is hypothesized that in algae, the TAG biosynthesis pathway may play a more active role in the stress response, in addition to functioning as carbon and energy storage under environmental stress conditions.[132]

Fatty acid biosynthesis in algae

Algae synthesize fatty acids as building blocks for the formation of various types of lipids. Similar to higher plants, the most commonly synthesized fatty acids have chain lengths ranging from C16 to C18 (Table 1).[136] In general, saturated and mono-unsaturated fatty acids are predominant in most algae examined. Indeed, the major saturated fatty acid is palmitic acid, while oleic acid is much less abundant than in higher plants.[132]

However, some algae and cyanobacteria are able to synthesize, as predominant FA species, medium-chain fatty acids (C10, C12 and C14), whereas others are able to produce VLC-FA (>C20). For instance, 27-50% of all fatty acids found in the filamentous cyanobacterium *Trichodesmium erythraeum* are composed by a C10 fatty acid and up to nearly 70% of all fatty acids in the golden alga *Prymnesium parvum* are composed of C14 fatty acid.[137] The distribution of individual fatty acids is quite distinct and tightly regulated. This control relates primarily to the function of acyl lipids in the membrane composition.[128]

Another distinguishing feature of some algae is the large amounts of VLC- PUFAs (>18C) as their major fatty acid components (Figure 5, Table 1), which resulted from several desaturation/elongation steps, promoted by

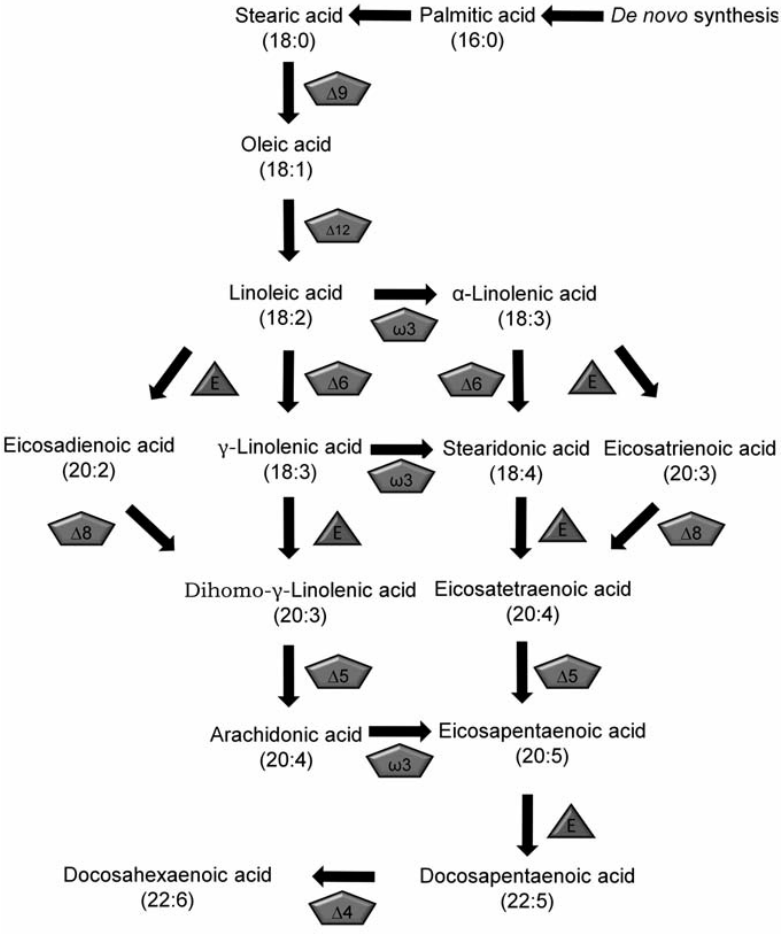


Figure 5. Very-long-chain polyunsaturated fatty acid (VLC-PUFA) biosynthesis in eukaryotic algae. Pentagons and triangles represent desaturases and elongases enzymes involved in algal VLC-PUFA biosynthesis, respectively. Adapted from Harwood and Guschina.[128]

large classes of desaturase and/or elongase enzymes (Figure 5). This is particularly true for marine species. Examples include the green alga *Parietochloris incise*,[131] the diatom *Phaeodactylum tricornutum* and the dinoflagellate *Crypthecodinium cohnii*[138] where the VLC-PUFAs ARA, EPA or DHA constituted the major fatty acid species representing about 33.6-42.5%, approximately 30% and 30-50% of the total fatty acid composition of these three species, respectively. In this scenario, the presence of 20% or more of three kinds of fatty acids in algae is attracting much economical interest because of the role of algae at the beginning of food chains and the perceived need for VLC-PUFAs in healthy diets.[128]

In addition, and alternatively to PUFAs, certain algae species may contain other unusual lipids in their oil composition, including chlorosulpholipids[139] and halogenated fatty acids.[140]

Table 1. Major fatty acid and polyunsaturated fatty acid (PUFA) in algae.

Algal Class	Major Fatty Acid	Major PUFAs
Bacillariophyceae	C16:0 and C16:1	C20:5 and C22:6
Chlorophyceae	C16:0 and C18:1	C18:2 and C18:3
Euglenophyceae	C16:0 and C18:1	C18:2 and C18:3
Chrysophyceae	C16:0 and C16:1 and C18:1	C20:5, C22:5 and C22:6
Chryotophyceae	C16:0 and C20:1	C18:3, C18:4 and C20:5
Eustigmatophyceae	C16:0 and C18:1	C20:3 and C20:4
Prasinophyceae	C16:0 and C18:1	C18:3 and C20:5
Dinophyceae	C16:0	C18:5 and C22:6
Prymnesiophyceae	C16:0 and C16:1 and C18:1	C18:2, C18:3 and C22:6
Rhodophyceae	C16:0	C18:2 and C20:5
Xanthophyceae	C14:0, C16:0 and C16:1	C 16:3 and C20:5
Cianobacteria	C16:0 and C16:1 and C18:1	C16:0, C18:2 and C18:3

C14, Myristic acid; C16:0, Palmitic acid; C:16:1, Palmitoleic acid; C16:3, Hexadecatrienoic acid; C18:1, Oleic acid; C18:2, Linoleic acid; C18:3, Linolenic acid; C18:4, Parinaric acid; C18:5, Octadecatetraenoic acid; C20:1, Eicosenoic acid; C20:3, Dihomo-gamma-linolenic acid; C20:4, Arachidonic acid; C20:5, Eicosapentaenoic acid; C22:5, Docosapentaenoic acid; C22:6, Docosahexaenoic acid. Adapted from Cobelas. [135]

Triacylglycerols accumulation in algae

TAG biosynthesis in algae may occur via the direct glycerol pathway (Kennedy pathway).[141] The fatty acids that are produced in the chloroplast are sequentially transferred from CoA to positions 1 and 2 of G3P, by GPAT and LPAAT enzymes, respectively, forming PA.[136] PA is then dephosphorylated by a specific action of PAP enzyme releasing DAG. In the final step of TAG synthesis, a third fatty acid is transferred to the vacant position 3 of DAG, by the action of DGAT. As in plants, the acyltransferases involved in the TAG synthesis process seem to exhibit preferences for specific acyl-CoA, and thus may play an important role in determining the final acyl composition of TAG molecules.[132]

Beyond the Kennedy pathway, the existence was suggested of an acyl-CoA-independent synthesis of TAG in algae, through PDAT and CPT activities, similar to that observed in plant and yeast. This pathway could play an important role in the regulation of the membrane lipid composition in response to various environmental and growth conditions, since under various stress conditions, algae usually undergo rapid degradation of the photosynthetic membrane with concomitant occurrence and accumulation of cytosolic TAG-enriched lipid bodies.[132]

Despite the occurrence and the levels of algal TAG production appear to be species/strain-specific, oleaginous algae produce only small quantities of TAG under optimal growth or favorable environmental conditions.[142] Under optimal conditions of growth, algae are able to synthesize fatty acids mainly for esterification into glycerol-based membrane lipids, which represent about 5-20% of their dry cell weight (DCW). The membrane lipid fatty acid composition includes medium-chain (C10-C14), long-chain (C16-18) and very-long-chain (>C20) fatty acid derivatives.[143]

In contrast, under unfavorable environmental or stress conditions, many algae species promote a shift in lipid metabolism from membrane lipid synthesis to the synthesis and storage of neutral lipids, especially in the form of TAGs. The *de novo* biosynthesis and conversion of certain membrane polar lipids into TAGs may contribute to the overall increase in TAG content. As a result, TAGs may account for as much as 80% of the total lipid content in the cell. As an example of the shift in the lipid metabolism, many microalgae have the ability to produce substantial amounts (20-50% DCW) of TAGs as lipid storage under photo-oxidative stress or other adverse environmental conditions.[144]

The major chemical stimuli to TAG accumulation in algae are nutrient starvation, salinity and growth-medium pH. On the other hand, the major physical stimuli are temperature and light intensity. In addition to chemical and physical factors, growth phase and/or aging of the culture also affects the content and fatty acid composition of TAG molecules.[132]

A significant variation in the algal lipid content and fatty acid profile in response to different growth conditions was observed.[145-147] In *Pavlova lutheri*, a marine Pavlovophyceae, it was reported that the proportions of PUFAs, especially EPA, were significantly higher under low light. By contrast, content of saturated fatty acids and DHA were significantly higher under strong light.[148] In addition, these authors also demonstrated that the growth and lipid composition presented a higher sensibility to variations in light intensity than in carbon source.

Another example of light effect in lipid and fatty acid composition includes *Chlorella zofingiensis*, a green algae that can grow well photoautotrophically as well as heterotrophically. It presented a 900% increase in lipid yield in heterotrophic cells compared with photoautotrophic cell culture. Moreover, about 80% of total lipid content was represented by neutral lipids in heterotrophic cells, with 88.7% being TAGs. On the other hand, photoautotrophic cells accumulated mainly membrane lipids (glycolipids and phospholipids).[149]

Guihéneuf and co-workers[150] investigated the effect of UV radiation (UV-R) on the lipid composition of two marine microalgae, *Pavlova lutheri* and *Odontella aurita*. The results indicated that the exposure to UV-R treatment led to a decrease in the proportions of PUFAs especially into structural lipids (glycolipids and phospholipids) in *P. lutheri*, whereas in *O. aurita*, exposure to UV-R did not change the fatty acid composition and lipid fractions of the cells, suggesting that this species is more resistant and seems to be able to partially acclimate to UV-R.[150]

Aplications of Triacylglycerols produced by plants and algae

Biodiesel

Biodiesel is a clean-burning fuel derived from vegetable oils or animal fats, which has been used as alternative to diesel fuel.[151] TAGs are the main components of the vegetable oils and animal fats that are used for biodiesel production.[152]

TAGs are the most similar chemical to fossil oil and, therefore, is the best potential replacement for the chemical industry. In fact, fossil oil is derived from ancient lipid-rich organic material, such as spores and planktonic algae that were sedimented and transformed under

high pressure and temperature over millions of years.[153] TAGs consist of several different fatty acids which present different physical and chemical properties. Therefore, the composition of these fatty acids will be the most important parameters influencing the corresponding properties of the biodiesel derived from TAGs.[152]

The advantages of the use of biodiesel include its higher oxygenated state compared to the conventional diesel, which leads to lower carbon monoxide (CO) production and reduced emission of particulate matter, and the few or lack of sulfur or aromatic compounds in its composition. These compounds are present in the conventional diesel and contribute to sulfur oxide and sulfuric acid formation, while aromatic compounds also increase particulate emissions and are considered carcinogens.[14] Furthermore, the use of biodiesel confers additional advantages, including a higher flashpoint (allows safer handling and storage), faster biodegradation (particularly advantageous in environmentally sensitive areas where fuel leakage poses great hazards) and greater lubricity.[14] However, the biggest environmental advantage of using biodiesel is that it is a renewable energy source, since the reduced hydrocarbon chains of biodiesel are derived from solar energy: plants and algae capture light energy during photosynthesis, converting carbon dioxide and water to the sugars, from which TAGs are derived.[154]

As a result of their rapid growth and substantial production of TAGs, algae could be employed as cell factories to produce oils and other lipids for biofuels and other biomaterials, such as biodiesel, methane, hydrogen, ethanol, among others.[132] In this context, biofuel production using microalgae offers several advantages in relation to oilseed crops: a) the high growth rate and the high photosynthetic conversion efficiency of microalgae makes it possible to satisfy the demand for biofuels using limited land resources; b) the cultivation of microalgae consumes less water than land crops; c) microalgae thrive in saline/brackish water/coastal seawater for which there are few competing demands; d) microalgae are able to synthesize and to accumulate large quantities of neutral lipids, mainly TAGs; e) microalgae can utilize nutrients such as nitrogen and phosphorus from a variety of wastewater sources, providing additional benefits through wastewater bio-remediation; and f) are capable of tolerating high CO_2 content in gas streams allowing high-efficiency of CO_2 mitigation[155,156] For these reasons, microalgae are capable of producing more oil per unit area of land, compared to terrestrial oilseed crops.[157]

On the other hand, one of the major drawbacks of microalgae for biofuel production is the low biomass concentration in the microalgal culture due to the limit of light penetration, which in combination with the small size of algal cells makes the harvest of algal biomasses relatively costly.[155]

The main limitation for the replacement of conventional diesel by biodiesel is that biodiesel still represents a small percentage of total diesel consumption, despite the large increase in its use. Two main factors have contributed to the limited adoption of biodiesel: problems with the fuel characteristics of biodiesel, namely poor cold-temperature properties, higher rates of oxidation and increased emission of nitrogen oxides (NOx) relative to the conventional diesel and the interrelated factors of cost and supply limitations, since the total world plant oil production, for example, would only satisfy approximately 80% of USA diesel demand.[14]

Genetic manipulation of fatty acid content

Modification of oil composition could contribute to the production of nutritionally and industrially desirable oils in crop plants and algae.[64] The increasing global demand for oils has intensified the research efforts to genetically modify the organism to boost the oil yield.[69] However, this requires not only a precise manipulation of fatty acid but also TAG synthesis in such a way that a specific synthesized fatty acid will be effectively incorporated into each position of TAG molecule.[61]

The natural diversity observed in seed and algae TAGs indicates that there should be no barriers to produce exotic FAs in domesticated oilseed crops and algae, and also provides a deep and potentially useful gene pool for genetic manipulation.[10]

Several unusual fatty acids described in this review, due to their unique chemical properties, have important industrial applications. However, to make these economically attractive, conventional oilseed crops will have to be genetically engineered to produce oils with a single predominant unusual fatty acid in its oil composition. The existence of wild species, such as castor, which contains oil with 90% hydroxylated fatty acid, suggests that this is a feasible task.[11] Thus, attractive targets for plant genetic engineering for altered TAG composition are the temperate oilseed crops, such as soybean, rapeseed, flax, and sunflower.[10]

Today, the generation of transgenic crop plants engineered to accumulate high levels of specific unusual fatty acids is a topic of enormous interest. However, the inability to specifically target unusual fatty acids to seed TAGs, and their excessive accumulation in membrane lipids, might disrupt seed membrane integrity and impair seed development or germination. To overcome this problem it is essential to prevent the accumulation of unusual fatty acids in seed membrane lipids through engineering projects aimed at generating viable, high-yielding transgenic plants by introducing genes involved in the exclusion of unusual fatty acids from membranes.[11]

Oilseeds provide an attractive platform for the production of high-value fatty acids that can replace non-sustainable petroleum special chemicals, such as diesel.[67] However, in the case of the production of fatty acids in transgenic plants, the conversion of plant oils into biodiesel and chemical feedstocks competes with their use as food. Therefore, it is extremely important to ensure that these new industrial productions do not affect the supply of food at a time when world food requirements are increasing rapidly.[14,110]

On the other hand, the production of PUFAs by marine and freshwater microalgae is the subject of intensive research and increasing commercial attention.[128] In this scenario, metabolic engineering through genetic manipulation might represent a promising strategy for the overproduction of algal oils.[129]

The focus of attention is the isolation and use of novel fatty acid metabolic genes, especially elongases and desaturases (Figure 5) that are responsible for PUFAs synthesis, from different species of algae and their transfer to plants in order to modify crops to create useful new products through genetic engeneering tools.[26] One example of this tendency is the metabolic engineering of omega-3 long-chain polyunsaturated fatty acids in plants using an acyl-CoA Δ6-desaturase from the marine microalgae *Micromonas pusilla*, which originated an artificial pathway that produced 26% EPA in *Nicotiana benthamiana* leaves.[158] The authors also demonstrated that this enzyme appears to function as an acyl-CoA desaturase that has preference for ω3 substrates, both in Arabidopsis and in yeast transgenic lines.

In this context, the identification and functional characterization of enzymes involved in biosynthesis of long-chain PUFAs in algae has an important biotechnological significance for the production of VLC-PUFAs in transgenic oilseed crops.[159,161]

In addition, it is believed that the most costly downstream processing steps in fuel production using microalgal is that most microalgae will not grow to a density higher than a few grams of biomass per liter of water. While there are several possible low-cost solutions to concentrate the biomass, these methods are slow and the resulting biomass may still require further dewatering. One possible solution is to manipulate the biology of microalgal cells to allow the secretion of fuels or feedstocks directly into the growth medium, through the manipulation of lipid secretory pathway.[156]

In conclusion, the complete clarification of enzyme activities responsible for controlling the flux of fatty acids to TAG will make a decisive contribution to the correct transfer of unusual fatty acids into storage oils and can generate new tools for genetic oil manipulation.[39,153] In this scenario, TFs regulatory action can ultimately affect several reactions in biochemical pathways that contribute to the production and accumulation of storage compounds. Thus, modification of the expression of genes encoding TFs represents another interesting strategy that can be adopted in order to increase the accumulation of desirable oils in target organisms.[18]

References

1. Dormann P, Benning C. Galactolipids rule in seed plants. Trends in Plant Science 2002; 7:112-8.
2. Bertrand M. Carotenoid biosynthesis in diatoms. Photosynth Res 2010;106:89-102.
3. Lemoine Y, Schoefs B. Secondary ketocarotenoid astaxanthin biosynthesis in algae: a multifunctional response to stress. Photosynth Res 2010;106:155-77.
4. Wang XM. Plant phospholipases. Annual Review of Plant Physiology and Plant Molecular Biology 2001;52:211-+.
5. Meijer HJG, Munnik T. Phospholipid-based signaling in plants. Annual Review of Plant Biology 2003;54:265-306.
6. Dyas L, Goad LJ. STERYL FATTY ACYL ESTERS IN PLANTS (VOL 34, PG 17, 1993). Phytochemistry 1993;34:1663-3.
7. Pata, MO, Hannun YA, Ng CKY. Plant sphingolipids: decoding the enigma of the Sphinx. New Phytologist 2010;185:611-30.
8. Dunn TM, Lynch DV, Michaelson LV, et al. A post-genomic approach to understanding sphingolipid metabolism in Arabidopsis thaliana. Annals of Botany 2004;93:483-97.
9. Lung SC, Weselake RJ. Diacylglycerol acyltransferase: a key mediator of plant triacylglycerol synthesis. Lipids 2006;41:1073-88.
10. Voelker T, Kinney AJ. Variations in the Biosynthesis of Seed-Storage Lipids. Annu Rev Plant Physiol Plant Mol Biol 2001;52:335-61.
11. Millar AA, Smith MA, Kunst L. All fatty acids are not equal: discrimination in plant membrane lipids. Trends Plant Sci 2000;5:95-101.
12. Somerville C, Browse J. Plant Lipids: Metabolism, Mutants, and Membranes. Science 1991;252:80-7.
13. Stymne SS. Triacylglycerol biosynthesis. In: The Biochemistry of Plants: A Comprehensive Treatise. Academic Press, Orlando, USA, 1987, pp. 175-214.
14. Durrett TP, Benning C, Ohlrogge J. Plant triacylglycerols as feedstocks for the production of biofuels. Plant J 2008;54:593-607.
15. Nikolau BJ, Ohlrogge JB, Wurtele ES. Plant biotin-containing carboxylases. Arch Biochem Biophys 2003;414:211-22.
16. Tai H, Jaworski JG. 3-Ketoacyl-acyl carrier protein synthase III from spinach (Spinacia oleracea) is not similar to other condensing enzymes of fatty acid synthase. Plant Physiol 1993;103:1361-7.
17. Dyer JM, Stymne S, Green AG, et al. High-value oils from plants. Plant J 2008;54:640-55.
18. Weselake RJ, Taylor DC, Rahman MH, et al. Increasing the flow of carbon into seed oil. Biotechnol Adv 2009;27:866-78.
19. Pollard MR, Anderson L, Fan C, et al. A specific acyl-ACP thioesterase implicated in medium-chain fatty acid production in immature cotyledons of Umbellularia californica. Arch Biochem Biophys 1991;284:306-12.
20. Mayer KM, Shanklin J. Identification of amino acid residues involved in substrate specificity of plant acyl-ACP thioesterases using a bioinformatics-guided approach. BMC Plant Biol 2007;7:1.
21. Salas JJ, Ohlrogge JB. Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. Arch Biochem Biophys 2002;403:25-34.
22. Somerville C, et al. Lipids: chapter 10 In BB Buchanan, W Gruissem, RL Jones, eds, Biochemistry and Molecular Biology of Plants. American Society of Plant Physiologists 2000:456–527.
23. Somerville C, et al. Lipids. In: BB Buchanan, W Gruissem, RL Jones (eds.) Biochemistry and Molecular Biology of Plants. American Society of Plant Physiologists, 2000, chap 10, p. 456-527.
24. van Erp H, Bates PD, Burgal J, et al. Castor phospholipid:diacylglycerol acyltransferase facilitates efficient metabolism of hydroxy fatty acids in transgenic Arabidopsis. Plant Physiol 2010;61:1092-106.
25. Damude HG, Kinney AJ. Enhancing plant seed oils for human nutrition. Plant Physiol 2008;147:962-8.
26. Napier JA. The production of unusual fatty acids in transgenic plants. Annu Rev Plant Biol 2007;58:295-319.
27. Napier JA, Graham IA. Tailoring plant lipid composition: designer oilseeds come of age. Curr Opin Plant Biol 2010;13:330-7.
28. Voelker TA, Worrell AC, Anderson L, et al. Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. Science 1992;257:72-4.
29. Voelker TA, Worrell AC, Anderson L, et al. Broad-range and binary-range acyl-acyl-carrier protein thioesterases suggest an alternative mechanism for medium-chain production in seeds. Plant Physiol 1997;114:669-77.
30. Dehesh K, Edwards P, Fillatti J, et al. KAS IV: a 3-ketoacyl-ACP synthase from Cuphea sp. is a medium chain specific condensing enzyme. Plant J 1998;15: 383-90.
31. Graham SA. Cuphea: a new plant source of medium-chain fatty acids. Crit Rev Food Sci Nutr 1989;28(2):139-73.
32. Harwood JL. Recent advances in the biosynthesis of plant fatty acids. Biochim Biophys Acta 1996;1301:7-56.
33. Napier JA, Sayanova O. The production of very-long-chain PUFA biosynthesis in transgenic plants: towards a sustainable source of fish oils. Proc Nutr Soc 2005;64:387-93.
34. Wallis JG, Watts JL, Browse J. Polyunsaturated fatty acid synthesis: what will they think of next? Trends Biochem Sci 2002;27:467.
35. Williams CM, Burdge G. Long-chain n-3 PUFA: plant v. marine sources. Proc Nutr Soc 2006;65:42-50.
36. Domergue F, Abbadi A, Heinz E. Relief for fish stocks: oceanic fatty acids in transgenic oilseeds. Trends Plant Sci 2005;10:112-6.
37. Girke T, Schmidt H, Zähringer U, et al. Identification of a novel delta 6-acyl-group desaturase by targeted gene disruption in Physcomitrella patens. Plant J 1998;15:39-48.
38. Sayanova O, Haslam R, Qi B, Lazarus CM, et al. The alternative pathway C20 Delta8-desaturase from the non-photosynthetic organism Acanthamoeba castellanii is an atypical cytochrome b5-fusion desaturase. FEBS Lett 2006;580:1946-52.
39. Abbadi A, Domergue F, Bauer J, et al. Biosynthesis of very-long-chain polyunsaturated fatty acids in transgenic oilseeds: constraints on their accumulation. Plant Cell 2004;16:2734-48.
40. Demaison L, Moreau D. Dietary n-3 polyunsaturated fatty acids and coronary heart disease-related mortality: a possible mechanism of action. Cell Mol Life Sci 2002;59:463-77.
41. Cahoon EB, Ohlrogge JB. Metabolic Evidence for the Involvement of a [delta]4-Palmitoyl-Acyl Carrier Protein Desaturase in Petroselinic Acid Synthesis in Coriander Endosperm and Transgenic Tobacco Cells. Plant Physiol 1994;104:827-37.
42. Metzger JO, Bornscheuer U. Lipids as renewable resources: current state of chemical and biotechnological conversion and diversification. Appl Microbiol Biotechnol 2006;71:13-22.
43. Li R, Yu K, Hildebrand DF. DGAT1, DGAT2 and PDAT expression in seeds and other tissues of epoxy and hydroxy fatty acid accumulating plants. Lipids 2010;45:145-57.
44. Bafor M, Smith MA, Jonsson L, et al. Biosynthesis of vernoleate (cis-12-epoxyoctadeca-cis-9-enoate) in microsomal preparations from developing endosperm of Eu-phorbia lagascae. Arch Biochem Biophys 1993; 303:145-51.
45. Alvarez HM, Kalscheuer R, Steinbuchel A. Accumulation and mobilization of storage lipids by Rhodococcus opacus PD630 and Rhodococcus ruber NCIMB 40126. Appl Microbiol Biotechnol 2000;54:218-23.
46. Alvarez HM, Souto MF, Viale A, et al. Biosynthesis of fatty acids and triacylglycerols by 2,6,10,14-tetramethyl pentadecane-grown cells of Nocardia globerula 432. FEMS Microbiol Lett 2001;200:195-200.
47. Alvarez HM, Steinbuchel A. Triacylglycerols in prokaryotic microorganisms. Appl Microbiol Biotechnol 2002;60:367-76.

48. Olukoshi ER, Packter MN. Importance of stored triacylglycerols in Streptomyces: possible carbon source for antibiotics. Microbiology 1994;140 (:931-43.
49. Akao T, Kusaka T. Solubilization of diglyceride acyltransferase from the membrane of Mycobacterium smegmatis. J Biochem 1976; 80:723-8.
50. Athenstaedt K, Daum G. The life cycle of neutral lipids: synthesis, storage and degradation. Cell Mol Life Sci 2006;63:1355-69.
51. Graham IA. Seed storage oil mobilization. Annu Rev Plant Biol 2008;59:115-42.
52. Hsieh K, Huang AH. Endoplasmic reticulum, oleosins, and oils in seeds and tapetum cells. Plant Physiol 2004;136:3427-34.
53. Murphy DJ, Vance J. Mechanisms of lipid-body formation. Trends Biochem Sci 1999;24: 109-15.
54. Banas A, Dahlqvist A, Ståhl U, et al. The involvement of phospholipid:diacylglycerol acyltransferases in triacylglycerol production. Biochem Soc Trans 2000;28:703-5.
55. Kim HU, Li Y, Huang AH. Ubiquitous and endoplasmic reticulum-located lysophosphatidyl acyltransferase, LPAT2, is essential for female but not male gametophyte development in Arabidopsis. Plant Cell 2005;17: 1073-89.
56. Cagliari A, Pinheiro-Margis M, Loss G, et al. Identification and expression analysis of castor bean (Ricinus communis) genes encoding enzymes from the triacylglycerol biosynthesis pathway. Plant Science 2010;179:499-509.
57. Kennedy EP. Biosynthesis of complex lipids. Fed Proc 1961;20:934-40.
58. Turnbull AP, Rafferty JB, Sedelnikova SE, et al. Analysis of the structure, substrate specificity, and mechanism of squash glycerol-3-phosphate (1)-acyltransferase. Structure 2001;9:347-53.
59. Bourgis F, Kader JC, Barret P, et al. A plastidial lysophosphatidic acid acyltransferase from oilseed rape. Plant Physiol 1999;120:913-22.
60. Yu B, Wakao S, Fan J, et al. Loss of plastidic lysophosphatidic acid acyltransferase causes embryo-lethality in Arabidopsis. Plant Cell Physiol 2004;45:503-10.
61. Weier D, Luhs WJ, Dettendorfer J, et al. sn-1-Acylglycerol-3-phosphate acyltransferase of Escherichia coli causes insertion of cis-11 eicosenoic acid into the sn-2 position of transgenic rapeseed oil. Molecular Breeding 1998;4:39-46.
62. Franca MG, Matos AR, Darcy-Lameta A, et al. Cloning and characterization of drought-stimulated phosphatidic acid phosphatase genes from Vigna unguiculata. Plant Physiol Biochem 2008;46:1093-100.
63. Nakamura Y, Tsuchiya M, Ohta H. Plastidic phosphatidic acid phosphatases identified in a distinct subfamily of lipid phosphate phosphatases with prokaryotic origin. J Biol Chem 2007;282:29013-21.
64. Kocsis MG, Weselake RJ. Phosphatidate phosphatases of mammals, yeast, and higher plants. Lipids 1996;31:785-802.
65. Pierrugues O, Brutesco C, Oshiro J, et al. Lipid phosphate phosphatases in Arabi-dopsis. Regulation of the AtLPP1 gene in response to stress. J Biol Chem 2001;276: 20300-8.
66. Siloto RM, Truska M, Brownfield D, et al. Directed evolution of acyl-CoA:diacylglycerol acyltransferase: Development and characterization of Brassica napus DGAT1 mutagenized libraries. Plant Physiol Biochem 2009;47:456-61.
67. Cahoon EB, Shockey JM, Dietrich CR, et al. Engineering oilseeds for sustainable production of industrial and nutritional feedstocks: solving bottlenecks in fatty acid flux. Curr Opin Plant Biol 2007;10:236-44.
68. Turkish AR, Henneberry AL, Cromley D, et al. Identification of two novel human acyl-CoA wax alcohol acyltransferases: members of the diacylglycerol acyltransferase 2 (DGAT2) gene superfamily. J Biol Chem 2005;280: 14755-64.
69. Saha S, Enugutti B, Rajakumari S, et al. Cytosolic triacylglycerol biosynthetic pathway in oilseeds. Molecular cloning and expression of peanut cytosolic diacylglycerol acyltransferase. Plant Physiol 2006;141:1533-43.
70. Durrett TP, McClosky DD, Tumaney AW, et al. A distinct DGAT with sn-3 acetyltransferase activity that synthesizes unusual, reduced-viscosity oils in Euonymus and transgenic seeds. Proc Natl Acad Sci USA 2010;107:9464-9.
71. Cases S, Smith SJ, Zheng YW, et al. Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. Proc Natl Acad Sci USA 1998;95:13018-23.
72. Bouvier-Nave P, Benveniste P, Oelkers P, et al. Expression in yeast and tobacco of plant cDNAs encoding acyl CoA:diacylglycerol acyltransferase. Eur J Biochem 2000;267:85-96.
73. Hobbs DH, Lu C, Hills MJ. Cloning of a cDNA encoding diacylglycerol acyltransferase from Arabidopsis thaliana and its functional expression. FEBS Lett 1999;452:145-9.
74. Katavic V, Reed DW, Taylor DC, et al. Alteration of seed fatty acid composition by an ethyl methanesulfonate-induced mutation in Arabidopsis thaliana affecting diacylglycerol acyltransferase activity. Plant Physiol 1995;108:399-409.
75. Zou J, Wei Y, Jako C, et al. The Arabidopsis thaliana TAG1 mutant has a mutation in a diacylglycerol acyltransferase gene. Plant J 1999;19:645-53.
76. Zhang M, Fan J, Taylor DC, et al. DGAT1 and PDAT1 acyltransferases have overlapping functions in Arabidopsis triacylglycerol biosynthesis and are essential for normal pollen and seed development. Plant Cell 2009;21:3885-901.
77. Lardizabal KD, Mai JT, Wagner NW, et al. DGAT2 is a new diacylglycerol acyltransferase gene family: purification, cloning, and expression in insect cells of two polypeptides from Mortierella ramanniana with diacylglycerol acyltransferase activity. J Biol Chem 2001:276:38862-9.
78. Cases S, Stone SJ, Zhou P, et al. Cloning of DGAT2, a second mammalian diacylglycerol acyltransferase, and related family members. J Biol Chem 2001;276:38870-6.
79. Beisson F, Koo AJ, Ruuska S, et al. Ara-bidopsis genes involved in acyl lipid metabolism. A 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a web-based database. Plant Physiol 2003;132:681-97.
80. Yu XX, Murray SF, Pandey SK, et al. Antisense oligonucleotide reduction of DGAT2 expression improves hepatic steatosis and hyperlipidemia in obese mice. Hepatology 2005;42:362-71.
81. Stone SJ, Myers HM, Watkins SM, et al. Lipopenia and skin barrier abnormalities in DGAT2-deficient mice. J Biol Chem 2004; 279:11767-76.
82. Orland MD, Anwar K, Cromley D, et al. Acyl coenzyme A dependent retinol esterification by acyl coenzyme A: diacylglycerol acyltransferase 1. Biochim Biophys Acta 2005;1737:76-82.
83. Oelkers P, Cromley D, Padamsee M, et al. The DGA1 gene determines a second triglyceride synthetic pathway in yeast. J Biol Chem 2002;277:8877-81.
84. Sandager L, Gustavsson MH, Stahl U, et al. Storage lipid synthesis is non-essential in yeast. J Biol Chem 2002;277:6478-82.
85. Kroon JT, Wei W, Simon WJ, et al. Identification and functional expression of a type 2 acyl-CoA:diacylglycerol acyltransferase (DGAT2) in developing castor bean seeds which has high homology to the major triglyceride biosynthetic enzyme of fungi and animals. Phytochemistry 2006;67:2541-9.
86. Shockey JM, Gidda SK, Chapital DC, et al. Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. Plant Cell 2006;18:2294-313.
87. Burgal J, Shockey J, Lu C, et al. Metabolic engineering of hydroxy fatty acid production in plants: RcDGAT2 drives dramatic increases in ricinoleate levels in seed oil. Plant Biotechnol J 2008;6:819-31.
88. Xu J, Francis T, Mietkiewska E, et al. Cloning and characterization of an acyl-CoA-dependent diacylglycerol acyltransferase 1 (DGAT1) gene from Tropaeolum majus, and a study of the functional motifs of the DGAT protein using site-directed mutagenesis to modify enzyme activity and oil content. Plant Biotechnol J 2008;6:799-818.
89. Kalscheuer R, Luftmann H, Steinbuchel A. Synthesis of novel lipids in Saccharomyces cerevisiae by heterologous expression of an unspecific bacterial acyltransferase. Appl Environ Microbiol 2004;70:7119-25.
90. Stoveken T, Kalscheuer R, Malkus U, et al. The wax ester synthase/acyl coenzyme A:diacylglycerol acyltransferase from Acineto-bacter sp. strain ADP1: characterization of a novel type of acyltransferase. J Bacteriol

2005;187:1369-76.
91. Mhaske V, Beldjilali K, Ohlrogge J, et al. Isolation and characterization of an Arabidopsis thaliana knockout line for phospholipid: diacylglycerol transacylase gene (At5g13640). Plant Physiol Biochem 2005;43:413-7.
92. Jako C, Kumar A, Wei Y, et al. Seed-specific over-expression of an Arabidopsis cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. Plant Physiol 2001;126:861-74.
93. Routaboul JM, Benning C, Bechtold N, et al. The TAG1 locus of Arabidopsis encodes for a diacylglycerol acyltransferase. Plant Physiol Biochem 1999;37:831-40.
94. Bates PD, Ohlrogge JB, Pollard M. Incorporation of newly synthesized fatty acids into cytosolic glycerolipids in pea leaves occurs via acyl editing. J Biol Chem 2007;282:31206-16.
95. Ghosal A, Banas A, Stahl U, et al. Saccharomyces cerevisiae phospholipid:diacylglycerol acyl transferase (PDAT) devoid of its membrane anchor region is a soluble and active enzyme retaining its substrate specificities. Biochim Biophys Acta 2007;1771: 1457-63.
96. Dahlqvist A, Stahl U, Lenman M, et al. Phospholipid:diacylglycerol acyltransferase: an enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. Proc Natl Acad Sci USA 2000; 97:6487-92.
97. Stahl U, Carlsson AS, Lenman M, et al. Cloning and functional characterization of a phospholipid:diacylglycerol acyltransferase from Arabidopsis. Plant Physiol 2004; 135:1324-35.
98. Oelkers P, Tinkelenberg A, Erdeniz N, et al. A lecithin cholesterol acyltransferase-like gene mediates diacylglycerol esterification in yeast. J Biol Chem 2000275:15609-12.
99. Stobart K, Mancha M, Lenman M, et al. Triacylglycerols are synthesized and utilized by transacylation reactions in microsomal preparations of developing safflower (Carthamus tinctorius L.). Planta,1997;203: 58-66.
100. Lehner R, and Kuksis A. Biosynthesis of triacylglycerols. Prog Lipid Res 1996;35:169-201.
101. McMaster CR, Bell RM. CDP-choline:1,2-diacylglycerol cholinephosphotransferase. Biochim Biophys Acta 1997;1348:100-10.
102. Stobart AK, and Stymne S. The interconversion of diacylglycerol and phosphatidylcholine during triacylglycerol production in microsomal preparations of developing cotyledons of safflower (Carthamus tinctorius L.). Biochem J 1985;232:217-21.
103. Lee J, Welti R, Schapaugh WT, et al. Phospholipid and triacylglycerol profiles modified by PLD suppression in soybean seed. Plant Biotechnol J 2011;9:359-72.
104. Parcy F, Valon C, Kohara A, et al. The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in concert to control multiple aspects of Arabidopsis seed development. Plant Cell 1997;9:1265-77.
105. Kagaya Y, Toyoshima R, Okuda R, et al. LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. Plant Cell Physiol 2005;46:399-406.
106. Meinke DW. A Homoeotic Mutant of Arabidopsis thaliana with Leafy Cotyledons. Science 1992;258:1647-50.
107. Meinke DW, et al. Leafy Cotyledon Mutants of Arabidopsis. Plant Cell 1994;6:1049-1064.
108. Raz V, Bergervoet JH, Koornneef M. Sequential steps for developmental arrest in Arabidopsis seeds. Development 2001;128:243-52.
109. Hills MJ. Control of storage-product synthesis in seeds. Curr Opin Plant Biol 2004;7:302-8.
110. Baud S, Lepiniec L. Physiological and developmental regulation of seed oil production. Prog Lipid Res, 2010;49:235-49.
111. Santos-Mendoza M, Dubreucq B, Baud S, et al. Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. Plant J 2008;54:608-20.
112. Gutierrez L, Van Wuytswinkel O, Catelain M, et al. Combined networks regulating seed maturation. Trends Plant Sci 2007;12:294-300.
113. Wang H, Guo J, Lambert Kn, et al. Deve-lopmental control of Arabidopsis seed oil biosynthesis. Planta 2007;226:773-83.
114. Cernac A, Benning C. WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. Plant J 2004;40:575-85.
115. Focks N, Benning C. wrinkled1: A novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. Plant Physiol 1998; 118:91-101.
116. Lara P, Onate-Sanchez L, Abraham Z, et al. Synergistic activation of seed storage protein gene expression in Arabidopsis by ABI3 and two bZIPs related to OPAQUE2. J Biol Chem 2003;278:21003-11.
117. Baud S, Mendoza MS, To A, et al. WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. Plant J 2007;50:825-38.
118. Mu J, Tan H, Zheng Q, et al. LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in Arabidopsis. Plant Physiol 2008; 148:1042-54.
119. Wang HW, Zhang HW, Hao YJ, et al. The soybean Dof-type transcription factor genes, GmDof4 and GmDof11, enhance lipid content in the seeds of transgenic Arabidopsis plants. Plant J 2007;52:716-29.
120. Hunter SC, Ohlrogge JB. Regulation of spinach chloroplast acetyl-CoA carboxylase. Arch Biochem Biophys 1998;359:170-8.
121. Andre C, Froehlich TE, Moll MR, et al. A heteromeric plastidic pyruvate kinase complex involved in seed oil biosynthesis in Arabidopsis. Plant Cell 2007;19:2006-22.
122. Baud S, Wiulleme S, Dubreucq B, et al. Function of plastidial pyruvate kinases in seeds of Arabidopsis thaliana. Plant J 2007;52:405-19.
123. Huang AH. Oleosins and oil bodies in seeds and other organs. Plant Physiol 1996;110: 1055-61.
124. Wallis JG, Browse J. Lipid biochemists salute the genome. Plant J 2010;61:1092-106.
125. Kim HU, Hsieh K, Ratnayake C, et al. A novel group of oleosins is present inside the pollen of Arabidopsis. J Biol Chem 2002;277:22677-84.
126. Murphy DJ. Structure, function and biogenesis of storage lipid bodies and oleosins in plants. Prog Lipid Res 1993;32:247-80.
127. Tzen JT, Lai YK, Chan KL, et al. Oleosin isoforms of high and low molecular weights are present in the oil bodies of diverse seed species. Plant Physiol 1990;94:1282-9.
128. Harwood JL, Guschina IA. The versatility of algae and their lipid metabolism. Biochimie 2009;91:679-84.
129. Guschina IA, Harwood JL. Lipids and lipid metabolism in eukaryotic algae. Prog Lipid Res 2006;45:160-86.
130. Thompson GA. Lipids and membrane function in green algae. Biochim Biophys Acta 1996;1302:17-45.
131. Bigogno C, Khozin-Goldberg I, Boussiba S, et al. Lipid and fatty acid composition of the green oleaginous alga Parietochloris incisa, the richest plant source of arachidonic acid. Phytochemistry 2002;60:497-503.
132. Hu Q, Sommerfield M, Jarvis E, et al. Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. Plant J 2008;54:621-39.
133. Ben-Amotz A, Shaish A, Avron M. Mode of Action of the Massively Accumulated beta-Carotene of Dunaliella bardawil in Protecting the Alga against Damage by Excess Irradiation. Plant Physiol 1989;91:1040-3.
134. Riekhof WR, Sears BB, Benning C. Annotation of genes involved in glycerolipid biosynthesis in Chlamydomonas reinhardtii: discovery of the betaine lipid synthase BTA1Cr. Eukaryot Cell 20054:242-52.
135. Cobelas MAL. Lipids in microalgae. I. Biochemistry. Grasas y Aceites 1989;40:118-145.
136. Ohlrogge J, Browse J. Lipid biosynthesis. Plant Cell 1995;7:957-70.
137. Parker, P.L., C. Van Baalen, L. Maurer, Fatty acids in eleven species of blue-green algae: geochemical significance. Science 1967;155:707-8.
138. De Swaaf ME, de Ruk TC, Eggink G, Sijtsma L. Optimisation of docosahexaenoic acid production in batch cultivation by Crypthecodinium cohnii. J Biotechnol 1999;70:185-92.
139. Haines T. Sulpholipids and halosulpholipids. In: Biomembranes of Eukaryotic Microorganisms. JA Erwin (ed.) New York, 1973, p. 197-232.
140. Dembitsky VM, Srebnik M. Natural halogenated fatty acids: their analogues and derivatives. Prog Lipid Res 2002;41:315-67.
141. Ratledge C. An overview of microbial lipids.

In: In Microbial
Lipids. Academic Press, New York, 1988, p. 3-11.
142. Hu Q. Environmental effects on cell composition. In: Handbook
of Microalgal Culture. Blackwell, Oxford, 2004, p. 83-93.
143. Pohl P, Wagner H. Control of fatty acid and lipid biosynthesis in Euglena gracilis by ammonia, light and DCMU. Z Naturforsch B 1972;27:53-61.
144. Tonon T. Long chain polyunsaturated fatty acid production and partitioning to triacylglycerols in four microalgae. Phytochemistry 2002;61:15-24.
145. Yongmanitchai W, Ward OP. Growth of and omega-3 fatty acid production by Phaeodactylum tricornutum under different culture conditions. Appl Environ Microbiol 1991; 57:419-25.
146. Carvalho AP, Pontes I, Gaspar H, et al. Metabolic relationships between macro- and micronutrients, and the eicosapentaenoic acid and docosahexaenoic acid contents of Pavlova lutheri. Enzyme and Microbial Technology 2006;38:358-66.
147. Petkov G, Garcia G. Which are fatty acids of the green alga Chlorella? Biochemical Systematics and Ecology 2007;35:281-5.
148. Guiheneuf F, Mimouni V, Ulmann L, et al. Combined effects of irradiance level and carbon source on fatty acid and lipid class composition in the microalga Pavlova lutheri commonly used in mariculture. Journal of Experimental Marine Biology and Ecology 2009;369:136-43.
149. Liu J, Huang J, Sun Z, et al. Differential lipid and fatty acid profiles of photoautotrophic and heterotrophic Chlorella zofingiensis: assessment of algal oils for biodiesel production. Bioresour Technol 2011;102:106-10.
150. Guiheneuf F, Fouqueray M, Mimouni V, et al. Effect of UV stress on the fatty acid and lipid class composition in two marine microalgae Pavlova lutheri (Pavlovophyceae) and Odontella aurita (Bacillariophyceae). Journal of Applied Phycology 2010;22:629-638.
151. Vasudevan PT, Briggs M. Biodiesel production--current state of the art and challenges. J Ind Microbiol Biotechnol 2008;35:421-30.
152. Lam MK, Lee KT, Mohamed AR. Homogeneous, heterogeneous and enzymatic catalysis for transesterification of high free fatty acid oil (waste cooking oil) to biodiesel: a review. Biotechnol Adv 2010;28:500-18.
153. Dyer JM, Mullen RT. Engineering plant oils as high-value industrial feedstocks for biorefining: the need for underpinning cell biology research. Physiol Plant 2008;132:11-22.
154. Hill J, Nelson E, Tilman D, et al. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. Proc Natl Acad Sci USA 2006;103:11206-10.
155. Li Y. Biofuels from microalgae. Biotechnol Prog 2008;24:815-20.
156. Radakovits, R., Jinkerson RE, Darzins A, et al., Genetic engineering of algae for enhanced biofuel production. Eukaryot Cell 2010; 9:486-501.
157. Huerlimann R, Nys R, Heimann K. Growth, lipid content, productivity, and fatty acid composition of tropical microalgae for scale-up production. Biotechnol Bioeng 2010;107: 245-57.
158. Petrie JR, Shrestha P, Mansour MP, et al. Metabolic engineering of omega-3 long-chain polyunsaturated fatty acids in plants using an acyl-CoA Delta6-desaturase with omega3-preference from the marine microalga Micromonas pusilla. Metab Eng 2010;12:233-40.
159. Meyer A, Kirsch H, Domergue F, et al. Novel fatty acid elongases and their use for the reconstitution of docosahexaenoic acid biosynthesis. J Lipid Res 2004;45:1899-909.
160. Domergue F, Abbadi A, Zähringer U, et al. In vivo characterization of the first acyl-CoA Delta6-desaturase from a member of the plant kingdom, the microalga Ostreococcus tauri. Biochem J 2005;389:483-90.
161. Drexler H, Spiekermann P, Meyer A, Domergue F, et al. Metabolic engineering of fatty acids for breeding of new oilseed crops: strategies, problems and first results. J Plant Physiol 2003;160:779-802.

4.2 Capítulo 2

O Presente capítulo foi publicado no livro ***Gene Duplication*** (2012).

11

The Evolutionary History of CBF Transcription Factors: Gene Duplication of CCAAT – Binding Factors NF-Y in Plants

Alexandro Cagliari, Andreia Carina Turchetto-Zolet, Felipe dos Santos Maraschin, Guilherme Loss, Rogério Margis and Marcia Margis-Pinheiro
Universidade Federal do Rio Grande do Sul/UFRGS
Brazil

1. Introduction

Eukaryotic gene expression is often controlled by complex and refined combinatorial transcription factor networks composed of multiprotein complexes that derive their gene regulatory capacity from both intrinsic properties and from their *trans*-acting partners (Singh, 1998; Wolberger, 1998; Remenyi *et al*., 2004). Participation in such higher complex order allows an organism to use single transcription factors to control multiple genes with different temporal and spatial expression patterns (Siefers *et al*., 2009).
In this chapter, we provide a synopsis of the genetic and genomic mechanisms that might be responsible for the gene copy diversification observed in the eukaryotic NF-Y transcription factor family. We identify the genes coding for NF-Y transcription factors in eukaryotes with an emphasis on the duplication of the NF-Y family in the plant lineage and discuss the important consequences of its gene diversification.

2. The CCAAT *cis*-element promoter

Eukaryotic genes contain numerous *cis*-regulatory elements that mediate their induction, repression or basal transcription (Dynan and Tjian, 1985; Myers *et al*., 1986; Maity and de Crombrugghe, 1998). These regulatory elements can be found in the proximity of transcribed genes, such as the promoter region and/or in distant regions of the genes where they may act as enhancers (de Silvio *et al*., 1999).
The transcriptional regulation of several eukaryotic genes is coordinated through sequence-specific binding of proteins to the promoter region located upstream of the gene. During evolution, many of these protein-binding sequences, which are found in a wide variety of organisms, have shown a high degree of conservation (Edwards *et al*., 1998).
The CCAAT box is one of the most common upstream elements, found in approximately 25–30% of eukaryotic promoters (Bucher, 1990; Mantovani, 1998). It is typically located between 60–100 bp upstream of the transcription start site and it can function in direct or in inverted orientations (Dorn *et al*., 1987b; Bucher, 1990; Edwards *et al*., 1998; Mantovani, 1998; Stephenson *et al*., 2007) with possible cooperative interactions between multiple boxes (Tasanen *et al*., 1992) or other conserved motifs (Muro *et al*., 1992; Rieping and Schoffl, 1992;

Edwards *et al.*, 1998). CCAAT boxes are highly conserved within homologous genes across species in terms of position, orientation, and flanking nucleotides (Mantovani, 1998). In addition, the spacing between the CCAAT box and other promoter-specific *cis*-elements is also conserved among species (Dorn *et al.*, 1987a; Chodosh *et al.*, 1988; Maity and de Crombrugghe, 1998).The expression of genes under the control of promoters that contain CCAAT boxes may be ubiquitous or tissue/stage specific, suggesting that the gene expression pattern is also determined by other *cis* and *trans* elements (Stephenson *et al.*, 2007).

In *Sacharomyces cerevisiae*, CCAAT boxes are found in the promoters of cytochrome genes, in genes coding for proteins that are activated by non-fermentable carbon sources (McNabb *et al.*, 1995) and in genes involved in nitrogen metabolism (Dang *et al.*, 1996). In the filamentous fungus *Aspergillus nidulans*, CCAAT boxes are present in genes involved with penicillin biosynthesis (Steidl *et al.*, 1999). In higher eukaryotes, a multitude of promoters contain CCAAT boxes, including those of developmentally controlled and tissue-specific genes (Berry *et al.*, 1992), housekeeping and inducible genes (Roy and Lee, 1995) and cell-cycle regulated genes (Mantovani, 1998). In addition, many cell-cycle regulated promoters lack a recognizable TATA-box, but contain more than one CCAAT box in a position close to and sometimes overlapping with the start site of transcription (Zwicker and Muller, 1997).

3. The CBF/NF-Y transcription factor

Several CCAAT-binding proteins have been isolated and described, including CBF/NF-Y (CCAAT Binding Factor/Nuclear Factor of the Y box), CTF/NF1 (CCAAT Transcription Factor/Nuclear Factor 1), C/EBP (CCAAT/Enhancer Binding Protein) and CDP (CCAAT Displacement Protein) (Mantovani, 1999). Among them, NF-Y is the most ubiquitous and specific one acting as a key proximal promoter factor in the transcriptional regulation of an array of different eukaryotic genes. Unlike other CCAAT-binding proteins, NF-Y requires a high degree of conservation of the CCAAT pentanucletide sequence and shows strong preference for specific flanking sequences (Dorn *et al.*, 1987a; Stephenson *et al.*, 2007). Therefore, the NF-YC transcription factor can be distinguished from the other CCAAT-binding proteins based on its DNA sequence requirements (Maity and de Crombrugghe, 1998).

The CBF/NF-Y transcription factor, which will be referenced in this chapter as NF-Y, is a conserved oligomeric transcription factor found in all eukaryotes that is involved in the regulation of diverse genes (Maity *et al.*, 1992; McNabb *et al.*, 1995; Edwards *et al.*, 1998; Mantovani, 1998; Siefers *et al.*, 2009). NF-Y typically acts in concert with other regulatory factors to modulate gene expression in a highly controlled manner (Nelson *et al.*, 2007). In many eukaryotic promoters, the functional NF-Y-binding sites are relatively close to the TATA motif (Bucher, 1990) and are invariably flanked by at least one additional functionally important *cis*-element. Several reports have shown that various factors, including transcription factors, co-activators, and TATA-binding proteins, interact with NF-Y or its subunits in promoting transcriptional regulation (Mantovani, 1999; Yazawa and Kamada, 2007). NF-Y was originally identified as the protein that recognizes the MHC class II conserved Y box in Ea promoters (Dorn *et al.*, 1987a; Matuoka and Chen, 2002). It specifically recognizes the consensus sequence 5′-CTGATTGGYYRR-3′ or 5′-YYRRCCAATCAG-3′(Y is 5 pyrimidines and R is 5 purines) present in the promoter region of eukaryotic genes. Bioinformatic analyses indicate that about 30% of mammalian promoters have predicted

NF-Y binding sites (Bucher, 1990; Testa *et al.*, 2005), and chromatin immunoprecipitation data have demonstrated additional widespread NF-Y binding in nonpromoter sites.
Suggesting the importance of binding context, NF-Y-regulated gene expression can be tissue specific, developmentally regulated, or constitutive (Maity and de Crombrugghe, 1998; Siefers *et al.*, 2009).The transcriptional activity of NF-Y can be regulated by differential expression, alternative splicing, protein–protein interactions, and cellular redox potential (Matuoka and Yu Chen, 1999).
NF-Y has been shown to be involved in the regulation of some G1/S genes whose expressions are attenuated during the senescence process (Matuoka and Yu Chen, 1999). NF-Y plays a pivotal role in the cell cycle regulation of the mammalian cyclin A, *cdc25C*, and *cdc2* genes, in the S-phase of the cell cycle (Currie, 1998). Additionally, there are a number of genes involved in the cellular response to damage and stress, including the phospholipid hydroperoxide glutathione peroxidase genes (Huang *et al.*, 1999), which are regulated by NF-Y, indicating its pivotal role in the removal of damaging agents from cells (Matuoka and Chen, 2002). Although NF-Y functions basically as a transactivator of gene expression, it is also involved, directly or indirectly, in the downregulation of transcription. For instance, NF-Y binds to the mouse CCAAT box renin enhancer and blocks the binding of positive regulatory elements (Shi *et al.*, 2001). In this case, NF-Y dysfunction would lead to the damage of systems that control blood pressure (Matuoka and Chen, 2002).
NF-Y is composed of three different subunits named NF-YA (also known as HAP-2 or CBF-B), NF-YB (HAP3 or CBF-A), and NF-YC (HAP5 or CBF-C) that interact to form a complex that can bind CCAAT DNA motifs and control the expression of target genes (Figure 1). Each subunit is required for DNA binding, subunit association and transcriptional regulation in both vertebrates and plants (Sinha *et al.*, 1995; Stephenson *et al.*, 2007). Yeast possesses a fourth subunit, called HAP4, which provides a transcriptional activation domain to the complex (Forsburg and Guarente, 1989; Lee *et al.*, 2003). The yeast HAP4 protein is not needed for DNA-binding but contains an acidic domain that is essential to promote transactivation when associated with the HAP2/HAP3/HAP5 complex (Olesen and Guarente, 1990; Serra *et al.*, 1998). In vertebrates, the function of this fourth domain was incorporated into other subunits (Forsburg and Guarente, 1989; Yazawa and Kamada, 2007).
Despite the wide cellular distribution and functional variability of NF-Y-regulated genes, most eukaryotic genomes have only one or two genes encoding each NF-Y subunit (Maity and de Crombrugghe, 1998; Riechmann and Ratcliffe, 2000). Fungi and animals, for example, present single genes encoding each protein subunit. Thus, there is minimal combinatorial diversity in the subunit composition of the heterotrimeric NF-Y in these organisms (Siefers *et al.*, 2009). In contrast, the NF-Y complex in vascular plants is generally encoded by gene families (Riechmann and Ratcliffe, 2000).

3.1 NF-Y subunits

NF-Y is the only transcription factor thus far identified for which the interaction of three heterologous subunits creates the DNA binding domain (Maity and de Crombrugghe, 1992; McNabb *et al.*, 1995; Sinha *et al.*, 1995). All three NF-Y subunits are essential for the DNA binding activity and one molecule of each subunit forms the NF-Y-DNA complex (Maity and De Crombrugghe, 1996). Each NF-Y subunit contains a conserved domain with identities greater than 70% across species. This highly conserved domain is located at the *C*-terminus of NF-YA; in the central part of NF-YB; and at the *N*-terminus of NF-YC (Li *et al.*, 1992).

The NF-YA conserved domain can be divided in two functionally distinct regions: an *N*-terminal region that is required for NF-YB and NF-YC association and a *C*-terminal region required for DNA-binding (Maity and de Crombrugghe, 1992). Additionally, NF-YA usually contains a glutamine (Q)-rich and a serine/threonine (S/T)-rich regions. There are numerous variants of NF-YA due to alternative splicing at the Q-S/T domains (Li *et al.*, 1992) and, although the expression of these isoforms is variable depending of tissue and cell types, they all seem intact in terms of transcriptional function (Matuoka and Chen, 2002).

Both NF-YB and NF-YC subunits possess the highly conserved histone-fold motif (HFM) and are structurally similar to core histone subunits H2B and H2A, respectively, and to the archaebacterial histone-like protein Hmf-2 (Arents and Moudrianakis, 1995; Baxevanis *et al.*, 1995; Mantovani, 1998). In terms of identity, NF-YB is 30% identical to H2B, 14% to H2A, 17% to H4 and 18% to H3; NF-YC is 21% identical to H2A, 15% to H4 and H3 and 20% to H2B (Liberati *et al.*, 1999). Other proteins showing a remarkable identity (25-30%) to both NF-YB and NF-YC are present in *Archaea*. These proteins homodimerize and associate with DNA, forming nucleosome-like structures (Sandman *et al.*, 1990). The NF-YB and NF-YC subunits also contain residues that are important for their contact with DNA (Romier *et al.*, 2003; Stephenson *et al.*, 2007). In contrast, the conserved segment of NF-YA has no homology with the histone-fold motif, or with any of the known dimerization motifs present in other heteromeric DNA-binding proteins (Maity and de Crombrugghe, 1992).

Some portions of NF-YA, NF-YB and NF-YC present a high degree of identity with yeast HAP3, HAP2 and HAP5, respectively. These HAP genes, which are components of the yeast CCAAT-binding protein, are necessary for the expression of genes encoding components of the electron transport chain. Yeast strains mutated for either of the three genes failed to grow on media containing a nonfermentable carbon source such as lactate or glycerol, a characteristic respiratory-defect phenotype (McNabb *et al.*, 1995).

Assembly of the NF-Y heterotrimer in mammals (where this complex is better studied) follows a strict, stepwise pattern (Sinha *et al.*, 1995; Sinha *et al.*, 1996) (Figure 1). Initially, the NF-YB and NF-YC subunits form a tight heterodimer (Figure 1a) similar to those of the HFM, a conserved protein–protein and DNA-binding interaction module (Luger *et al.*, 1997) composed by 65 amino acid stretch common to all histones that is required for nucleosome formation (Baxevanis *et al.*, 1995; Luger *et al.*, 1997; de Silvio *et al.*, 1999). This dimer then moves to the nucleus, where the third subunit (NF-YA, Figure 1b) is recruited to generate the complete, heterotrimeric NF-Y (Figure 1c). Interestingly, NF-YA is unable to interact with the NF-YB or NF-YC alone, interacting only with the NF-YB-NF-YC heterodimer (Serra *et al.*, 1998). The complete NF-Y is able to bind promoters containing the core pentamer nucleotide sequence CCAAT (Figure 1d) with high specificity and affinity resulting in either positive or negative transcriptional regulation (Figure 1e) (Peng and Jahroudi, 2002; 2003; Ceribelli *et al.*, 2008; Siefers *et al.*, 2009).

Because the NF-Y transcription factor contains H2B-like and H2A-like molecules (NF-YB and NF-YC, respectively), the complex presents all the core histone components and could mimic the interaction of the nucleosome core with genomic DNA (Struhl and Moqtaderi, 1998). In this scenario, it has been demonstrated that the NF-YA/NF-YB/NF-YC trimer or the NF-YB/NF-YC dimer can bind to H3/H4 tetramer during nucleosome assembly (Caretti *et al.*, 1999). In addition, the NF-Y complex also can bind to the chromatin even after nucleosome formation, indicating the ability of NF-Y to interact with genomic DNA assembled in the nucleosome. The interaction between the NF-Y transcription factor and the DNA molecule causes local disruption of the nucleosomal architecture (Coustry *et al.*, 2001).

This disruption results in a partial dissociation of DNA from the histone core, which might enable the access of the general transcription machinery to initiate the transcription process.

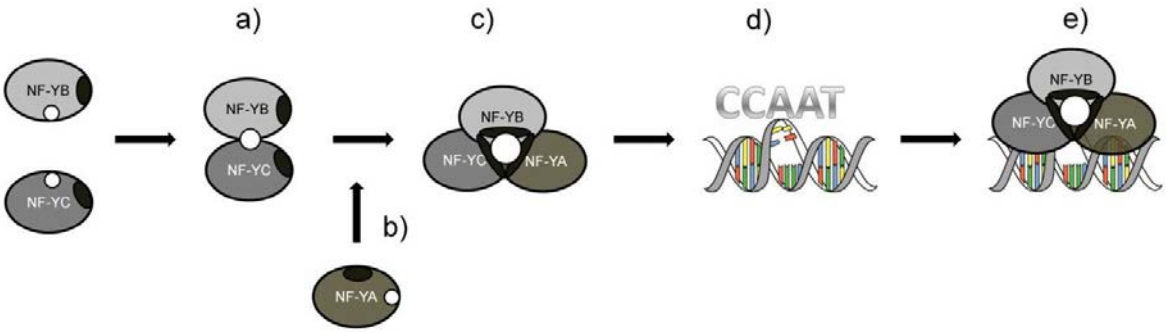


Fig. 1. Assembly of NF-Y subunits and its binding to DNA. Initially, the NF-YB and NF-YC subunits form a tight heterodimer via protein–protein interactions **(a)**. The dimer then moves to the nucleus, where is recruited the third subunit (NF-YA) **(b)** to generate the complete, heterotrimeric NF-Y **(c)** that is able to bind promoters containing the core pentamer nucleotide sequence CCAAT **(d)** resulting in either positive or negative transcriptional regulation **(e)**. Adapted from Mantovani (1999). White circles and oblong black circles into each NF-Y subunit represent the DNA-binding domain and the NF-Y interaction domain of NF-Y subunits, respectively.

4. Gene duplication and evolution

DNA duplication act as one of the main forces driving the evolution of organisms by creating the raw genetic material that natural selection can subsequently modify. Gene duplications arise in eukaryotes at a rate of 0.01 paralogs per gene per million years (Lynch and Conery, 2000), the same order of magnitude of the mutation rate per nucleotide per year (De Grassi *et al.*, 2008). Duplication of individual genes, chromosomal segments, or entire genomes represent the primary source for the origin of evolutionary novelties, including new gene functions and expression patterns (Holland *et al.*, 1994; Sidow, 1996; Lynch and Conery, 2000). However, how duplicated genes successfully evolve from an initial state of complete redundancy, wherein one copy is likely to be expendable, to a stable situation in which both copies are maintained by natural selection, is unclear (Sidow, 1996; Lynch and Conery, 2000; Ober, 2010).

In the evolutionary history of plants, genome duplications have been relatively common, leading to the hypothesis that most angiosperms are to some extent polyploidal (Soltis, 2005). The genome of Arabidopsis, for example, possesses traces of at least three polyploidy events (Vision *et al.*, 2000; Simillion *et al.*, 2002), followed by subsequent gene loss (Bowers *et al.*, 2003; Ober, 2010).

Similar to a point mutation, a duplication that occurs in an individual can be fixed or lost in the population. Compared with pre-existing alleles, if a new allele of the duplicate gene is selectively neutral, it has a small probability (1/2N) to be fixed in a diploid population (where N is the effective population size). This suggests that the majority of duplicated genes will be lost. For those duplicated genes that do become fixed, the fixation time averages is 4N generations (Kimura, 1989; Zhang, 2003).

On an evolutionary scale, gene duplication may result in new functions via different scenarios. Although the most likely outcome is a loss of function in one of the two gene copies (nonfunctionalization, Figure 2a), in rare instances one copy may acquire a novel evolutionarily advantageous function and become preserved by natural selection (neofunctionalization, Figure 2b), while the other copy retains the original function. Alternatively, after duplication, mutations may occur in both genes leading to specialization to perform complementary functions (subfunctionalization, Figure 2c) (Lynch and Conery, 2000; Lynch and Force, 2000). This process produces novel genetic variants that drive genetic innovation (Lynch and Conery, 2000; Conrad and Antonarakis, 2007). Because gene duplication generates functional redundancy, it is often not advantageous to the organism to possess two identical genes. In nonfunctionalization (Figure 2a), the accumulation of deleterious mutations might lead to the loss of the original function of one paralogue. Alternatively, instead of being completely lost, many duplicated genes are silenced or become pseudogenes and are thus either unexpressed or functionless (Gallagher *et al.*, 2004; Nicole *et al.*, 2006; Yang *et al.*, 2006; Beisswanger and Stephan, 2008; Xiong *et al.*, 2009). Pseudogenization is the most frequent fate of duplicated genes. In *Caenorhabditis elegans*, for example, genomic analyses have identified 2168 pseudogenes or approximately one pseudogene for every eight functional genes (Harrison *et al.*, 2001). In humans, one pseudogene was identified for approximately every two functional genes (Harrison *et al.*, 2002). As pseudogenes generally do not confer a selective advantage, they have a low probability of being fixed in large populations (Ober, 2010).

Unless the presence of an extra amount of gene product is advantageous, it is unlikely that two genes with the same function will be stably maintained in the genome of the organism (Nowak *et al.*, 1997). In subfunctionalization (Figure 2c), both duplicated copies may become, by accumulation of mutations, partially compromised to the point at which their total capacity is reduced to the level of the single-copy ancestral gene (Force *et al.*, 1999; Stoltzfus, 1999; Lynch and Force, 2000). Subfunctionalization can occur through the modification of the regulatory elements by mutations (Force *et al.*, 1999; Hinman and Davidson, 2007) or by epigenetic silencing (Rodin and Riggs, 2003). In an evolutionary scale, one of the most important forms of subfunctionalization is the division of gene expression after duplication (Force *et al.*, 1999). For example, zebrafish ENGRAILED 1 and ENGRAILED 1-B, generated by a chromosomal segmental duplication, are a pair of transcription factors that occurred in the lineage of ray-finned fish. While ENGRAILED-1 is expressed in the pectoral appendage bud, ENGRAILED 1-B is expressed in a specific set of neurons in the hindbrain/spinal cord (Force *et al.*, 1999). In yeast, more than 40% of gene pairs exhibit significant expression divergence (Gu *et al.*, 2002). Also, the comparison of 17 fungal genomes revealed that duplicated genes rarely diverge with respect to biochemical function, but typically diverge with respect to regulatory control (Wapinski *et al.*, 2007). On the other hand, if two redundant gene copies were retained without significant functional divergence in the genome, the organism may acquire increased genetic robustness against harmful mutations (Figure 1d) (Conrad and Antonarakis, 2007).

In neofunctionalization (Figure 2b), the ancestral gene keeps its ancestral function, while the duplicated gene gains a new function under positive selection for advantageous mutations (De Grassi *et al.*, 2008). However, in many cases, rather than an entirely new function, a related function evolves after gene duplication. For example, the red and green-sensitive opsin genes of humans where the result of a gene duplication that occurred in hominoids and Old World monkeys (Yokoyama and Yokoyama, 1989). After the duplication process,

functional divergence of the two opsins resulted in a 30-nanometer difference in their maximum absorption wavelength. This difference conferred a sensitivity to a wide range of colors for humans and related primates (Zhang, 2003).

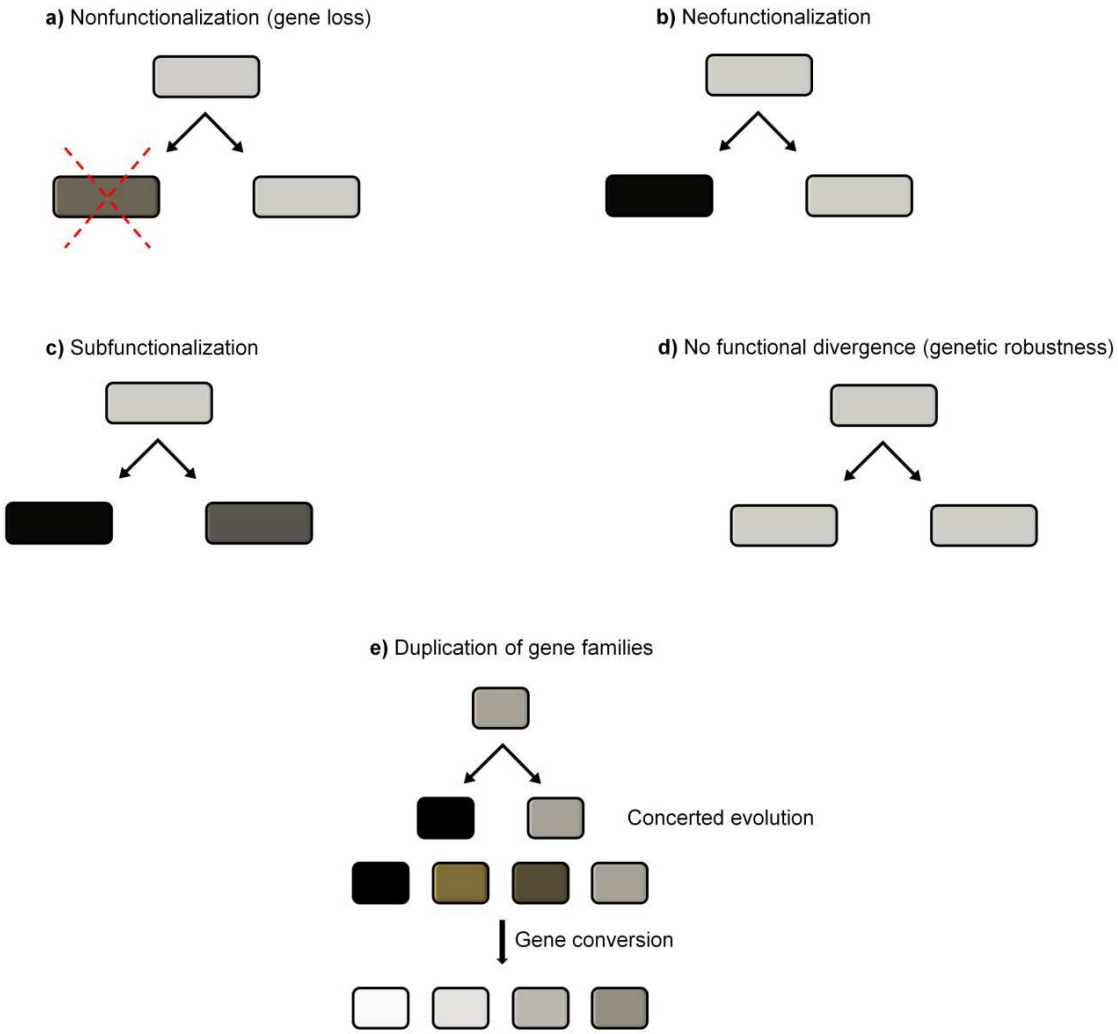


Fig. 2. Evolutionary fate of duplicated genes. Gene duplication may result in new functions via different scenarios. **(a)** nonfunctionalization; **(b)** neofunctionalization; **(c)** subfunctionalization; **(d)** genetic robustness and **(e)** gene conversion. Adapted from Conrad and Antonarakis (2007).

The fate of a gene that suffers duplication seems to be the result of diverse and, in some cases, interdependent factors (Taylor *et al.*, 2001). These variables include its functional category (Papp *et al.*, 2003; Kondrashov and Koonin, 2004; Marland *et al.*, 2004), degree of conservation (Conant and Wagner, 2002; Davis and Petrov, 2004; Jordan *et al.*, 2004; Braybrook and Harada, 2008), sensitivity to dosage effects (Kondrashov and Koonin, 2004), as well as its regulatory and architectural complexity (He and Zhang, 2005). Some observations indicate that natural selection created a preferential association of duplications with certain gene categories. For example, genes encoding proteins that interact with the environment are more frequently retained after the duplication process than genes which interact at intracellular compartments

(Li *et al.*, 2003; Marland *et al.*, 2004). In addition, genomes tend to retain duplicated genes involved in signal transduction and transcription, but to lose duplicated DNA repair genes (Blanc and Wolfe, 2004; Maere *et al.*, 2005; Paterson *et al.*, 2010).

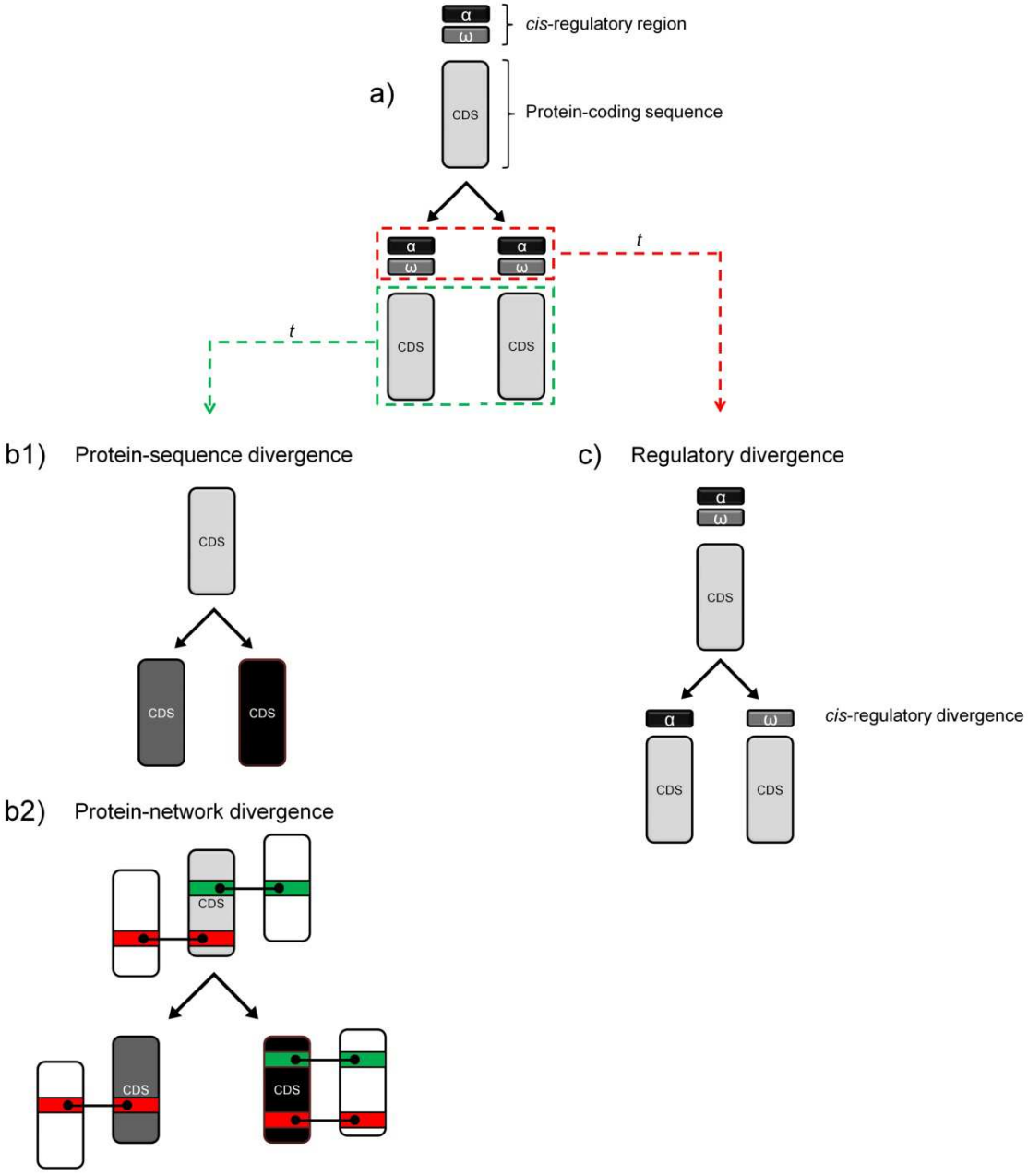


Fig. 3. Functional divergence of duplicated genes. **(a)** The independent evolution of *cis*-regulatory and the protein-coding regions. **(b1)** protein sequence divergence after duplication; **(b2)** protein network divergence with increase or loss of partners and **(c)** DNA sequence regulatory divergence after duplication (certain regulatory motifs are lost in one copy of the duplicated gene sequence). *t;* evolutionary time. Adapted from Conrad and Antonarakis (2007).

It has been shown that shortly after duplication the protein-coding sequence and *cis*-regulatory regions of some duplicated genes can evolve independently (Figure 3a) (Wagner, 2000). This independent evolution can generate protein sequence divergence of duplicated genes (Figure 3b1) or protein network divergence (Figure 3b2), where the protein interaction domains (*cis*-regulatory elements) of the original sequence evolve by maintenance, gain, or loss of interacting partners. Alternatively, the divergence of *cis*-regulatory motifs in the promoter-proximal region (Figure 3c) can generate expression divergence between the duplicated genes (Conrad and Antonarakis, 2007).

5. Gene duplication of NF-Y in plants

While duplication of NF-Y genes is poorly understood in the plant lineage, many of the functional mechanistic details are likely conserved across plant, animal and fungal lineages. This inference comes from strong cross-kingdom conservation of functional important amino acid residues in mammalian and yeast NF-Ys (Maity and de Crombrugghe, 1992; Maity *et al.*, 1992; Sinha *et al.*, 1995; Coustry *et al.*, 1996; Kim *et al.*, 1996; Sinha *et al.*, 1996; Mantovani, 1998; Romier *et al.*, 2003). CCAAT-like motifs are found in several plant promoters, and binding activity to CCAAT sequences has been identified in plant nuclear extracts (Yazawa and Kamada, 2007). Besides, at least some plant NF-YA and NF-YB subunits have been shown to complement yeast mutant strains lacking the corresponding NF-Y subunit. Additionally, several groups have demonstrated that each of the three plant NF-Y proteins can substitute their yeast counterparts in gene expression assays (Edwards *et al.*, 1998; Masiero *et al.*, 2002; Ben-Naim *et al.*, 2006; Siefers *et al.*, 2009). These observations indicate that plant NF-Y subunits might act as general transcription factors, as in mammals (Yamamoto *et al.*, 2009).

Although a complete functional plant NF-Y complex has not yet been described, the individual subunits are known to be involved in a number of important physiological processes, such as specific developmental processes and response to environmental stimuli (Lotan *et al.*, 1998; Kusnetsov *et al.*, 1999; Miyoshi *et al.*, 2003; Ben-Naim *et al.*, 2006; Combier *et al.*, 2006; Wenkel *et al.*, 2006; Cai *et al.*, 2007; Nelson *et al.*, 2007; Warpeha *et al.*, 2007; Siefers *et al.*, 2009). A well-established example is the NF-YB subunit gene called LEAFY COTYLEDON-1 (LEC1), which specifically controls embryo development, especially the maturation phase. LEC1 plays specialized roles not only because of its developmentally regulated expression but also due to its distinct molecular activity, as the *in vivo* function of LEC1 cannot be replaced by other NF-YB subunits, except for the most closely related Leafy Cotyledon 1 Like (L1L) (Kwong *et al.*, 2003; Lee *et al.*, 2003; Yamamoto *et al.*, 2009). In Arabidopsis, many NF-Y subunit genes are expressed ubiquitously, although some are differentially expressed. For example, while the AtNF-YC-4 transcript accumulates in seeds 7 days after germination, AtNF-YB-9 is only expressed in green siliques (Gusmaroli *et al.*, 2001).

Plant NF-Y function also appears to be important for responses to drought stress. Although a specific mechanism of action remains unclear, overexpression of the AtNF-YB1 subunit and its orthologue in maize (*Zea mays*), ZmNF-YB2, leads to enhanced drought resistance (Nelson *et al.*, 2007). Another study showed that overexpression of maize NF-YA5 reduced drought susceptibility, anthocyanin production and stomatal aperture, while *nf-ya5* mutants had the expected opposite phenotype in each situation (Li *et al.*, 2008). In addition, several

publications strongly suggest that NF-Y transcription factors are also involved in photoperiod-regulated flowering (Ben-Naim *et al.*, 2006; Wenkel *et al.*, 2006; Siefers *et al.*, 2009).

We adopted a high throughput comparative genomic approach to conduct a broad survey of fully sequenced genomes, including representatives of amoebozoa, yeasts, fungi, algae, mosses, plants, vertebrate and invertebrate species to identify the presence of homologous genes coding for each of the three subunits that form the NF-Y transcription factor (Table 1). NF-Y gene and protein sequences were obtained through blast searches (blastp, blastx and tblastx) against the Protein and Genome databases with the default parameters at the NCBI (National Center for Biotechnology Information - http://www.ncbi.nlm.nih.gov) and against completed genome projects database at the JGI (Joint Genome Institute - http://www.jgi.doe.gov).

The results point to a scenario where all fungi and the majority of metazoa possess single genes coding for each of the NF-Y subunits (Table 1). The metazoa exceptions include the amphioxus *Branchiostoma floridae,* the nematode *Caenorhabditis elegans* and the gastropod *Lottia gigantea,* all of each present a proportional duplication of the three subunits, possessing two genes for each subunit (Table 1).

In contrast, plants possess gene families coding for each NF-Y subunit (Table 1). For instance, in the model plant *Arabidopsis thaliana* 10 genes coding for NF-YA, 13 for NFY-B, and 13 for NF-YC were identified. Because of the heterotrimeric composition, the 36 Arabidopsis NF-Y subunits could theoretically combine to generate 1.690 unique transcription factors (Siefers *et al.*, 2009). This Arabidopsis NF-Y expansion is a general feature of the plant lineage, including monocots and eudicots. In rice (*Oryza sativa*), for example, 11 genes were identified coding for the NF-YA subunit, 10 for NF-YB and 8 for NFY-C. Four of the rice NF-YB subunits have been characterized and at least one of these genes is involved in chloroplast development (Miyoshi *et al.*, 2003; Yazawa and Kamada, 2007). Interestingly, the moss *Physcomitrella patens* and the lycophyte *Selaginella mollendorffii* possess single genes coding for NF-YA subunits whereas the other subunits are encoded by multiple genes (Table 1).

Since the evolutionary rates can be species dependent, the difference observed in the number of genes of NF-Y subunits in eukaryotic class (Table 1), especially in vascular plants, can be result of recent duplication process that contribute to the establishment of genes families coding each NF-Y subunit. However, some duplicated genes might have suffered high level of diversification what could be responsible to prevent their identification in our analyses.

Representative plants genes (monocot and eudicot) were selected to perform phylogenetic analyses of the NF-Y subunits. The phylogenetic analysis was reconstructed after protein sequence alignments using a Bayesian approach in MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). The mixed amino acid substitution model plus gamma and invariant sites was used in two independent runs of 5,000,000 generations each with two Metropolis-coupled Monte Carlo Markov chains (MCMCMC) that were run in parallel (starting each from a random tree). Markov chains were sampled every 100 generations, and the first 25% of the trees were discarded as burn-in. The remaining ones were used to compute the majority rule consensus tree, the posterior probability of clades and branch lengths (Figure 4 to 6).

Phylo	Specie	Code	Subunit A Genes	Subunit B Genes	Subunit C Genes
Metazoa	*Homo sapiens*	Hsa	1	1	1
	Mus musculus	Mmu	1	1	1
	Rattus norvegicus	Rno	1	1	1
	Canis familaris	Cfa	1	1	1
	Monodelphis domestica	Mdo	1	1	1
	Gallus gallus	Gga	1	1	1
	Xenopus tropicalis	Xtr	1	1	1
	Gasterosteus aculeatus	Gac	1	1	1
	Oryzias latipes	Ola	1	1	1
	Takifugu rubripes	Tru	1	1	1
	Danio rerio	Dre	1	1	1
	Ciona savignyi	Csa	1	1	1
	Branchiostoma floridae	Bfl	2	2	2
	Strongylocentrotus purpuratus	Spu	1	1	1
	Drosophila melanogaster	Dme	1	1	1
	Anopheles gambiae	Aga	1	1	1
	Tribolium castaneum	Tca	1	1	1
	Caenorhabditis elegans	Cel	2	2	2
	Lottia gigantea	Lgi	2	2	2
	Nematostella vectensis	Nve	1	1	1
Fungi	*Neurospora crassa*	Ncr	1	1	1
	Candida tropicalis	Ctr	1	1	1
	Tuber melanosporum	Tme	1	1	1
	Pyrenophora teres	Pte	1	1	1
	Aspergillus nidulans	Ani	1	1	1
	Chaetomium globosum	Cgl	1	1	1
	Penicillium marneffei	Pma	1	1	1
	Talaromyces stipitatus	Tst	1	1	1
	Sordaria macrospora	Sma	1	1	1
Heterolobosea	*Naegleria gruberi*	Ngr	1	1	1
Metaphyta	*Manihot esculenta*	Mes	12	15	9
	Ricinus communis	Rco	6	12	7
	Populus tricocharpa	Ptr	8	17	9
	Medicago truncatula	Mtr	5	10	5
	Glycine max	Gma	21	25	11
	Cucumis sativus	Csa	6	11	3
	Prunus persica	Ppe	6	13	6
	Arabidopsis thaliana	Ath	10	13	13
	Carica papaya	Cpa	5	9	3
	Vitis vinifera	Vvi	7	12	5
	Sorghum bicolor	Sbi	9	10	7
	Zea mays	Zma	10	20	14
	Oryza sativa	Osa	11	10	8
	Brachipodyum	Bdi	7	13	10

	distachyon				
	Selaginella mollendorffii	Smo	1	5	3
	Physcomitrella patens	Ppa	1	6	6
Heterokonta	*Phaeodactylum tricornutum*	Ptri	1	1	1

Table 1. NF-Y genes identified in the fully eukaryotic sequenced genomes.

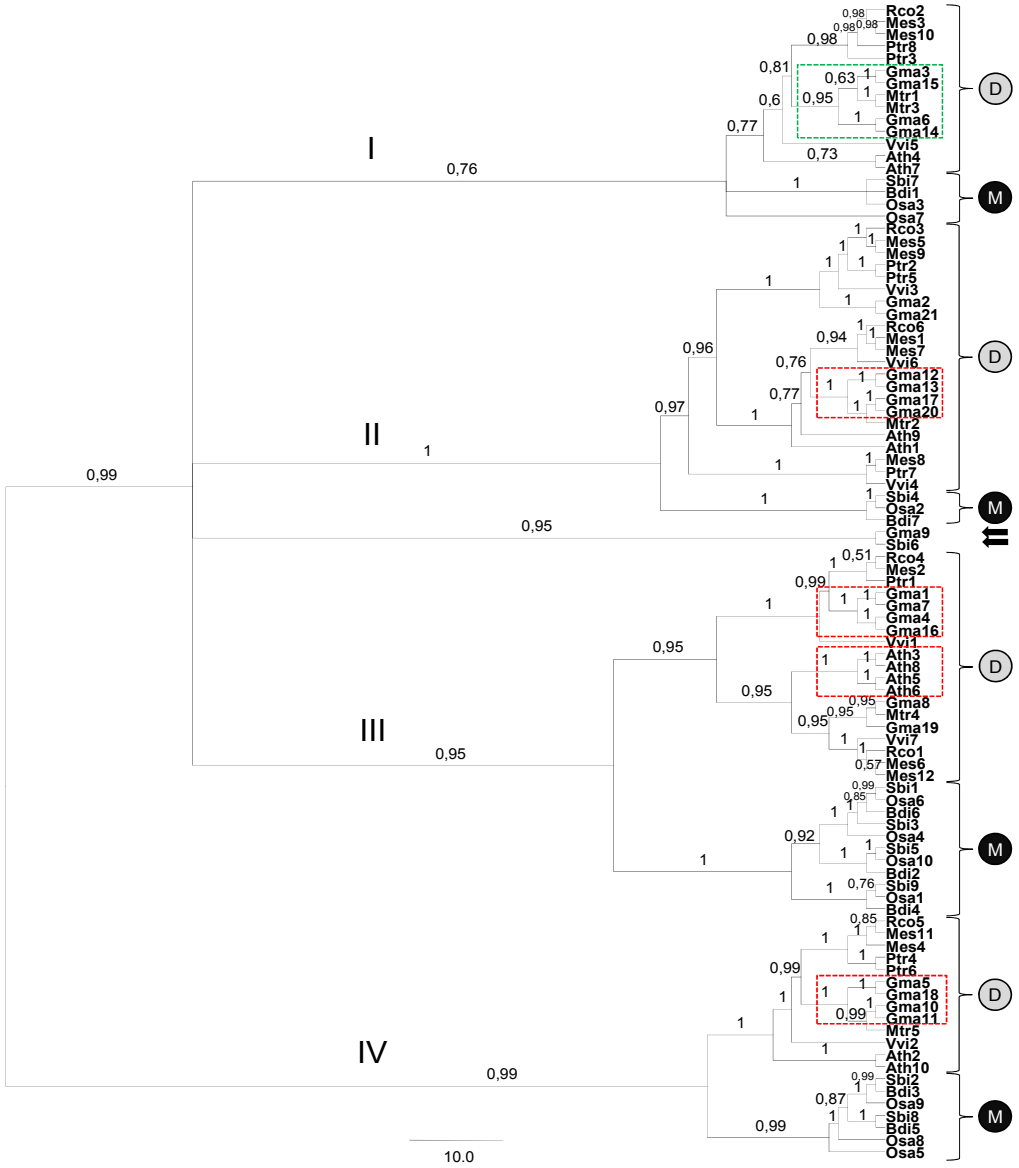


Fig. 4. Phylogenetic tree of monocot and eudicot representatives of NF-YA subunit.M: monocots; D: eudicots; Rco: *Ricinus communis*; Mes: *Manihot esculenta*; Ptr: *Populus tricocharpa*; Gma: *Glycine max*; Mtr: *Medicago truncatula*; Vvi: *Vitis vinifera*; Ath: *Arabidopsis thaliana*; Sbi: *Sorghum bicolor;* Bdi: *Brachipodyum distachyon*; Osa: *Oryza sativa*; red square: event of duplication inside the specie; green square: event of duplication inside the same plant family; black arrows: genes that possess an unresolved position in the phylogenetic tree; I to IV: independent phylogenetic gene clusters.

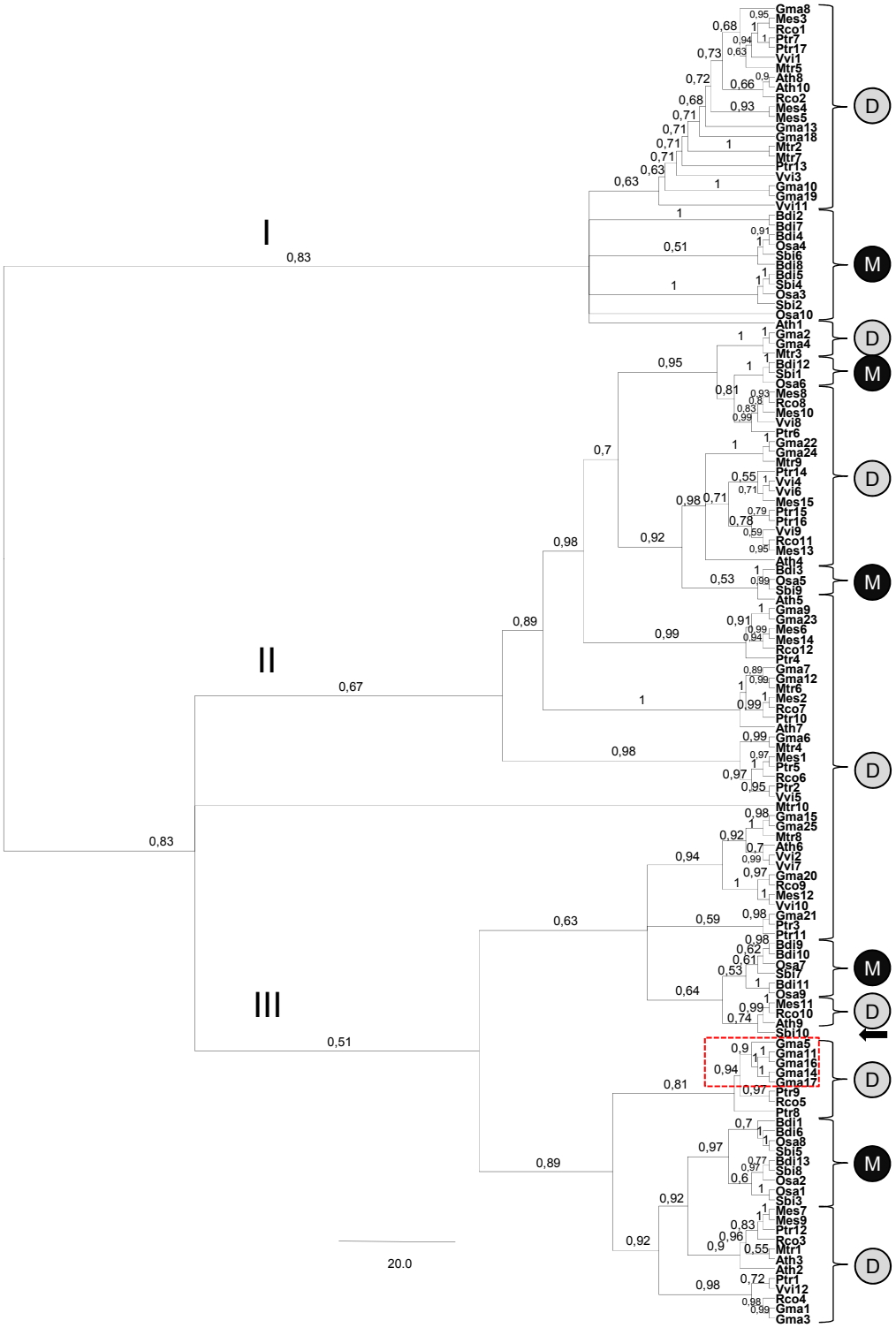


Fig. 5. Phylogenetic tree of monocot and eudicot representatives of NF-YB subunit. For details see legend of Figure 4.

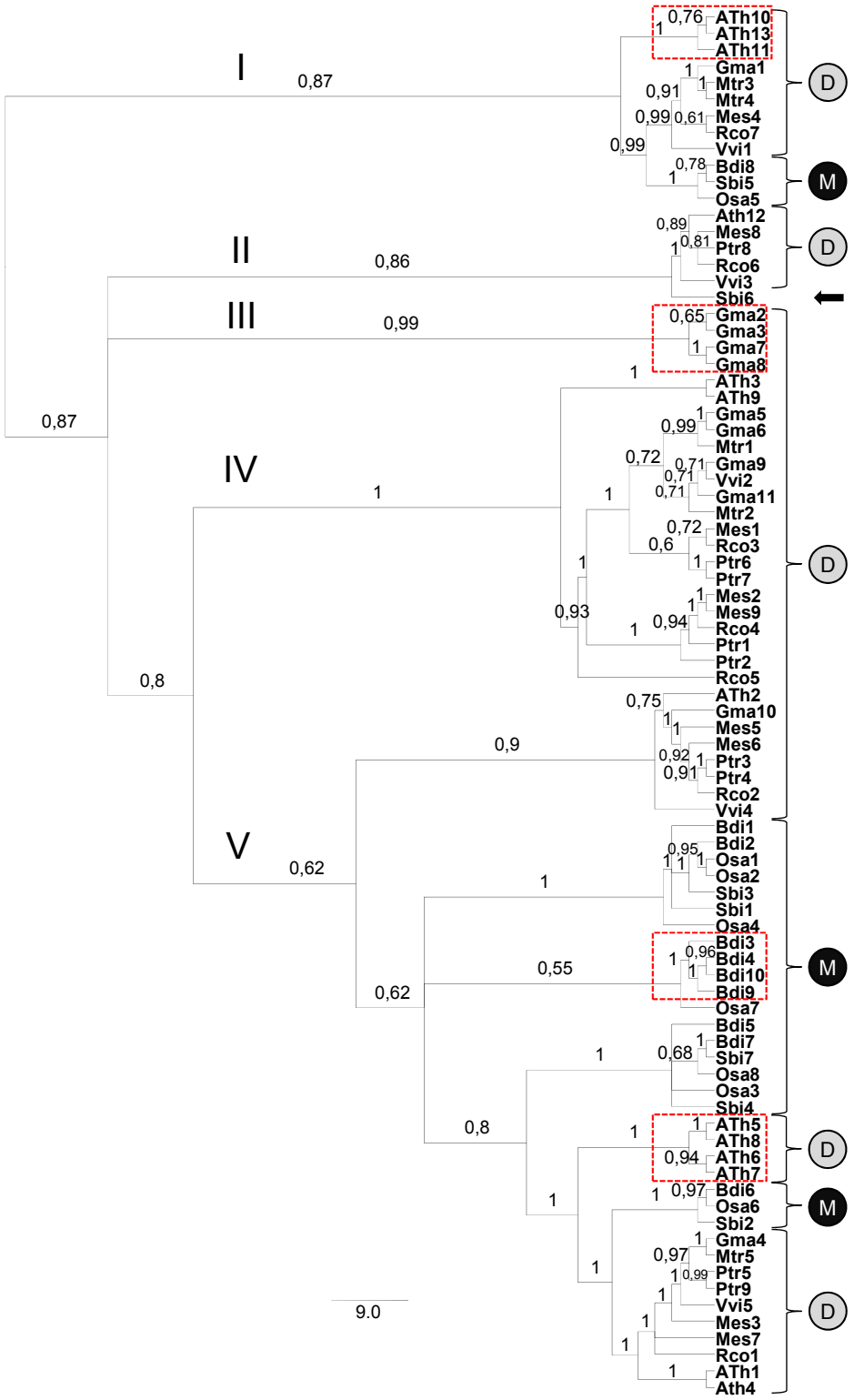


Fig. 6. Phylogenetic tree of monocot and eudicot representatives of NF-YC subunit. For details see legend of Figure 4.

Phylogenetic analysis showed that the gene diversification of all NF-Y subunits likely resulted from several duplication events along evolution and diversification of plants (Figure 4 to 6). It was possible to observe the formation of four independent highly supported clusters for the NF-YA subunit (I to IV, Figure 4), three for NF-YB (I to III, Figure 5) and five for NF-YC (I to V, Figure 6). Based on these results, we suggest that each cluster might possess an independent ancestral subunit that the duplicated members of each group originated from. However, independent duplication events have occurred in many species after the divergence of monocots and eudicots. For example, the soybean and Arabidopsis genomes have experienced a series of recent duplication events (red squares in figures 4 to 6) that could be the result of chromosome duplication or could be derived from polyploidization events (soybean is a good example of a recent polyploidization). These duplications can help us to explain the differences observed in the number of genes coding for the NF-Y subunits in plants (Table 1). Additionally, these duplications seem to be relatively recent and can provide the raw material for neofunctionalization (Figure 2b) and functional divergence of duplicated genes (Figure 3). With few exceptions (genes that possess an unresolved position in the phylogenetic tree are plotted with black arrows, Figures 4 to 6), all clusters of a specific NF-Y subunit are formed by well-defined sub-clusters of monocot and eudicot representatives (Figure 4 to 6). Events of duplication inside a specific plant family were also observed between the two fabaceae species *Glycine max* and *Medicago truncatula* (green square, Figure 4), which could indicate concerted evolution of duplicated genes between these related species (Figure 2e). This is similar to the clade-specific shifts in selective constraint following concerted duplication events observed for MADS box transcription factors in angiosperms (Shan *et al.*, 2009).

The duplication process is a prominent feature of plant genomic architecture (Figure 1). This has led many researchers to speculate that gene duplication may have played an important role in the evolution of phenotypic novelty within the plant lineage (Flagel and Wendel, 2009). As a result of pervasive and recurring small-scale duplications, which may be followed by functional divergence, many nuclear genes in plants are members of gene families and may exhibit copy number variation lineages (Blanc and Wolfe, 2004; Schlueter *et al.*, 2004), as can be observed in table 1.

Evidence for frequent gene duplication has also been observed in the evolutionary history of numerous gene families that have expanded during the diversification of the angiosperms (De Bodt *et al.*, 2005; Zahn *et al.*, 2005; Duarte *et al.*, 2010). In multigene families descended from a common ancestor, individual genes in the group exert similar functions and have similar DNA sequences (Conrad and Antonarakis, 2007). One concept, concerted evolution, applies particularly to localized and typically tandem copies of a gene. The concept posits that all genes in a given group evolve coordinately, and that homogenization is the result of gene conversion (Figure 2e) (Conrad and Antonarakis, 2007).

The emerging picture points to plant NF-Y complexes acting as essential regulatory hubs for many processes. Multiple NF-Y subunits in vascular plants may associate with each other in various combinations that regulate the expression of specific gene sets and might provide similar levels of combinatorial diversity for transcriptional fine-tuning (Siefers *et al.*, 2009).The amplification observed in the plant lineage (Table 1) raises the possibility that new and divergent functions of heterotrimeric complexes have evolved in plants (Nelson *et al.*, 2007) indicating a more complex regulatory role for the various NF-Y proteins in plants than in other organisms (Stephenson *et al.*, 2007).

The existence of multiple genes for each subunit in the plant genome indicates that the specificity of subunit interaction may be determined by preferential protein–protein interaction, tissue or cell-specific expression of each gene or a combination of both (Yazawa and Kamada, 2007). The large number of possible combinations has hindered the analysis of plant NF-Y complexes and suggests that they might act in a more intricate system than in vertebrates and yeast, which have only one gene that encodes each HAP subunit (Yazawa and Kamada, 2007). Additionally, the multiple copies for each NF-Y subunit raises a question if a specific NF-Y subunit interacts with any other two NF-Y subunits or if the NF-Y subunit interacts with only specific member(s) of the other two subunits (Thirumurugan *et al.*, 2008).

Although the presence of many genes encoding NF-Y subunits suggests a high degree of genetic redundancy in plants, the analysis of mutants in single NF-Y genes in Arabidopsis has been associated with defects in development and enhanced stress sensitivity, suggesting a specialized function for each member (Lotan *et al.*, 1998; Kwong *et al.*, 2003; Lee *et al.*, 2003; Zanetti *et al.*, 2010). This could indicate that duplicated genes have passed through a neofunctionalization process (Figure 2b).

Some proteins may require several key substitutions before acquiring a new function, while others may be more mutationally labile. An example includes the terpene synthase gene family in Norway spruce (*Picea abies*). These genes appear to have undergone repeated rounds of neofunctionalization (Figure 2b) (Keeling *et al.*, 2008) and a small number of key amino acid substitutions among paralogs was sufficient to alter the substrate specificity and terpenoid product profiles (Flagel and Wendel, 2009). Another example of neofunctionalization (Figure 2b) in plants is observed in Arabidopsis, where a specific amino acid residue identified in LEC1 and LEC1-LIKE (L1L) is responsible for differentiating their functions (seed development) from those of other NF-YB members (Kwong *et al.*, 2003; Lee *et al.*, 2003; Yamamoto *et al.*, 2009). In addition, the analysis of amino acid substitution rates in plants has been appointed for the asymmetric evolution of certain duplicates of NF-YB and NF-YC subunits, which appears to be coupled with the asymmetric divergence in gene function (Yang *et al.*, 2005; Yamamoto *et al.*, 2009).

With respect to expression patterns, the Arabidopsis NF-Y gene family presents some members that are ubiquitously expressed and others that are tissue specific or induced only after the switch to reproductive growth in flowers and siliques (Gusmaroli *et al.*, 2001; 2002; Yazawa and Kamada, 2007). The difference observed in the expression pattern of these genes could represent an example of *cis*-regulatory divergence (Figure 3c), where the *cis*-element of gene evolves independently from the other members of gene family, and becomes regulated by different stimuli and/or *trans*-activators.

Because genes that harbor NF-Y binding domains include genes that are constitutive, inducible, and cell-cycle-dependent, the regulation of the expression of these genes cannot be exclusively due to NF-Y binding to DNA. In this scenario, the interaction with other transcription factors, either functionally or physically, will contribute to the NF-Y action (Matuoka and Chen, 2002). In addition, the independent evolution of protein-binding domains present in duplicated gene architecture can contribute to protein network divergence (Figure 3b2), increasing the numbers of possible interacting partners of NF-Y genes.

When compared with other forms of mutation, a notable feature of duplication is that it creates genetic redundancy. This redundancy fosters evolutionary innovation, creating the opportunity for duplicates to explore new evolutionary terrain (Flagel and Wendel, 2009).

The most important contribution of gene duplication to evolution is to supply new genetic material for mutation, drift and selection to act upon. This leads to the creation of new genes and new gene functions (Hurley *et al*., 2005; Woollard, 2005; Schmidt and Davies, 2007), two important factors in the origin of genomic and organismal complexity (Gu *et al*., 2002; Taylor and Raes, 2004; Sterck *et al*., 2007).

The plasticity of a genome or species in adapting to environmental changes would be severely limited without gene duplication, because no more than two variants (alleles) exist at any locus within a diploid individual. A good example is the dozens of duplicated immunoglobulin genes that constitute the vertebrate adaptive immune system. It seems difficult to imagine how this system could have acquired this high complexity level without gene duplication (Zhang, 2003).

Plant gene families are largely conserved even over evolutionary time scales that encompass the diversification of all angiosperms and nonflowering plants (Rensing *et al*., 2008). This property of plant genomes indicates that plants have not created new gene families, but have been endowed with a basic genetic toolkit of ancient origin. Despite the evolutionary conservation of gene families, lineage-specific fluctuations in gene family size are frequently observed among taxa (Velasco *et al*., 2007; Ming *et al*., 2008; Rensing *et al*., 2008), which suggests that the diversity and lineage-specific phenotypic variation observed in land plants may not be explained by an equally diverse set of entirely novel genes. Indeed, much of plant diversity may have arisen from the duplication and adaptive specialization processes of pre-existing genes (neofunctionalization and subfunctionalization, Figure 2b and c, respectively). This perspective assigns gene duplication a central role in plant diversification, being a key process that generates the raw material necessary for adaptive evolution (Flagel and Wendel, 2009).

6. Conclusions

Whereas various classes of structural and metabolic genes preferentially return to a single copy state following whole-genome duplication (Paterson *et al*., 2010), transcription factors tend to be preferentially retained among the duplicated genes in *A. thaliana* (Flagel and Wendel, 2009). Our findings support the hypotheses that this preference seems to be true for all plant species, based on the number of genes identified for each NF-Y subunit. Certainly, further studies encompassing functional assays are required to ascertain the role of these genes in plant metabolism.

The number of interacting partners in a molecular network (connectivity) of a particular gene also influences the probability of duplication gene retention (Flagel and Wendel, 2009). In this scenario, the high number of genes coding for the three subunits of NF-Y transcription factor in higher plants leads to numerous interaction possibilities among different genes of each subunit and among these genes and other transcription factors what could contribute to gene retention of the NF-Y transcription factor family in plants.

7. Acknowledgement

This work was supported by a CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), FAPERGS (Fundação de Amparo a Pesquisa do Estado do Rio Grande do Sul), FINEP (Financiadora de Projetos) and MCT (Ministério de Ciência e Tecnologia). A. Cagliari

received a Ph.D. fellowship from CNPq. and G. Loss a Ph.D. fellowship from CAPES. A. Turchetto-Zolet and F. Maraschin have a PNPD-CAPES fellowship and M. Margis-Pinheiro and R. Margis are recipients of CNPq research fellowships number 308708/2006-7 and 303967/2008-0, respectively.

8. References

Arents, G. and Moudrianakis, E.N. (1995) The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc Natl Acad Sci U S A*, 92, 11170-11174.

Baxevanis, A.D., Arents, G., Moudrianakis, E.N. and Landsman, D. (1995) A variety of DNA-binding and multimeric proteins contain the histone fold motif. *Nucleic Acids Res*, 23, 2685-2691.

Beisswanger, S. and Stephan, W. (2008) Evidence that strong positive selection drives neofunctionalization in the tandemly duplicated polyhomeotic genes in Drosophila. *Proc Natl Acad Sci U S A*, 105, 5447-5452.

Ben-Naim, O., Eshed, R., Parnis, A., Teper-Bamnolker, P., Shalit, A., Coupland, G., Samach, A. and Lifschitz, E. (2006) The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA. *Plant J*, 46, 462-476.

Berry, M., Grosveld, F. and Dillon, N. (1992) A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin. *Nature*, 358, 499-502.

Blanc, G. and Wolfe, K.H. (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell*, 16, 1667-1678.

Bowers, J.E., Chapman, B.A., Rong, J. and Paterson, A.H. (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, 422, 433-438.

Braybrook, S.A. and Harada, J.J. (2008) LECs go crazy in embryo development. *Trends Plant Sci*, 13, 624-630.

Bucher, P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J Mol Biol*, 212, 563-578.

Cai, X., Ballif, J., Endo, S., Davis, E., Liang, M., Chen, D., DeWald, D., Kreps, J., Zhu, T. and Wu, Y. (2007) A putative CCAAT-binding transcription factor is a regulator of flowering timing in Arabidopsis. *Plant Physiol*, 145, 98-105.

Caretti, G., Motta, M.C. and Mantovani, R. (1999) NF-Y associates with H3-H4 tetramers and octamers by multiple mechanisms. *Mol Cell Biol*, 19, 8591-8603.

Ceribelli, M., Dolfini, D., Merico, D., Gatta, R., Vigano, A.M., Pavesi, G. and Mantovani, R. (2008) The histone-like NF-Y is a bifunctional transcription factor. *Mol Cell Biol*, 28, 2047-2058.

Chodosh, L.A., Baldwin, A.S., Carthew, R.W. and Sharp, P.A. (1988) Human CCAAT-binding proteins have heterologous subunits. *Cell*, 53, 11-24.

Combier, J.P., Frugier, F., de Billy, F., Boualem, A., El-Yahyaoui, F., Moreau, S., Vernie, T., Ott, T., Gamas, P., Crespi, M. and Niebel, A. (2006) MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula. *Genes Dev*, 20, 3084-3088.

Conant, G.C. and Wagner, A. (2002) GenomeHistory: a software tool and its application to fully sequenced genomes. *Nucleic Acids Res*, 30, 3378-3386.

Conrad, B. and Antonarakis, S.E. (2007) Gene duplication: a drive for phenotypic diversity and cause of human disease. *Annu Rev Genomics Hum Genet*, 8, 17-35.

Coustry, F., Maity, S.N., Sinha, S. and de Crombrugghe, B. (1996) The transcriptional activity of the CCAAT-binding factor CBF is mediated by two distinct activation domains, one in the CBF-B subunit and the other in the CBF-C subunit. *J Biol Chem*, 271, 14485-14491.

Coustry, F., Hu, Q., de Crombrugghe, B. and Maity, S.N. (2001) CBF/NF-Y functions both in nucleosomal disruption and transcription activation of the chromatin-assembled topoisomerase IIalpha promoter. Transcription activation by CBF/NF-Y in chromatin is dependent on the promoter structure. *J Biol Chem*, 276, 40621-40630.

Currie, R.A. (1998) NF-Y is associated with the histone acetyltransferases GCN5 and P/CAF. *J Biol Chem*, 273, 1430-1434.

Dang, V.D., Bohn, C., Bolotin-Fukuhara, M. and Daignan-Fornier, B. (1996) The CCAAT box-binding factor stimulates ammonium assimilation in Saccharomyces cerevisiae, defining a new cross-pathway regulation between nitrogen and carbon metabolisms. *J Bacteriol*, 178, 1842-1849.

Davis, J.C. and Petrov, D.A. (2004) Preferential duplication of conserved proteins in eukaryotic genomes. *PLoS Biol*, 2, E55.

De Bodt, S., Maere, S. and Van de Peer, Y. (2005) Genome duplication and the origin of angiosperms. *Trends Ecol Evol*, 20, 591-597.

De Grassi, A., Lanave, C. and Saccone, C. (2008) Genome duplication and gene-family evolution: the case of three OXPHOS gene families. *Gene*, 421, 1-6.

de Silvio, A., Imbriano, C. and Mantovani, R. (1999) Dissection of the NF-Y transcriptional activation potential. *Nucleic Acids Res*, 27, 2578-2584.

Dorn, A., Bollekens, J., Staub, A., Benoist, C. and Mathis, D. (1987a) A multiplicity of CCAAT box-binding proteins. *Cell*, 50, 863-872.

Dorn, A., Durand, B., Marfing, C., Le Meur, M., Benoist, C. and Mathis, D. (1987b) Conserved major histocompatibility complex class II boxes--X and Y--are transcriptional control elements and specifically bind nuclear proteins. *Proc Natl Acad Sci U S A*, 84, 6249-6253.

Duarte, J.M., Wall, P.K., Edger, P.P., Landherr, L.L., Ma, H., Pires, J.C., Leebens-Mack, J. and dePamphilis, C.W. (2010) Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and Oryza and their phylogenetic utility across various taxonomic levels. *BMC Evol Biol*, 10, 61.

Dynan, W.S. and Tjian, R. (1985) Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature*, 316, 774-778.

Edwards, D., Murray, J.A. and Smith, A.G. (1998) Multiple genes encoding the conserved CCAAT-box transcription factor complex are expressed in Arabidopsis. *Plant Physiol*, 117, 1015-1022.

Flagel, L.E. and Wendel, J.F. (2009) Gene duplication and evolutionary novelty in plants. *New Phytol*, 183, 557-564.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L. and Postlethwait, J. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 151, 1531-1545.

Forsburg, S.L. and Guarente, L. (1989) Identification and characterization of HAP4: a third component of the CCAAT-bound HAP2/HAP3 heteromer. *Genes Dev*, 3, 1166-1178.

Gallagher, C.E., Matthews, P.D., Li, F. and Wurtzel, E.T. (2004) Gene duplication in the carotenoid biosynthetic pathway preceded evolution of the grasses. *Plant Physiol*, 135, 1776-1783.

Gu, Z., Nicolae, D., Lu, H.H. and Li, W.H. (2002) Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet*, 18, 609-613.

Gusmaroli, G., Tonelli, C. and Mantovani, R. (2001) Regulation of the CCAAT-Binding NF-Y subunits in Arabidopsis thaliana. *Gene*, 264, 173-185.

Gusmaroli, G., Tonelli, C. and Mantovani, R. (2002) Regulation of novel members of the Arabidopsis thaliana CCAAT-binding nuclear factor Y subunits. *Gene*, 283, 41-48.

Harrison, P.M., Echols, N. and Gerstein, M.B. (2001) Digging for dead genes: an analysis of the characteristics of the pseudogene population in the Caenorhabditis elegans genome. *Nucleic Acids Res*, 29, 818-830.

Harrison, P.M., Hegyi, H., Balasubramanian, S., Luscombe, N.M., Bertone, P., Echols, N., Johnson, T. and Gerstein, M. (2002) Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. *Genome Res*, 12, 272-280.

He, X.L. and Zhang, J.Z. (2005) Gene complexity and gene duplicability. *Current Biology*, 15, 1016-1021.

Hinman, V.F. and Davidson, E.H. (2007) Evolutionary plasticity of developmental gene regulatory network architecture. *Proc Natl Acad Sci U S A*, 104, 19404-19409.

Holland, P.W., Garcia-Fernandez, J., Williams, N.A. and Sidow, A. (1994) Gene duplications and the origins of vertebrate development. *Dev Suppl*, 125-133.

Huang, H.S., Chen, C.J. and Chang, W.C. (1999) The CCAAT-box binding factor NF-Y is required for the expression of phospholipid hydroperoxide glutathione peroxidase in human epidermoid carcinoma A431 cells. *FEBS Lett*, 455, 111-116.

Hurley, I., Hale, M.E. and Prince, V.E. (2005) Duplication events and the evolution of segmental identity. *Evol Dev*, 7, 556-567.

Jordan, I.K., Wolf, Y.I. and Koonin, E.V. (2004) Duplicated genes evolve slower than singletons despite the initial rate increase. *BMC Evol Biol*, 4, 22.

Keeling, C.I., Weisshaar, S., Lin, R.P. and Bohlmann, J. (2008) Functional plasticity of paralogous diterpene synthases involved in conifer defense. *Proc Natl Acad Sci U S A*, 105, 1085-1090.

Kim, I.S., Sinha, S., de Crombrugghe, B. and Maity, S.N. (1996) Determination of functional domains in the C subunit of the CCAAT-binding factor (CBF) necessary for formation of a CBF-DNA complex: CBF-B interacts simultaneously with both the CBF-A and CBF-C subunits to form a heterotrimeric CBF molecule. *Mol Cell Biol*, 16, 4003-4013.

Kimura, M. (1989) The neutral theory of molecular evolution and the world view of the neutralists. *Genome*, 31, 24-31.

Kondrashov, F.A. and Koonin, E.V. (2004) A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. *Trends Genet*, 20, 287-290.

Kusnetsov, V., Landsberger, M., Meurer, J. and Oelmuller, R. (1999) The assembly of the CAAT-box binding complex at a photosynthesis gene promoter is regulated by light, cytokinin, and the stage of the plastids. *J Biol Chem*, 274, 36009-36014.

Kwong, R.W., Bui, A.Q., Lee, H., Kwong, L.W., Fischer, R.L., Goldberg, R.B. and Harada, J.J. (2003) LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell*, 15, 5-18.

Lee, H., Fischer, R.L., Goldberg, R.B. and Harada, J.J. (2003) Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. *Proc Natl Acad Sci U S A*, 100, 2152-2156.

Li, W.H., Gu, Z., Cavalcanti, A.R. and Nekrutenko, A. (2003) Detection of gene duplications and block duplications in eukaryotic genomes. *J Struct Funct Genomics*, 3, 27-34.

Li, W.X., Oono, Y., Zhu, J., He, X.J., Wu, J.M., Iida, K., Lu, X.Y., Cui, X., Jin, H. and Zhu, J.K. (2008) The Arabidopsis NFYA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance. *Plant Cell*, 20, 2238-2251.

Li, X.Y., Mantovani, R., Hooft van Huijsduijnen, R., Andre, I., Benoist, C. and Mathis, D. (1992) Evolutionary variation of the CCAAT-binding transcription factor NF-Y. *Nucleic Acids Res*, 20, 1087-1091.

Liberati, C., di Silvio, A., Ottolenghi, S. and Mantovani, R. (1999) NF-Y binding to twin CCAAT boxes: role of Q-rich domains and histone fold helices. *J Mol Biol*, 285, 1441-1455.

Lotan, T., Ohto, M., Yee, K.M., West, M.A., Lo, R., Kwong, R.W., Yamagishi, K., Fischer, R.L., Goldberg, R.B. and Harada, J.J. (1998) Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell*, 93, 1195-1205.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, 389, 251-260.

Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science*, 290, 1151-1155.

Lynch, M. and Force, A. (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics*, 154, 459-473.

Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M. and Van de Peer, Y. (2005) Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci U S A*, 102, 5454-5459.

Maity, S.N. and de Crombrugghe, B. (1992) Biochemical analysis of the B subunit of the heteromeric CCAAT-binding factor. A DNA-binding domain and a subunit interaction domain are specified by two separate segments. *J Biol Chem*, 267, 8286-8292.

Maity, S.N., Sinha, S., Ruteshouser, E.C. and de Crombrugghe, B. (1992) Three different polypeptides are necessary for DNA binding of the mammalian heteromeric CCAAT binding factor. *J Biol Chem*, 267, 16574-16580.

Maity, S.N. and De Crombrugghe, B. (1996) Purification, characterization, and role of CCAAT-binding factor in transcription. *Methods Enzymol*, 273, 217-232.

Maity, S.N. and de Crombrugghe, B. (1998) Role of the CCAAT-binding protein CBF/NF-Y in transcription. *Trends Biochem Sci*, 23, 174-178.

Mantovani, R. (1998) A survey of 178 NF-Y binding CCAAT boxes. *Nucleic Acids Res*, 26, 1135-1143.

Mantovani, R. (1999) The molecular biology of the CCAAT-binding factor NF-Y. *Gene*, 239, 15-27.

Marland, E., Prachumwat, A., Maltsev, N., Gu, Z. and Li, W.H. (2004) Higher gene duplicabilities for metabolic proteins than for nonmetabolic proteins in yeast and E. coli. *J Mol Evol*, 59, 806-814.

Masiero, S., Imbriano, C., Ravasio, F., Favaro, R., Pelucchi, N., Gorla, M.S., Mantovani, R., Colombo, L. and Kater, M.M. (2002) Ternary complex formation between MADS-box transcription factors and the histone fold protein NF-YB. *J Biol Chem*, 277, 26429-26435.

Matuoka, K. and Yu Chen, K. (1999) Nuclear factor Y (NF-Y) and cellular senescence. *Exp Cell Res*, 253, 365-371.

Matuoka, K. and Chen, K.Y. (2002) Transcriptional regulation of cellular ageing by the CCAAT box-binding factor CBF/NF-Y. *Ageing Res Rev*, 1, 639-651.

McNabb, D.S., Xing, Y. and Guarente, L. (1995) Cloning of yeast HAP5: a novel subunit of a heterotrimeric complex required for CCAAT binding. *Genes Dev*, 9, 47-58.

Miyoshi, K., Ito, Y., Serizawa, A. and Kurata, N. (2003) OsHAP3 genes regulate chloroplast biogenesis in rice. *Plant J*, 36, 532-540.

Muro, A.F., Bernath, V.A. and Kornblihtt, A.R. (1992) Interaction of the -170 cyclic AMP response element with the adjacent CCAAT box in the human fibronectin gene promoter. *J Biol Chem*, 267, 12767-12774.

Myers, R.M., Tilly, K. and Maniatis, T. (1986) Fine structure genetic analysis of a beta-globin promoter. *Science*, 232, 613-618.

Nelson, D.E., Repetti, P.P., Adams, T.R., Creelman, R.A., Wu, J., Warner, D.C., Anstrom, D.C., Bensen, R.J., Castiglioni, P.P., Donnarummo, M.G., Hinchey, B.S., Kumimoto, R.W., Maszle, D.R., Canales, R.D., Krolikowski, K.A., Dotson, S.B., Gutterson, N., Ratcliffe, O.J. and Heard, J.E. (2007) Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. *Proc Natl Acad Sci U S A*, 104, 16450-16455.

Nicole, M.C., Hamel, L.P., Morency, M.J., Beaudoin, N., Ellis, B.E. and Seguin, A. (2006) MAP-ping genomic organization and organ-specific expression profiles of poplar MAP kinases and MAP kinase kinases. *BMC Genomics*, 7, 223.

Nowak, M.A., Boerlijst, M.C., Cooke, J. and Smith, J.M. (1997) Evolution of genetic redundancy. *Nature*, 388, 167-171.

Ober, D. (2010) Gene duplications and the time thereafter - examples from plant secondary metabolism. *Plant Biol (Stuttg)*, 12, 570-577.

Olesen, J.T. and Guarente, L. (1990) The HAP2 subunit of yeast CCAAT transcriptional activator contains adjacent domains for subunit association and DNA recognition: model for the HAP2/3/4 complex. *Genes Dev*, 4, 1714-1729.

Papp, B., Pal, C. and Hurst, L.D. (2003) Dosage sensitivity and the evolution of gene families in yeast. *Nature*, 424, 194-197.

Paterson, A.H., Freeling, M., Tang, H. and Wang, X. (2010) Insights from the comparison of plant genome sequences. *Annu Rev Plant Biol*, 61, 349-372.

Peng, Y. and Jahroudi, N. (2002) The NFY transcription factor functions as a repressor and activator of the von Willebrand factor promoter. *Blood*, 99, 2408-2417.

Peng, Y. and Jahroudi, N. (2003) The NFY transcription factor inhibits von Willebrand factor promoter activation in non-endothelial cells through recruitment of histone deacetylases. *J Biol Chem*, 278, 8385-8394.

Remenyi, A., Scholer, H.R. and Wilmanns, M. (2004) Combinatorial control of gene expression. *Nat Struct Mol Biol*, 11, 812-815.

Riechmann, J.L. and Ratcliffe, O.J. (2000) A genomic perspective on plant transcription factors. *Curr Opin Plant Biol*, 3, 423-434.

Rieping, M. and Schoffl, F. (1992) Synergistic effect of upstream sequences, CCAAT box elements, and HSE sequences for enhanced expression of chimaeric heat shock genes in transgenic tobacco. *Mol Gen Genet*, 231, 226-232.

Rodin, S.N. and Riggs, A.D. (2003) Epigenetic silencing may aid evolution by gene duplication. *J Mol Evol*, 56, 718-729.

Romier, C., Cocchiarella, F., Mantovani, R. and Moras, D. (2003) The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. *J Biol Chem*, 278, 1336-1345.

Ronquist, F. and Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, 1572-1574.

Roy, B. and Lee, A.S. (1995) Transduction of calcium stress through interaction of the human transcription factor CBF with the proximal CCAAT regulatory element of the grp78/BiP promoter. *Mol Cell Biol*, 15, 2263-2274.

Sandman, K., Krzycki, J.A., Dobrinski, B., Lurz, R. and Reeve, J.N. (1990) HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon Methanothermus fervidus, is most closely related to histones. *Proc Natl Acad Sci U S A*, 87, 5788-5791.

Schlueter, J.A., Dixon, P., Granger, C., Grant, D., Clark, L., Doyle, J.J. and Shoemaker, R.C. (2004) Mining EST databases to resolve evolutionary events in major crop species. *Genome*, 47, 868-876.

Schmidt, E.E. and Davies, C.J. (2007) The origins of polypeptide domains. *Bioessays*, 29, 262-270.

Serra, E., Zemzoumi, K., di Silvio, A., Mantovani, R., Lardans, V. and Dissous, C. (1998) Conservation and divergence of NF-Y transcriptional activation function. *Nucleic Acids Res*, 26, 3800-3805.

Shan, H., Zahn, L., Guindon, S., Wall, P.K., Kong, H., Ma, H., DePamphilis, C.W. and Leebens-Mack, J. (2009) Evolution of plant MADS box transcription factors: evidence for shifts in selection associated with early angiosperm diversification and concerted gene duplications. *Mol Biol Evol*, 26, 2229-2244.

Shi, Q., Gross, K.W. and Sigmund, C.D. (2001) NF-Y antagonizes renin enhancer function by blocking stimulatory transcription factors. *Hypertension*, 38, 332-336.

Sidow, A. (1996) Gen(om)e duplications in the evolution of early vertebrates. *Curr Opin Genet Dev*, 6, 715-722.

Siefers, N., Dang, K.K., Kumimoto, R.W., Bynum, W.E.t., Tayrose, G. and Holt, B.F., 3rd (2009) Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. *Plant Physiol*, 149, 625-641.

Simillion, C., Vandepoele, K., Van Montagu, M.C., Zabeau, M. and Van de Peer, Y. (2002) The hidden duplication past of Arabidopsis thaliana. *Proc Natl Acad Sci U S A*, 99, 13627-13632.

Singh, K.B. (1998) Transcriptional regulation in plants: the importance of combinatorial control. *Plant Physiol*, 118, 1111-1120.

Sinha, S., Maity, S.N., Lu, J. and de Crombrugghe, B. (1995) Recombinant rat CBF-C, the third subunit of CBF/NFY, allows formation of a protein-DNA complex with CBF-A and CBF-B and with yeast HAP2 and HAP3. *Proc Natl Acad Sci U S A*, 92, 1624-1628.

Sinha, S., Kim, I.S., Sohn, K.Y., de Crombrugghe, B. and Maity, S.N. (1996) Three classes of mutations in the A subunit of the CCAAT-binding factor CBF delineate functional domains involved in the three-step assembly of the CBF-DNA complex. *Mol Cell Biol*, 16, 328-337.

Soltis, P.S. (2005) Ancient and recent polyploidy in angiosperms. *New Phytol*, 166, 5-8.

Steidl, S., Papagiannopoulos, P., Litzka, O., Andrianopoulos, A., Davis, M.A., Brakhage, A.A. and Hynes, M.J. (1999) AnCF, the CCAAT binding complex of Aspergillus nidulans, contains products of the hapB, hapC, and hapE genes and is required for activation by the pathway-specific regulatory gene amdR. *Mol Cell Biol*, 19, 99-106.

Stephenson, T.J., McIntyre, C.L., Collet, C. and Xue, G.P. (2007) Genome-wide identification and expression analysis of the NF-Y family of transcription factors in Triticum aestivum. *Plant Mol Biol*, 65, 77-92.

Sterck, L., Rombauts, S., Vandepoele, K., Rouze, P. and Van de Peer, Y. (2007) How many genes are there in plants (... and why are they there)? *Curr Opin Plant Biol*, 10, 199-203.

Stoltzfus, A. (1999) On the possibility of constructive neutral evolution. *J Mol Evol*, 49, 169-181.

Struhl, K. and Moqtaderi, Z. (1998) The TAFs in the HAT. *Cell*, 94, 1-4.

Tasanen, K., Oikarinen, J., Kivirikko, K.I. and Pihlajaniemi, T. (1992) Promoter of the gene for the multifunctional protein disulfide isomerase polypeptide. Functional significance of the six CCAAT boxes and other promoter elements. *J Biol Chem*, 267, 11513-11519.

Taylor, J.S., Van de Peer, Y. and Meyer, A. (2001) Genome duplication, divergent resolution and speciation. *Trends Genet*, 17, 299-301.

Taylor, J.S. and Raes, J. (2004) Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet*, 38, 615-643.

Testa, A., Donati, G., Yan, P., Romani, F., Huang, T.H., Vigano, M.A. and Mantovani, R. (2005) Chromatin immunoprecipitation (ChIP) on chip experiments uncover a widespread distribution of NF-Y binding CCAAT sites outside of core promoters. *J Biol Chem*, 280, 13606-13615.

Thirumurugan, T., Ito, Y., Kubo, T., Serizawa, A. and Kurata, N. (2008) Identification, characterization and interaction of HAP family genes in rice. *Mol Genet Genomics*, 279, 279-289.

Vision, T.J., Brown, D.G. and Tanksley, S.D. (2000) The origins of genomic duplications in Arabidopsis. *Science*, 290, 2114-2117.

Wagner, A. (2000) Decoupled evolution of coding region and mRNA expression patterns after gene duplication: implications for the neutralist-selectionist debate. *Proc Natl Acad Sci U S A*, 97, 6579-6584.

Wapinski, I., Pfeffer, A., Friedman, N. and Regev, A. (2007) Natural history and evolutionary principles of gene duplication in fungi. *Nature*, 449, 54-61.

Warpeha, K.M., Upadhyay, S., Yeh, J., Adamiak, J., Hawkins, S.I., Lapik, Y.R., Anderson, M.B. and Kaufman, L.S. (2007) The GCR1, GPA1, PRN1, NF-Y signal chain mediates both blue light and abscisic acid responses in Arabidopsis. *Plant Physiol*, 143, 1590-1600.

Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierec, J., Samach, A. and Coupland, G. (2006) CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis. *Plant Cell*, 18, 2971-2984.

Wolberger, C. (1998) Combinatorial transcription factors. *Curr Opin Genet Dev*, 8, 552-559.

Woollard, A. (2005) Gene duplications and genetic redundancy in C. elegans. *WormBook*, 1-6.

Xiong, A.S., Peng, R.H., Zhuang, J., Gao, F., Zhu, B., Fu, X.Y., Xue, Y., Jin, X.F., Tian, Y.S., Zhao, W. and Yao, Q.H. (2009) Gene duplication, transfer, and evolution in the chloroplast genome. *Biotechnol Adv*, 27, 340-347.

Yamamoto, A., Kagaya, Y., Toyoshima, R., Kagaya, M., Takeda, S. and Hattori, T. (2009) Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. *Plant J*, 58, 843-856.

Yang, J., Xie, Z. and Glover, B.J. (2005) Asymmetric evolution of duplicate genes encoding the CCAAT-binding factor NF-Y in plant genomes. *New Phytol*, 165, 623-631.

Yang, X., Tuskan, G.A. and Cheng, M.Z. (2006) Divergence of the Dof gene families in poplar, Arabidopsis, and rice suggests multiple modes of gene evolution after duplication. *Plant Physiol*, 142, 820-830.

Yazawa, K. and Kamada, H. (2007) Identification and characterization of carrot HAP factors that form a complex with the embryo-specific transcription factor C-LEC1. *J Exp Bot*, 58, 3819-3828.

Yokoyama, S. and Yokoyama, R. (1989) Molecular evolution of human visual pigment genes. *Mol Biol Evol*, 6, 186-197.

Zahn, L.M., Leebens-Mack, J., DePamphilis, C.W., Ma, H. and Theissen, G. (2005) To B or Not to B a flower: the role of DEFICIENS and GLOBOSA orthologs in the evolution of the angiosperms. *J Hered*, 96, 225-240.

Zanetti, M.E., Blanco, F.A., Beker, M.P., Battaglia, M. and Aguilar, O.M. (2010) A C subunit of the plant nuclear factor NF-Y required for rhizobial infection and nodule development affects partner selection in the common bean-Rhizobium etli symbiosis. *Plant Cell*, 22, 4142-4157.

Zhang, J.Z. (2003) Evolution by gene duplication: an update. *Trends in Ecology & Evolution*, 18, 292-298.

Zwicker, J. and Muller, R. (1997) Cell-cycle regulation of gene expression by transcriptional repression. *Trends Genet*, 13, 3-6.

4.3 Capítulo 3

O presente capítulo foi submetido ao periódico ***Genomics.***

New insights on the evolution of Leafy cotyledon1 (LEC1) type genes in vascular plants

Alexandro Cagliari[1], Andreia Carina Turchetto-Zolet[1,2], Ana Paula Korbers[1], Felipe dos Santos Maraschin[3], Rogerio Margis[1,2,4] and Marcia Margis-Pinheiro[1,2§].

[1]Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Brazil.

[2]Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Brazil.

[3]Departamento de Botânica, Universidade Federal do Rio Grande do Sul, Brazil.

[4]Departamento de Biofísica, Universidade Federal do Rio Grande do Sul, Brazil.

[§]Corresponding author:

Marcia Margis-Pinheiro. Departamento de Genética. Universidade Federal do Rio Grande do Sul. Avenida Bento Gonçalves 9500. Prédio 43.312. CEP 91501-970. Porto Alegre, RS, Brazil. Phone: +55 (51) 33089814. E-mail: marcia.margis@ufrgs.br. Homepage: http://www.ufrgs.br/rnai/ngfp/

ABSTRACT

NF-Y is a conserved oligomeric transcription factor found in all eukaryotes. In plants, this regulator evolved with a broad diversification of the genes coding for its three subunits (NF-YA, NF-YB and NF-YC). The NF-YB members can be divided into Leafy Cotyledon1 (LEC1) and non-LEC1 types. Here we presented a comparative genomic study using phylogenetic analyses to validate an evolutionary model for the origin of LEC-type genes in plants and their emergence from non-LEC1-type genes. We identified LEC1-type members in all vascular plant genomes, but not in amoebozoa, algae, fungi, metazoa and non-vascular plant representatives, which present exclusively non-LEC1-type genes as constituents of their NF-YB subunits. The non-synonymous to synonymous nucleotide substitution rates (Ka/Ks) between LEC1 and non-LEC1-type genes indicates the presence of positive selection acting on LEC1-type members to the fixation of LEC1-specific amino acid residues. The phylogenetic analyses demonstrated that plant LEC1-type genes are evolutionary divergent from the non-LEC1-type genes of plants, fungi, amoebozoa, algae and animals. Our results point to a scenario in which LEC1-type genes have originated in vascular plants after gene expansion in plants. We suggest that processes of neofunctionalization and/or subfunctionalization were responsible for the emergence of a versatile role for LEC1-type genes in vascular plants, especially in seed plants. LEC1-type genes besides being phylogenetic divergent also present different expression profile when compared with non-LEC1-type genes. Altogether, our data provide new insights about the LEC1 and non-LEC1 evolutionary relationship during the vascular plant evolution.

KEY WORDS: Leafy cotyledon1 (LEC1)-type genes evolution, LEC1, LEC1-Like (L1L), Seed maturation, Transcription Factor.

1 INTRODUCTION

Plant seed development can be conceptually divided into three phases: embryogenesis, maturation, and the induction of desiccation tolerance and seed dormancy (Yazawa and Kamada, 2007). Maturation is characterized by the massive accumulation of storage compounds, especially lipids, proteins, starches and sugars. These storage molecules provide energy sources for the establishment of seedlings after germination (Bewley, 1997; Harada, 1997).

In Arabidopsis, seed maturation and dormancy induction are genetically controlled by a network of master regulatory transcription factors, including LEAFY COTYLEDON1 (LEC1) and LEC1-LIKE (L1L). These genes encode proteins that play a central role in seed development as transcriptional regulators of embryogenesis and seed maturation (Giraudat *et al.*, 1992; Gusmaroli *et al.*, 2002; Lotan *et al.*, 1998; Meinke *et al.*, 1994; Parcy *et al.*, 1997; Stone *et al.*, 2008; Wobus and Weber, 1999; Yazawa *et al.*, 2004).

Both LEC1 and L1L are members of a conserved oligomeric transcription factor family found in all eukaryotes named NF-Y (Nuclear Factor of the Y box), which acts as a transcriptional regulator for different sets of genes (Dorn *et al.*, 1987; Stephenson *et al.*, 2007). NF-Y is composed of three different subunits named NF-YA (also known as HAP-2 or CBF-B), NF-YB (HAP3 or CBF-A), and NF-YC (HAP5 or CBF-C). While most eukaryotic genomes have only one or two genes encoding each NF-Y subunit (Maity and de Crombrugghe, 1998; Riechmann and Ratcliffe, 2000) vascular plants NF-Y subunits are generally encoded by gene families (Riechmann and Ratcliffe, 2000).

NF-YB subunit proteins typically consist of three domains: the *N*-terminal A domain, the central B domain, and the *C*-terminal C domain. The B domain is highly conserved across eukaryotes and is required for both DNA binding and interaction with other NF-Y subunits (Sinha *et al.*, 1996; Xing *et al.*, 1993). NF-YB subunit members can be divided into two classes: the LEC1-type and the non-LEC1-type. In Arabidopsis, the LEC1-type is composed of LEC1 and the closely related subunit L1L, while the other members of NF-YB are designated non-LEC1-type genes. This classification is based on the presence of 16 amino acids in the B

domain that are only shared by LEC1 and L1L (Holdsworth *et al.*, 2008; Kwong *et al.*, 2003). The results from domain swapping and site-directed mutagenesis experiments have demonstrated that the B domain is the protein portion that functionally differentiates LEC1-type from non-LEC1-type proteins in the NF-YB family (Lee *et al.*, 2003). Moreover, the *in vivo* function of the Arabidopsis LEC1 cannot be replaced by other NF-YB subunits except L1L (Kwong *et al.*, 2003; Lee *et al.*, 2003; Watanabe and Yamamoto, 2009), demonstrating the distinctive molecular activity of LEC1 and L1L genes. Additionally, while non-LEC1-type genes are ubiquitously expressed in Arabidopsis, the expression of LEC1-type genes is restricted to seeds and developing siliques (Cao *et al.*, 2011; Kwong *et al.*, 2003; Lotan *et al.*, 1998).

Although the importance of LEC1-type genes in transcriptional regulatory networks of seed development has been studied, the evolutionary history and origin of these genes was discussed in a single study (Xie et al. 2008). These authors performed a phylogenetic analysis of LEC1-type genes from a limited number of species (*Arabidopsis thaliana*, *Oryza sativa* and *Selaginella moellendorffii*; and non-LEC1 type of *Physcomitrella patens, Chlamydomonas reinhardtii, Volvox carteri*, and *Cyanidioschyzon merolae)*, and proposed that LEC1-type genes have originated in the vascular plant genome prior to the divergence of seed plants. However, in light of the availability of new sequenced genomes and given the importance of LEC1 and L1L genes in plants, further studies involving a larger number of taxa may be useful to advance in the understanding of unsolved questions about their evolutionary history and the physiological meaning of their evolution.

In order to identify plant homologous coding LEC1-type and non-LEC1-type and to expand the current understanding on the emergence and evolution of LEC1-type genes in seed plants and their evolutionary relationship with non-LEC1-types, we adopted a comparative genomic approach to conduct a broad survey on fully sequenced genomes. Our analyses included representatives of amoebozoa, yeasts, fungi, algae, mosses, plants, vertebrate and invertebrate species, providing a broad representation of eukaryotes. Phylogenetic analyses using

Bayesian inference were reconstructed, and an evolutionary model of the emergence of LEC1-type genes in vascular plants is proposed and discussed.

2 RESULTS AND DISCUSSION

2.1 LEC1-type and non-LEC1-type genes have diverged during plant evolution

We have adopted an extensive data mining approach (see Materials and Methods) focusing on the identification of LEC1 and non LEC1-type genes in all fully sequenced eukaryotic genomes. A set of 29 completely sequenced and 4 partially sequenced genomes (conifers) was investigated.

We identified LEC1-type members in all vascular plant genomes (Table 1), but not in amoebozoa, algae, fungi, metazoa and non-vascular plant representatives, which present exclusively non-LEC1-type genes as constituents of their NF-YB subunits. The number of non-LEC1-type genes identified in plants varied in the analyzed species, ranging from one gene in algae species to more than 20 genes in some angiosperm species (Table 1). A total of 64 LEC1-type genes were identified in the genomes of vascular plants, with gene numbers ranging from 1 to 6 copies in each species (Table 1). We have identified LEC1-type genes in the lycophyte *Selaginella moellendorffii* but not in the moss *Physcomitrella patens,* both important model organisms for comparative genomics (Table 1).

We performed a Bayesian analysis using the complete coding region sequences to further investigate the phylogenetic relationships among the identified plant LEC1-type genes (Figure 1). A total of 165 positions were included in the final dataset. Phylogenetic analysis demonstrated that LEC1-type members grouped in a well-supported clade (Posterior probability =1) in relation to the *S. moellendorffii* and *P. patens* non-LEC1-type genes. The non-LEC1-type gene from the green algae *Volvox carteri* was used as outgroup (Figure 1). One of the five sequences identified in the lycophyte representative (*S. moellendorffii*) contained amino acids exclusive to LEC1-type genes and grouped into the LEC1-type clade by phylogeny (red star, Figure 1). This result is in agreement with the results obtained by Xie et al. (2008) and indicates that the origin of LEC1-type genes

occurred after the emergence of vascular plants, as the non-vascular plant *P. patens* had no LEC1-type sequences.

Within the LEC1-type clade, we identified four main subclades, named groups A, B, C and D (Figure 1). Group A includes the LEC1-type sequences from conifers that clustered apart from angiosperm LEC1-type genes, supported with strong posterior probability. The function of LEC1-type genes in gymnosperms is not well-known and could be related with stress response and desiccation tolerance in vegetative tissues as observed in lycophytes (Xie *et al.*, 2008).

Groups C and D are moderately supported with posterior probabilities of 0.72 and 0.85, respectively, and include sequences from angiosperm species. These groups most likely represent duplication events that occurred during angiosperm evolution. The well characterized Arabidopsis LEC1 (Ath9) and L1L (Ath6) members grouped separately in Clades C and D, respectively (asterisks, Figure 1). Each of these clades includes sequences from other eudicot species, implying that the duplication events occurred during eudicot emergence. Our results suggest that the Arabidopsis LEC1 and L1L genes originated from two independent duplication events, which occurred after the emergence of LEC1-type genes in the vascular plant lineage (Figure 1).

The monocot LEC1-type members (Group B, Figure 1) formed a separate cluster within the L1L Clade (group C) with moderate posterior probability. No monocot LEC1-type genes were observed within LEC1 Clade (group D, figure 1), suggesting that the duplication that caused the Arabidopsis LEC1 and L1L emergence was posterior to the monocot and eudicot divergence. We also observed the formation of two sub clusters within the monocot cluster, which indicates that these species also passed through duplication events that could have originated the monocot LEC1 and L1L genes.

2.2 LEC1-type genes display a high degree of conservation of essential B domain amino acid residues

The classification of NF-YB proteins between LEC1-type and non-LEC1-type members is based on the specific amino acid residues of their central B

domain. The conserved core consensus sequence of the B domain of plant LEC1-type genes is shown in Figure 2A. All 16 amino acid residues, described as exclusive to LEC1-type genes (Holdsworth *et al.*, 2008; Kwong *et al.*, 2003), exhibited a high level of conservation among the species analyzed (arrows in 2A). Moreover, it was previously determined that one specific Aspartate (Asp) residue, which is replaced by a Lysine (K) in non-LEC1-type members, is required for LEC1 activity during embryogenesis and is able to confer partial LEC1 activity (embryogenesis induction) to a non-LEC1-type protein (Asp, red arrow in Figure 2A and 2B) (Lee *et al.*, 2003).

On an evolutionary scale, some proteins may require several key substitutions before acquiring a new function, while others may be more mutationally labile (Keeling *et al.*, 2008). To understand how the evolution of LEC1-type genes occurred we have analyzed the non-synonymous to synonymous nucleotide substitution rates (Ka/Ks) between the B domains of LEC1 and non-LEC1-type genes (Figure 2B) in vascular plant sequences. Despite the high level of conservation between the B domains of NF-YB genes, we have verified that some specific residues, especially Asp, are very conserved in LEC1-type genes but not in non-LEC1-type genes (Figure 2B). This observation is in agreement with the importance of this residue for LEC1-type activity.

A high Ka/Ks ratio between LEC1 and non-LEC1-type members was verified, indicating the presence of positive selection acting on LEC1-type members to the fixation of LEC1-specific residues (arrows in Figure 2B indicate the position of LEC1-specific residues), contributing to the maintenance of the LEC1-type and non-LEC1-type amino acid differences (Figure 2B). On the other hand, we observed a low Pi(a)/Pi(s) ratio (non-synonymous to synonymous nucleotide substitution rates inside each group), indicating that both LEC1 and non-LEC1-type members are under purifying selection (Figure 2B).

2.3 Neofunctionalization and/or subfunctionalization processes as evolutionary raw material for the origin of LEC1-type genes in vascular plants

To gain insight into plant LEC1-type gene emergence, we performed a Bayesian phylogenetic reconstitution including LEC1 and non-LEC1-type genes from representatives of all the analyzed genomes. The analyses demonstrated that plant LEC1-type genes are evolutionary divergent from the non-LEC1-type genes of plants, fungi, amoebozoa, algae and animals (Figure 3A). Our results point to a scenario in which LEC1-type genes have originated in vascular plants after gene expansion in plants, which is in agreement with previously results in which the function of the most ancient LEC1-type gene identified (lycophyte) was confirmed by its ability to functionally complement an Arabidopsis *lec1-1* mutant (Xie *et al.*, 2008). In this context, the NF-Y expansion in the plant lineage, through several independent duplication events into the NF-Y family during the evolution and diversification of plants (Cagliari *et al.*, 2011), provided the raw material from which LEC1-type genes originated.

Based on the fact that that exclusively in plants we observed gene families coding for NF-YB subunit (Riechmann and Ratcliffe, 2000) and that only 16 amino acid residues in the B domain are sufficient to differentiate LEC1-type from non-LEC1-type genes (Holdsworth *et al.*, 2008; Kwong *et al.*, 2003), we postulate that LEC1-type genes originated from an ancient duplicated non-LEC1-type gene (present in a primitive eukaryotic organism, Figure 3B) through the following: 1) a neofunctionalization process, in which one duplicated copy may have acquired a novel evolutionarily advantageous function that was preserved by natural selection, while the other copy retained its original function (De Grassi *et al.*, 2008) and/or 2) a subfunctionalization process in which, after duplication, mutations may have occurred in both duplicated genes, leading to specialization that enabled the genes to perform complementary functions (Figure 8B) (Lynch and Conery, 2000; Lynch and Force, 2000). In this context, LEC1-type neofunctionalization and/or subfunctionalization would be responsible for conferring the specific expression patterns and functions observed in LEC1-type genes, which differ from the expression patterns and functions of non-LEC1-type genes in Arabidopsis and

other vascular plants. An example of this type of genetic diversification can be observed in the terpene synthase gene family of the Norway spruce (*Picea abies*) in which a small number of key amino acid substitutions among paralogs was sufficient to alter their substrate specificity and terpenoid product profiles (Flagel and Wendel, 2009; Keeling *et al.*, 2008).

Even if we just consider LEC1-type genes we also observed that they have been passed through a neofunctionalization and/or subfunctionalization, since they acquired different functions in vascular plants, being involved with embryogenesis and seed maturation in seed plants (Kwong *et al.*, 2003; Lee *et al.*, 2003; Lotan *et al.*, 1998; Watanabe and Yamamoto, 2009; Yu *et al.*, 2011) and desiccation tolerance in vegetative tissues in non-seeds plants (Xie *et al.*, 2008). It was proposed that the expression of LEC1-type genes under drought stress in non-seed plants was recruited to play novel functions in early stages of seed plant evolution, being expressed during seed development and maturation (Xie *et al.*, 2008).

To test this hypothesis we run out an expression analysis of LEC1 and non-LEC1-type genes of monocot (rice and maize) and eudicot (Arabidopsis and soybean) representatives using the Genevestivator database (https://www.genevestigator.com) to gain insights about the expression profile of these genes. We observed that the expression of LEC1-type genes, when detected in the available conditions of Genevestigator database, was restricted to seed compartments (or structures that contain seeds, e.g. inflorescence), while the expression of non-LEC1-type genes was spread throughout different organs and tissues (Figure 4). Altogether, these results indicate that the phylogenetic divergence observed for LEC1-type also reflect a different and exclusive seed expression profile when compared with non-LEC1-type genes.

In an evolutionary perspective, the emergence of LEC1-type genes could have conferred selective advantages to vascular plants that were important to the spread of these organisms throughout different habitats and environmental conditions. In this scenario, the emergence of LEC1-type genes in vascular plants and their posterior neofunctionalization/subfunctionalization represented an

important step in the control of seed maturation process and, consequently could have contributed for the reproductive success of seed plants.

In conclusion, we present a comparative genomic study using robust phylogenetic analyses including a vast number of taxa to validate an evolutionary model for the origin of LEC-type genes in plants and to help to elucidate the LEC1 and L1L relationship. We proposed a model in which the LEC1-type genes originated from an ancient Eukaryotic non-LEC1-type gene after duplication events occurred in plants and the emergence of a different role for LEC1-type genes, especially during seed development in vascular plants occurred through a process of neofunctionalization and/or subfunctionalization.

3 MATERIAL AND METHODS

3.1 Sequence search

Exhaustive data mining searches were performed using Arabidopsis LEC1 and L1L coding sequences as queries (blastp, blastx and tblastx) against Protein and Genome databases with the default parameters and an e-value threshold of 1.0 E-50 at NCBI (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov), the completed genome projects database at the JGI (Joint Genome Institute http://www.jgi.doe.gov) and the phytozome (http://www.phytozome.net/) to identify LEC1-type genes in fully sequenced genomes, including representatives of amoebozoa, yeast, fungi, algae, mosses, plants, and vertebrate and invertebrate species. Due to the lack of complete sequenced genomes for conifers, we performed an additional search on the incomplete genome of conifers that were represented in NCBI. A gene was classified as a LEC1-type if it possessed the majority of LEC1-type specific residues in its conserved histone-fold motif.

3.2 Sequence alignment and phylogenetic analysis

Nucleotide and protein sequences were aligned and inspected using MUSCLE (Edgar, 2004) implemented in Molecular Evolutionary Genetics Analysis (MEGA version 5.0) (Tamura *et al.*, 2011). Taxa terminologies are abbreviated using the first letter of the genus and two letters of the species name (e.g., Rco corresponds to *Ricinus communis*).

The phylogenetic analyses were reconstructed with either complete protein sequences or with the conserved central B domain of NF-YB genes. Bayesian analysis was conducted in MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). The mixed amino acid substitution model plus gamma and invariant sites was used in two independent runs of 20,000,000 generations each with two Metropolis-coupled Monte Carlo Markov chains (MCMC) that were run in parallel (starting each from a random tree). Markov chains were sampled every 100 generations, and the first 25% of the trees were discarded as burn-in. The remaining trees were used to

compute the majority rule consensus tree, the posterior probability of clades and branch lengths.

The pairwise synonymous (Ks) and non-synonymous (Ka) numbers of substitutions corrected for multiple hits were calculated using the DnaSP software (DNA polymorphism analysis) (Librado and Rozas, 2009).

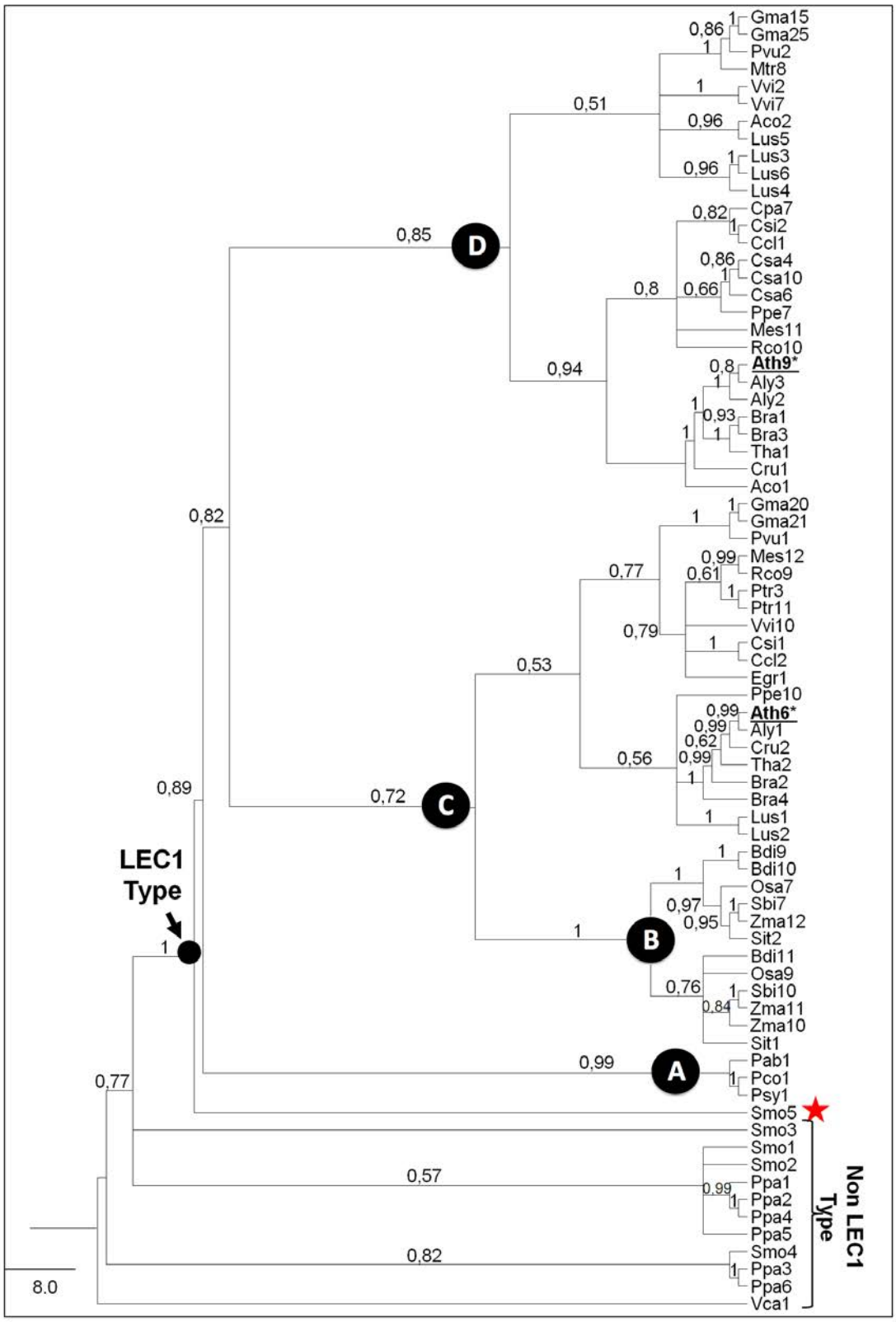


Figure 1. Phylogenetic relationships between plant LEC1-type genes reconstructed by the Bayesian method. All the 64 protein sequences from LEC1-type genes identified in vascular plants were used for Bayesian analysis. The non-

LEC1-type protein sequences from moss (*Physcomitella patens)* and lycophyte (*Selaginella mollendorffii)* were included in the analysis, and an algae (*Volvox carteri)* non-LEC1 sequence was used as an outgroup. The posteriori probabilities are labeled above the branches. Taxa terminologies are abbreviated using the first letter of the genus and two letters of the species name: *Manihot esculenta* (Mes), *Ricinus communis* (Rco), *Linum usitatissimum* (Lus), *Populus trichocarpa* (Ptr), *Medicago truncatula* (Mtr), *Phaseolus vulgaris* (Pvu), *Glycine max* (Gma), *Cucumis sativus* (Csa), *Prunus persica (Ppe), Arabidopsis thaliana* (Ath)*, Arabidopsis lyrata* (Aly), *Capsella rubella* (Cru), *Brassica rapa* (Bra), *Thellungiella halophila* (Tha), *Carica papaya* (Cpa), *Citrus sinensis* (Csi), *Citrus clementina* (Ccl), *Eucalyptus grandis* (Egr), *Vitis vinifera* (Vvi), *Aquilegia coerulea* (Aco), *Sorghum bicolor* (Sbi), *Zea mays* (Zma), *Setaria italica* (Sit), *Oryza sativa* (Osa), *Brachypodium distachyon* (Bdi), *Pinus Sylvestris (Psy), Picea abies* (Pab)*, Pinus contorta* (Pco)*, Selaginella mollendorffii* (Smo)*, Physcomitella patens* (Ppa*)* and *Volvox carteri* (Vca). Red star represents the *S. moellendorffii* LEC1-type gene. Black circle A includes the LEC1-type sequences from conifers. Black circles B and C include sequences from angiosperm species. The well characterized Arabidopsis L1L (Ath6) and LEC1 (Ath9) members grouped separately in Clades B and C and are represented by asterisks.

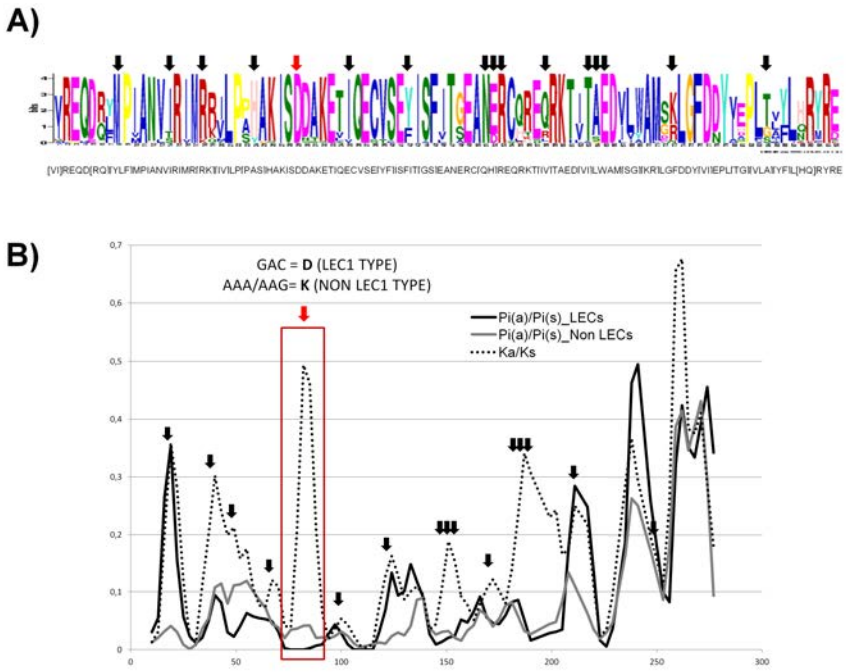


Figure 2. Selection pressure acting on the NF-YB conserved B domain. **A)** A LEC1-type conserved core consensus sequence logo was created with all sequences collected from fully sequenced genomes. Below the logos is a text representation of the majority amino acid consensus sequence created from the species. The overall height in each stack indicates the sequence conservation at that position; the height of each residue letter indicates the relative frequency of the corresponding residue. Amino acids are colored according to their chemical properties: green for polar, non-charged, non-aliphatic residues (NQST), blue for most hydrophobic (A, C, F, I, L, V, W and M), red for positively charged (KR), magenta for acidic (DE), pink for histidine (H), orange for glycine (G), yellow for proline (P) and turquoise for tyrosine (Y) amino acids. Black arrows indicate LEC1-type conserved residues. The red arrow indicates the crucial Asp residue. B**)** Estimation of Ka/Ks rates of the plant NF-YB conserved domain. Comparison of substitution rates (y axis) performed between full-length cDNA sequences of plant LEC1 and non-LEC1-type sequences (x axis). Black arrows indicate LEC1-type conserved residues. The red arrow indicates the crucial Asp residue.

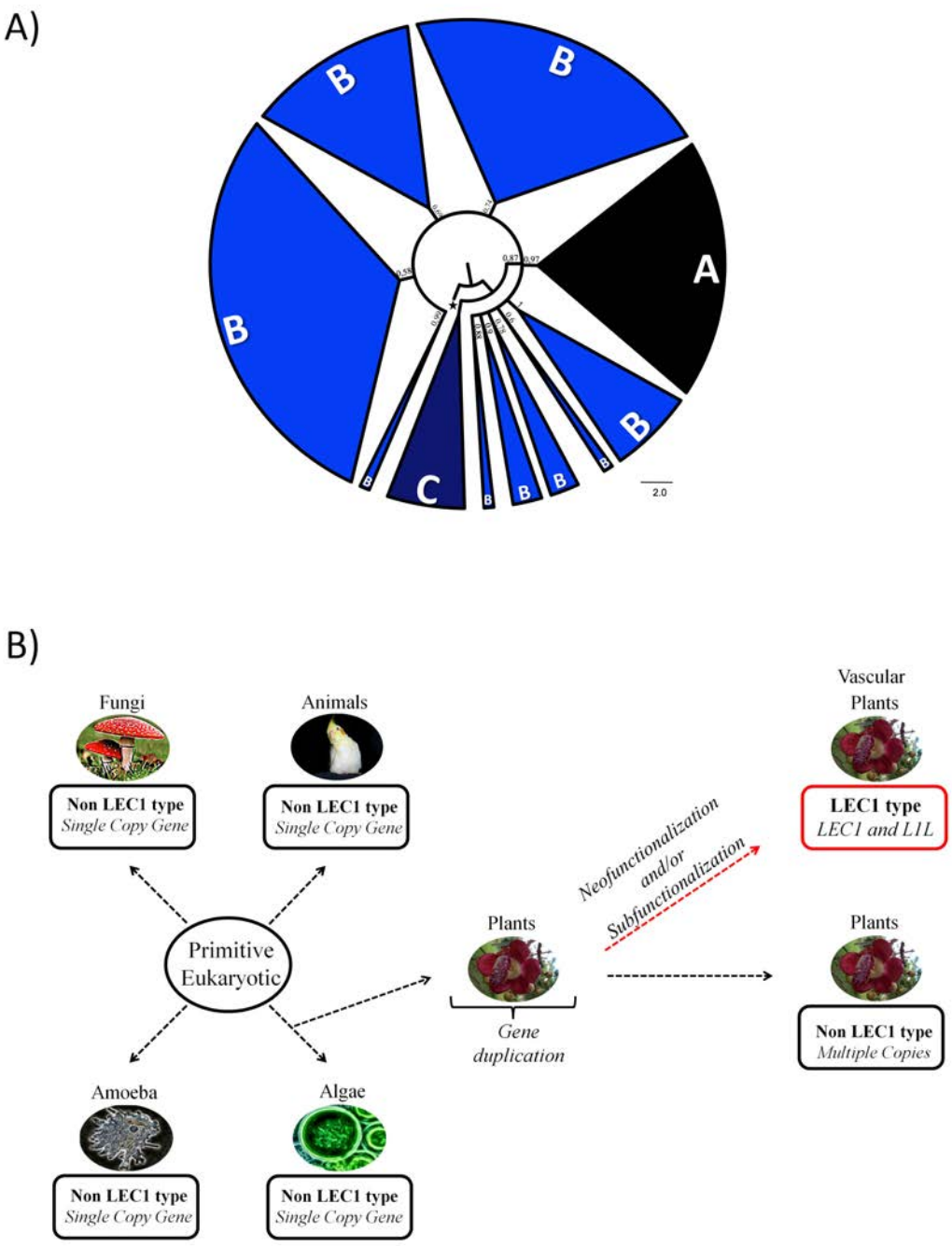


Figure 3. Evolutionary divergence between LEC1 and non-LEC1-type genes. A) LEC1-type genes (A - black) are evolutionary divergent from non-LEC1-type genes of vascular (B - gray) and non vascular plants and algae (C - gray) and fungi and animals (D - light gray). A total of 165 protein sequences representing the fully sequenced eukaryotic genomes were included in the Bayesian analyses. The non-LEC1-type genes from the moss *Physcomitella patens,* the lycophyte *Selaginella mollendorffii* and the algae *Volvox carteri* (C) were included in the analysis, and an amoebozoa (*Dictyostelium purpureum,* black star) non-LEC1 type sequence was used as an

outgroup. The posteriori probabilities are labeled on the collapsed branches. The plant representative species include *Manihot esculenta*, *Ricinus communis*, *Populus trichocarpa*, *Medicago truncatula*, *Glycine max*, *Cucumis sativus*, *Prunus persica, Arabidopsis thaliana*, *Sorghum bicolor*, *Oryza sativa*, *Pinus Sylvestris, Picea abies, Pinus contorta.* The fungi and animal species include *Candida tropicalis, Aspergillus nidulans, Caenorhabditis elegans, Anopheles gambiae, Danio rerio, Mus musculus, Xenopus tropicalis* and *Gallus gallus.*

B) Evolutionary model proposed for LEC1-type genes in vascular plants. LEC1-type genes originated from an ancient non-LEC1-type present in a primitive eukaryote and were maintained as a single non-LEC1 copy gene in amoebozoa, animals, fungi and algae. Gene expansion events resulted in multiple copies of non-LEC1-type genes in plants. LEC1-type genes originated from an ancient duplication of a non-LEC1-type gene through neofunctionalization and/or subfunctionalization processes, becoming preserved by natural selection in vascular plants.

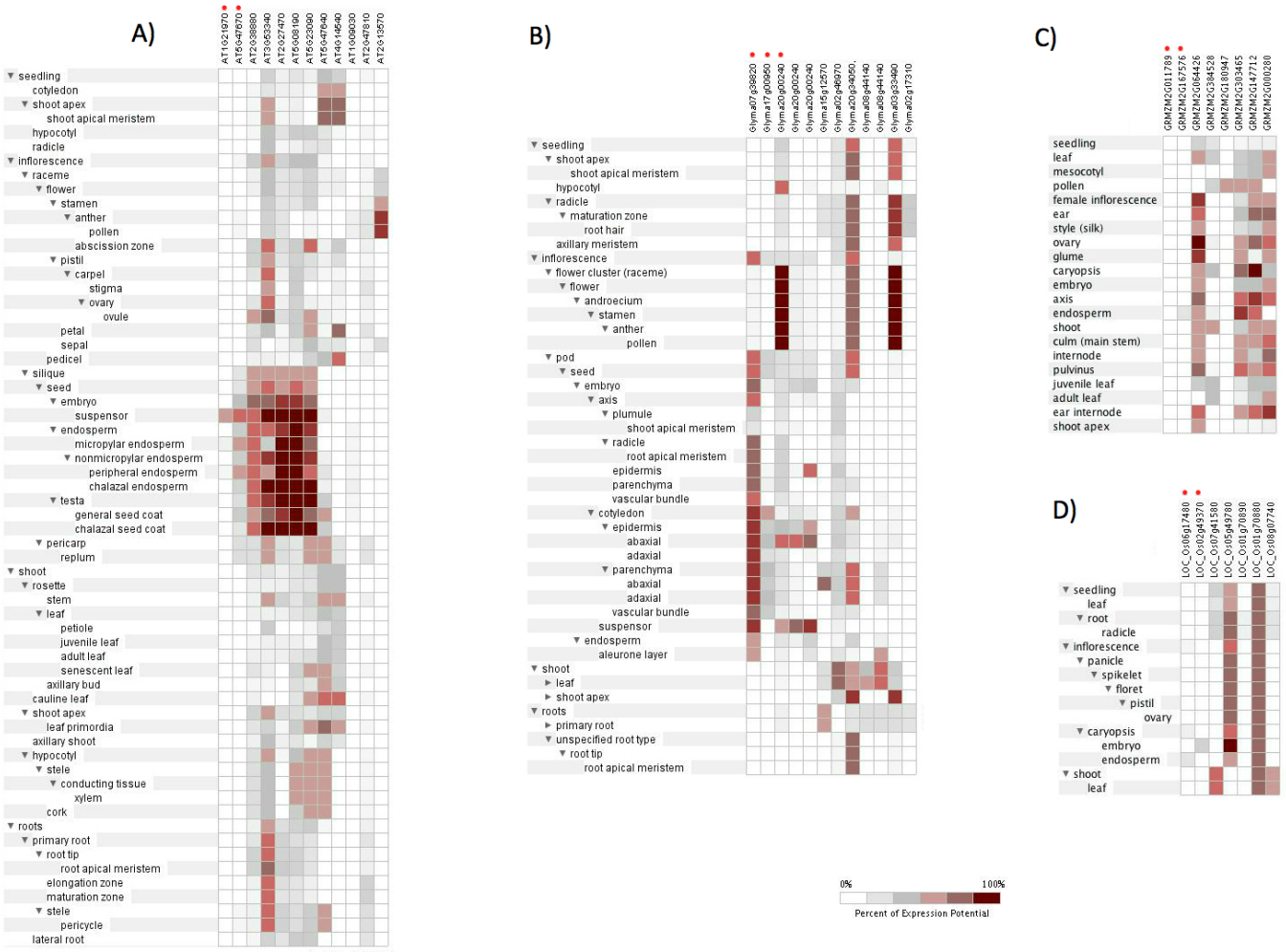


Figure 4. Expression profile of LEC1-type and non-LEC1-type genes in A) Arabidopsis, B) Soybean, C) Maize and D) Rice. The expression analysis was run out using Genevestigator software (https://www.genevestigator.com). Note that while the expression of non-LEC1-type genes is spread throughout several organs and tissues, the expression of LEC1-type genes (when detect under the Genevestigator conditions) is restricted to seeds or structures that will generate seeds (e.g. inflorescences).

Table 1. LEC1-type genes identified in plant genomes

Division	Species	NF-YB genes	LEC1 type genes	LEC1 type Accession number
Tracheophyta	*Manihot esculenta*	15	2	cassava4.1_032889m; cassava4.1_025848m
	Ricinus communis	12	2	29629.m001369; 57991.m000014
	Linum usitatissimum	23	6	Lus10003909; Lus10001914; Lus10008981; Lus10008980; Lus10008978; Lus10028845
	Populus trichocarpa	17	2	POPTR_0016s00740; POPTR_0006s00690
	Medicago truncatula	10	1	Medtr1g046630
	Phaseolus vulgaris	16	2	Phvulv091003579; Phvulv091021480
	Glycine max	25	4	Glyma03g18670; Glyma07g39820; Glyma17g00950; Glyma20g00240
	Cucumis sativus	11	3	Cucsa.042160; Cucsa.042150; Cucsa.243990
	Prunus persica	13	2	ppa026173m; ppa012143m
	Arabidopsis thaliana	13	2	AT1G21970 (LEC1); AT5G47670 (L1L);
	Arabidopsis lyrata	11	3	Alyrata 494206; Alyrata 921186; Alyrata 472432
	Capsella rubella	10	2	Carubv10011036m; Carubv10028250m
	Brassica rapa	20	4	Bra031356; Bra024924; Bra012301; Bra017471
	Thellungiella halophila	10	2	Thhalv10009593m; Thhalv10001185m
	Carica papaya	9	1	evm.TU.supercontig_12.138
	Citrus sinensis	16	2	orange1.1g026469m; orange1.1g038325m
	Citrus clementina	10	2	clementine0.9_029519m; clementine0.9_020395m
	Eucalyptus grandis	17	1	Eucgr.E03857
	Vitis vinifera	12	3	GSVIVG01014690001; GSVIVG01014689001; GSVIVG01002895001
	Aquilegia coerulea	5	2	Aquca_005_00328; Aquca_110_00016
	Sorghum bicolor	10	2	Sb10g010520; Sb04g029350
	Zea mays	18	3	GRMZM2G124663; GRMZM2G011789; GRMZM2G167576
	Setaria italica	9	2	Si020091m; Si008357m
	Oryza sativa	10	2	Os06g17480; Os02g49370
	Brachypodium distachyon	13	3	Bradi1g43480; Bradi1g43460; Bradi3g56400
	Pinus Sylvestris *	1	1	Gb JF280795.1
	Picea abies *	1	1	Gb JF280794.1
	Pinus contorta *	1	1	Gi 312861910
	Selaginella mollendorffii	5	1	Smoellindorffii 19595
Bryophyta	*Physcomitella patens*	6	0	-
Clorophyte	*Volvox carteri*	1	0	-

* Complete genome unavailable.

REFERENCES

[1] K. Yazawa, H. Kamada, Identification and characterization of carrot HAP factors that form a complex with the embryo-specific transcription factor C-LEC1, J Exp Bot, 58 (2007) 3819-3828.
[2] J.D. Bewley, Seed Germination and Dormancy, The Plant cell, 9 (1997) 1055-1066.
[3] J.J. Harada, Cellular and Molecular Biology of Plant Seed Development, in: Cellular and molecular Dordrecht: Kluwer Academic, 1997, pp. 545-592.
[4] J. Giraudat, B.M. Hauge, C. Valon, J. Smalle, F. Parcy, H.M. Goodman, Isolation of the Arabidopsis ABI3 gene by positional cloning, Plant Cell, 4 (1992) 1251-1261.
[5] G. Gusmaroli, C. Tonelli, R. Mantovani, Regulation of novel members of the Arabidopsis thaliana CCAAT-binding nuclear factor Y subunits, Gene, 283 (2002) 41-48.
[6] T. Lotan, M. Ohto, K.M. Yee, M.A. West, R. Lo, R.W. Kwong, K. Yamagishi, R.L. Fischer, R.B. Goldberg, J.J. Harada, Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells, Cell, 93 (1998) 1195-1205.
[7] D.W. Meinke, L.H. Franzmann, T.C. Nickle, E.C. Yeung, Leafy Cotyledon Mutants of Arabidopsis, Plant Cell, 6 (1994) 1049-1064.
[8] F. Parcy, C. Valon, A. Kohara, S. Misera, J. Giraudat, The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in concert to control multiple aspects of Arabidopsis seed development, Plant Cell, 9 (1997) 1265-1277.
[9] S.L. Stone, S.A. Braybrook, S.L. Paula, L.W. Kwong, J. Meuser, J. Pelletier, T.F. Hsieh, R.L. Fischer, R.B. Goldberg, J.J. Harada, Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis, Proc Natl Acad Sci U S A, 105 (2008) 3151-3156.
[10] U. Wobus, H. Weber, Seed maturation: genetic programmes and control signals, Curr Opin Plant Biol, 2 (1999) 33-38.
[11] K. Yazawa, K. Takahata, H. Kamada, Isolation of the gene encoding Carrot leafy cotyledon1 and expression analysis during somatic and zygotic embryogenesis, Plant Physiol Biochem, 42 (2004) 215-223.
[12] T.J. Stephenson, C.L. McIntyre, C. Collet, G.P. Xue, Genome-wide identification and expression analysis of the NF-Y family of transcription factors in Triticum aestivum, Plant Mol Biol, 65 (2007) 77-92.
[13] A. Dorn, J. Bollekens, A. Staub, C. Benoist, D. Mathis, A multiplicity of CCAAT box-binding proteins, Cell, 50 (1987) 863-872.

[14] J.L. Riechmann, O.J. Ratcliffe, A genomic perspective on plant transcription factors, Current opinion in plant biology, 3 (2000) 423-434.
[15] S.N. Maity, B. de Crombrugghe, Role of the CCAAT-binding protein CBF/NF-Y in transcription, Trends in biochemical sciences, 23 (1998) 174-178.
[16] Y. Xing, J.D. Fikes, L. Guarente, Mutations in yeast HAP2/HAP3 define a hybrid CCAAT box binding domain, EMBO J, 12 (1993) 4647-4655.
[17] S. Sinha, I.S. Kim, K.Y. Sohn, B. de Crombrugghe, S.N. Maity, Three classes of mutations in the A subunit of the CCAAT-binding factor CBF delineate functional domains involved in the three-step assembly of the CBF-DNA complex, Mol Cell Biol, 16 (1996) 328-337.
[18] R.W. Kwong, A.Q. Bui, H. Lee, L.W. Kwong, R.L. Fischer, R.B. Goldberg, J.J. Harada, LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development, Plant Cell, 15 (2003) 5-18.
[19] M.J. Holdsworth, L. Bentsink, W.J. Soppe, Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination, New Phytol, 179 (2008) 33-54.
[20] H. Lee, R.L. Fischer, R.B. Goldberg, J.J. Harada, Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor, Proc Natl Acad Sci U S A, 100 (2003) 2152-2156.
[21] Y. Watanabe, S. Yamamoto, Hōjinzei kihon tsūtatsu no gimonten, 4-teiban. ed., Gyōsei, Tōkyō, 2009.
[22] S. Cao, R.W. Kumimoto, C.L. Siriwardana, J.R. Risinger, B.F. Holt, 3rd, Identification and characterization of NF-Y transcription factor families in the monocot model plant Brachypodium distachyon, PLoS One, 6 (2011) e21805.
[23] Z. Xie, X. Li, B.J. Glover, S. Bai, G.Y. Rao, J. Luo, J. Yang, Duplication and functional diversification of HAP3 genes leading to the origin of the seed-developmental regulatory gene, LEAFY COTYLEDON1 (LEC1), in nonseed plant genomes, Mol Biol Evol, 25 (2008) 1581-1592.
[24] C.I. Keeling, S. Weisshaar, R.P. Lin, J. Bohlmann, Functional plasticity of paralogous diterpene synthases involved in conifer defense, Proc Natl Acad Sci U S A, 105 (2008) 1085-1090.
[25] A. Cagliari, A.C. Turchetto-Zolet, F.S. Maraschin, G. Loss, R. Margis, M. Margis-Pinheiro, The Evolutionary History of CBF Transcription Factors: Gene Duplication of CCAAT – Binding Factors NF-Y in Plants, in: Gene Duplication, Intech, 2011.
[26] A. De Grassi, C. Lanave, C. Saccone, Genome duplication and gene-family evolution: the case of three OXPHOS gene families, Gene, 421 (2008) 1-6.

[27] M. Lynch, J.S. Conery, The evolutionary fate and consequences of duplicate genes, Science, 290 (2000) 1151-1155.
[28] M. Lynch, A. Force, The probability of duplicate gene preservation by subfunctionalization, Genetics, 154 (2000) 459-473.
[29] L.E. Flagel, J.F. Wendel, Gene duplication and evolutionary novelty in plants, New Phytol, 183 (2009) 557-564.
[30] Y. Yu, Y. Li, G. Huang, Z. Meng, D. Zhang, J. Wei, K. Yan, C. Zheng, L. Zhang, PwHAP5, a CCAAT-binding transcription factor, interacts with PwFKBP12 and plays a role in pollen tube growth orientation in Picea wilsonii, J Exp Bot, 62 (2011) 4805-4817.
[31] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res, 32 (2004) 1792-1797.
[32] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, Mol Biol Evol, 28 (2011) 2731-2739.
[33] F. Ronquist, J.P. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under mixed models, Bioinformatics, 19 (2003) 1572-1574.
[34] P. Librado, J. Rozas, DnaSP v5: a software for comprehensive analysis of DNA polymorphism data, Bioinformatics, 25 (2009) 1451-1452.

4.4 Capítulo 4

Expression of LEC1 and L1L genes and their heterodimer assembly through castor bean (*Ricinus communis*) seed development.

Alexandro Cagliari[1], Ana Paula Korbers[1], Andreia Carina Turchetto-Zolet[1,2] Felipe dos Santos Maraschin[3], Rogerio Margis[1,2,4] and Marcia Margis-Pinheiro[1,2§].

[1]Departamento de Genética, Universidade Federal do Rio Grande do Sul, Brazil.

[2]Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Brazil.

[3]Departamento de Botânica, Universidade Federal do Rio Grande do Sul, Brazil.

[4]Departamento de Biofísica, Universidade Federal do Rio Grande do Sul, Brazil.

[§]Corresponding author:

Marcia Margis-Pinheiro. Departamento de Genética. Universidade Federal do Rio Grande do Sul. Avenida Bento Gonçalves 9500. Prédio 43.312. CEP 91501-970. Porto Alegre, RS, Brazil. Phone: +55 (51) 33089814. E-mail: marcia.margis@ufrgs.br. Homepage: http://www.ufrgs.br/rnai/ngfp/

ABSTRACT

Leafy Cotyledon1 (LEC1) and LEC1-Like (L1L) are members of a conserved oligomeric transcription factor family found in all eukaryotes named NF-Y (Nuclear Factor of the Y box), which acts as a transcriptional regulator for different sets of genes. Beyond the known functions in embryogenesis and seed maturation, there are evidences that LEC1 and L1L play either direct or indirect roles in the activation of seed storage genes, resulting in the accumulation of storage macromolecules. LEC1 and L1L are also known to function as key regulators of fatty acid biosynthesis. Here we identified and characterized LEC1 and L1L genes in castor bean (*Ricinus communis*), an oleaginous species that is emerging as an important oilseed model plant. We evaluated castor LEC1 (RcLEC1) and L1L (RcL1L) genes by analyzing their expression patterns and heterodimer assembly with castor NF-YC (RcNF-YC) members. Our results support the existence of different RcNF-YC partners between the RcLEC1 and RcL1L proteins, indicating that RcLEC1 and RcL1L are not redundant and could have evolved to play different roles during castor seed development. Moreover, the specificity of interactions and differences in expression patterns could result in different NF-Y transcription factors controlling different sets of genes. Altogether, our data provide new insights about the LEC1 and L1L functional divergence throughout castor seed development.

KEY WORDS: Leafy cotyledon1 (LEC1), LEC1-Like (L1L), Seed maturation, Seed Transcription Factor, lipid biosynthesis.

INTRODUCTION

Seed maturation is characterized by the massive accumulation of storage compounds, especially lipids, proteins, starches and sugars. The relative proportion and tissue localization of these compounds vary greatly depending on the species considered (Baud and Lepiniec, 2010). These storage molecules provide energy sources for the establishment of seedlings after germination. At the end of the maturation phase, the embryo acquires the ability to withstand desiccation and becomes metabolically quiescent (Bewley, 1997; Harada, 1997).

Gene expression programs are activated during the maturation phase and switched off during vegetative phases of plant development (Baud and Lepiniec, 2010). A network of master regulatory transcription factors, including LEAFY COTYLEDON1 (LEC1), LEC1-LIKE (L1L), LEAFY COTYLEDON2 (LEC2), FUSCA3 (FUS3), WRINKLED1 (WRI1), ABSCISIC ACID INSENSITIVE3 (ABI3), are responsible, in Arabidopsis, for the genetic control of seed maturation and dormancy induction. These genes encode proteins that play a central role in seed development as transcriptional regulators of embryogenesis (Giraudat *et al.*, 1992; Gusmaroli *et al.*, 2002; Lotan *et al.*, 1998; Meinke *et al.*, 1994; Parcy *et al.*, 1997; Sanjaya *et al.*, 2011; Stone *et al.*, 2008; Wobus and Weber, 1999; Yazawa *et al.*, 2004). Ectopic expression of these transcription factors cause the emergence of embryonic characteristics in vegetative and reproductive tissues (Braybrook and Harada, 2008; Meinke *et al.*, 1994; Satoh *et al.*, 1986; Shiota *et al.*, 1998; Stone *et al.*, 2008; Wagner *et al.*, 1987), while loss-of-function mutations result in the appearance of defects in seed development and maturation (Giraudat *et al.*, 1992; Gutierrez *et al.*, 2007; Holdsworth *et al.*, 2008; Kwong *et al.*, 2003; Lotan *et al.*, 1998; Luerssen *et al.*, 1998; Stone *et al.*, 2001; Yazawa and Kamada, 2007). The functions of these master transcription factors are partially redundant and the intricate network they define controls and/or cross-talks with hormonal and metabolic signaling pathways (Gutierrez *et al.*, 2007).

LEC1 and L1L are genes that coding for one of the subunits that compound the NF-Y (Nuclear Factor of the Y box) eukaryotic transcription factor (Dorn *et al.*, 1987; Stephenson *et al.*, 2007). NF-Y transcription factor is

composed of three different subunits named NF-YA (also known as HAP-2 or CBF-B), NF-YB (HAP3 or CBF-A), and NF-YC (HAP5 or CBF-C), each one required for DNA binding, subunit association and transcriptional regulation in mammals (Sinha *et al.*, 1995). There are evidences that plant NF-Y subunits can form a trimeric complex and function as transcription factors as in mammals (Yamamoto *et al.*, 2009): i) some plant NF-YA and NF-YB subunits are able to complement yeast strains mutated for their corresponding NF-Y subunits (Masiero *et al.*, 2002); ii) plant NF-Y subunits are also able to form a trimeric complex *in vitro* with mammalian cognate subunits (Ben-Naim *et al.*, 2006; Edwards *et al.*, 1998; Masiero *et al.*, 2002); and iii) some NF-YC proteins (e.g., NF-YC3, NF-YC4, and NF-YC9) physically interact with the floral-promoting NF-YB2 and NF-YB3 proteins, thereby participating in CONSTANS-mediated photoperiod-dependent flowering in Arabidopsis (Kumimoto *et al.*, 2010). Recently, it was demonstrated, by *in vitro* experiments, the conserved capacity of Arabidopsis NF-Y subunits to form a trimeric complex and to bind the specific CCAAT DNA sequence, as in mammals (Calvenzani *et al.*, 2012).

Both NF-YB and NF-YC subunits are structurally similar to the core histone subunits H2B and H2A, respectively, in that they possess the highly conserved histone-fold motif (HFM) (Arents and Moudrianakis, 1995; Baxevanis *et al.*, 1995; Mantovani, 1998). NF-YB subunit proteins typically consist of three domains: the *N*-terminal A domain, the central B domain, and the *C*-terminal C domain, being the B domain highly conserved across eukaryotes and required for both DNA binding and interaction with other NF-Y subunits (Sinha *et al.*, 1996; Xing *et al.*, 1993).

In mammals, for which the assembly of the NF-Y complex has been better studied, the NF-YB and NF-YC subunits form a tight heterodimer that moves to the nucleus, where the third subunit (NF-YA) is recruited to generate the complete NF-Y complex (Sinha *et al.*, 1996; Sinha *et al.*, 1995), resulting in either positive or negative transcriptional regulation (Ceribelli *et al.*, 2008; Peng and Jahroudi, 2002; Peng and Jahroudi, 2003; Siefers *et al.*, 2009).

The NF-YB subunit proteins can be divided into LEC1 and non-LEC1 type based on the presence of 16 amino acids in the B domain that are only observed in LEC1 type members (Holdsworth *et al.*, 2008; Kwong *et al.*, 2003)

and based on the expression pattern since the expression of LEC1-type genes is restricted to seeds and developing siliques while non-LEC1-type genes are ubiquitously expressed (Cao *et al.*, 2011; Kwong *et al.*, 2003; Lotan *et al.*, 1998). In Arabidopsis, the LEC1-type family is composed of LEC1 and the closely related subunit L1L.

We have previously presented a comparative genomic study using robust phylogenetic analyses including a vast number of taxa to validate an evolutionary model for the origin of LEC-type genes in plants and to help to elucidate the LEC1 and L1L relationship. We proposed a model in which the LEC1-type genes originated from an ancient Eukaryotic non-LEC1-type gene after duplication events occurred in plants and the emergence of a different role for LEC1-type genes, especially during seed development in vascular plants occurred through a process of neofunctionalization and/or subfunctionalization (Unpublished data).

Beyond the known functions of LEC1-type genes in embryogenesis and seed maturation (Kwong *et al.*, 2003; Lee *et al.*, 2003; Lotan *et al.*, 1998; Watanabe and Yamamoto, 2009; Yu *et al.*, 2011), there are evidences that LEC1 and L1L play either direct or indirect roles in the activation of seed storage genes, resulting in the accumulation of storage macromolecules (Braybrook and Harada, 2008). LEC1 and L1L are also known to function as key regulators of fatty acid biosynthesis: the inducible overexpression of LEC1 in Arabidopsis results in the induction of triacylglycerol (TAG) biosynthesis pathway (Mu *et al.*, 2008), and the expression of the oilseed rape (*Brassica napus*) or maize (*Zea mays*) orthologs of Arabidopsis LEC1 and L1L was enough to increase the fatty acid content in these species (Shen *et al.*, 2010; Tan *et al.*, 2011).

Castor bean (*Ricinus communis*) is an important oilseed whose seed oil composition (90% ricinoleic acid) is responsible for its unique chemical and physical properties, making castor oil a vital industrial raw material for numerous bioproducts, such as soaps, lubricants, paints, dyes, plastics, waxes, nylon, pharmaceuticals, cosmetics and biodiesel (Caupin, 1997).

Here we evaluated castor LEC1 (RcLEC1) and L1L (RcL1L) genes by

analyzing their expression patterns and heterodimer assembly with castor NF-YC (RcNF-YC) members. The possible implications of the interaction specificity between LEC1 proteins and their RcNF-YC partners on their activity as transcriptional factors throughout castor seed development were also discussed.

RESULTS

Expression of RcLEC1 and RcL1L genes through castor seed development

In order to gain insights about the acquisition of new functions during seed development by LEC1-type genes in vascular plant and to contribute to the understanding of the control of transcriptional regulation during seed development we identified and characterized LEC1-type genes in castor bean. We have identified two putative castor bean LEC1-type genes that have the amino acid characteristics of LEC1-type members (Figure 1and S1A, Table S1). These newly identified genes were designated as RcLEC1 and RcL1L genes based on their phylogenetic similarity with Arabidopsis orthologs (Figure 1).

The analysis of the expression of RcLEC1 and RcL1L through castor seed development (stages S1 to S5, see material and methods) was performed by quantitative real-time polymerase chain reaction (RT-qPCR). RcLEC1 presented a linear increase in expression until stage S4, followed by an accentuated decrease in stage S5 (mature seed) (Figure 2A). For the RcL1L gene, we observed a bell-shaped expression pattern (Figure 2B), characterized by a linear rise in the initial stages (S1 and S2), reaching maximal expression at stage S3, followed by a decrease during the last stages of seed development (S4 and S5), when the endosperm and embryo are totally expanded and the water loss process is initiated (Cagliari *et al.*, 2010).

All RcNF-YC members are expressed in castor seeds

In mammals, NF-YB and NF-YC first form a heterodimer structure during NF-Y assembly (Sinha *et al.*, 1996; Sinha *et al.*, 1995); therefore, we focused efforts on the identification of castor RcNF-YC members that could interact with RcLEC1 and RcL1L throughout NF-YB/NF-YC heterodimer assembly.

BLAST searches allowed the *in silico* identification of seven putative RcNF-YC members (Figure S1B, Table S1 and Figure 4). All of these genes contain the HFM conserved domain that is the signature of this gene family

(Kim *et al.*, 1996) (Figure S1B). DNA sequencing of RcNF-YC genes isolated from cDNA of seeds showed that one of the putative sequences did not contain the conserved HFM domain (data not shown). Therefore, this gene was not considered for further analysis.

The expression pattern of the six RcNF-YC genes was analyzed by RT-qPCR. Transcripts for all RcNF-YC genes were detected through seed development, although they had diverged in their expression profiles (Figure 2C to 2H). RcNF-YC1 displayed a biphasic pattern, with higher expression in stages S2 and S4 of seed development (Figure 2C). RcNF-YC2 and RcNF-YC3 showed very similar expression profiles, characterized by low levels of mRNA accumulation during the early stages (S1 to S3), followed by an accentuated increase in stage S4 and a subsequent decline in stage S5 (Figure 2D and 2E, respectively). RcNF-YC4 presented a linear increase in expression until stage S3, followed by a decrease in stage S4 and a posterior increase in stage S5 (Figure 2F). RcNF-YC5 mRNA accumulated at high levels in the earlier stages, when the embryo is not totally expanded (Cagliari *et al.*, 2010), and then decreased to undetectable levels over time (Figure 2G). RcNF-YC6 was expressed at the same level at stages S2 and S3, when the seed (endosperm) is totally expanded but the embryo is not fully formed (Cagliari *et al.*, 2010), followed by a severe reduction in later stages, becoming undetectable at these stages under our experimental conditions (Figure 2H).

The differential expression patterns shown by the RcNF-YC members are indicative of divergent temporal control during seed development. While the majority of RcNF-YC genes are expressed in all stages, the RcNF-YC5 and RcNF-YC6 genes are only expressed in the initial steps, which suggest that these genes play a role in the early stages of castor seed development.

In our conditions, a partially overlapping expression pattern for the RcNF-YB and RcNF-YC genes was observed, with a tendency toward a maximum expression levels at the middle stages of seed development (i.e., 2^{nd}, 3^{rd} and 4^{th} stages) (Figure 3). These stages are the most active periods in seed formation, during which the embryo and endosperm become totally expanded and where we previously observed the maximum expression of several enzymes involved in castor bean triacylglycerol (TAG) biosynthesis (Cagliari *et al.*, 2010). The

overlapping expression with TAG enzymes is in agreement with the role of some NF-Y subunits in seed oil accumulation (Shen *et al.*, 2010; Tan *et al.*, 2011).

While gene expression can tell us one part of the story, protein stability and kinetic activity can also play significant roles throughout castor seed development (Cagliari *et al.*, 2010). Therefore, additional studies will be particularly important to elucidate the potential roles of RcNF-YB and RcNF-YC in the control of lipid accumulation in castor seeds.

RcLEC1, RcL1L and RcNF-YC castor genes possess seed specific *cis*-regulatory elements in their promoter regions

Extensive analyses of the RcLEC1, RcL1L and RcNF-YC promoter regions (2 KB upstream of the transcription start site) allowed the identification of numerous *cis*-regulatory elements potentially that are responsive to different transcriptional regulatory networks (Table S2). The *cis*-regulatory elements, which were previously described as closely related to seed development and maturation, are depicted on Figure 4.

The best-characterized *cis*-elements in the promoters of genes that encode proteins involved in seed development and maturation are the prolamin box (P box TGT/CAAAG) and the GCN4 motif (TGAG/CTCA) (Takaiwa *et al.*, 1996; Vicente-Carbajosa *et al.*, 1997; Wu *et al.*, 1998; Xi and Zheng, 2011). The GCN4 *cis*-element was found in the promoter regions of RcLEC1, RcL1L, RcNF-YC1, RcNF-YC3, RcNF-YC4 and RcNF-YC5, while the P box *cis*-element was identified in RcNF-YC1, RcNF-YC3 and RcNF-YC6 and was absent in the RcLEC1 and RcL1L promoters (Figure 4). The involvement of these two motifs in regulating endosperm-specific expression was previously demonstrated (Albani *et al.*, 1997; Hammond-Kosack *et al.*, 1993; Marzabal *et al.*, 1998). We also identified RY motifs (CATGCA) (Baumlein *et al.*, 1992) in the promoters of RcNF-YC1 and RcNF-YC2 (Figure 4). Deletion of the RY motif abolished most of the seed-specific promoter activities, supporting a putative role for the RY box in the endosperm-specific gene expression (Xi and Zheng, 2011). Additionally, we found a Skn-1 motif, which has also been described as required

for endosperm expression (Blackwell *et al.*, 1994). Skn-1 was the most abundant *cis*-element identified in our analyses (two or more copies per promoter), being present in all promoters of the genes analyzed except RcL1L (Figure 4). Nevertheless, it must be noted that we were able to analyze only a partial RcL1L promoter sequence (344 bp) because the complete promoter region of RcL1L is not available in the castor bean genome database.

We have identified an ABA response *cis*-element (ABRE) (Michel *et al.*, 1993; Mundy *et al.*, 1990; Yamaguchi-Shinozaki *et al.*, 1990) in the RcNF-YC2, RcNF-YC4, RcNF-YC5 and RcNF-YC6 promoters (Figure 4). It is well known that the hormones abscisic acid (ABA) and gibberellic acid (GA) play important roles controlling seed maturation, germination and seedling development. Additionally, we have identified a GA response element (GARE) in the promoter regions of the RcNF-YC1, RcNF-YC2 and RcNF-YC6 genes (Figure 4). This *cis*-element is involved in the increased transcription level induced by GA, which is suppressed by high ABA content (Rogers and Rogers, 1992; Skriver *et al.*, 1991).

The fact that the RcNF-YB and RcNF-YC genes have *cis*-regulatory elements in their promoter sequences that are known to be responsive to hormones involved with seed development and to confer seed specific expression (Figure 4), is in agreement with the fact that the expression of all genes was detected in seeds by the RT-qPCR approach. Therefore, the confirmation of the functionality of these *cis*-regulatory elements associated with the identification of the *trans* elements that potentially bind these promoter *cis*-elements could open new insights in the regulation mechanisms of RcNF-YB (RcLEC1 and RcL1L) and RcNF-YC genes throughout castor seed development.

RcLEC1 and RcL1L interact with RcNF-YC members in the castor bean

The partial overlapping expression pattern observed for the RcLEC1 and RcL1L and RcNF-YC subunits throughout castor seed development (Figure 4) suggests that protein complexes could be formed. To investigate interactions

between specific protein pairs, Yeast Two Hybrid (Y2H) experiments were performed.

We initially examined the Y2H interactions between RcLEC1 and RcNF–YC subunits. We found that RcLEC1 interacts strongly with RcNF-YC2, RcNF-YC3, RcNF-YC4 and RcNF-YC5 and were unable to interact with either RcNF-YC1 or RcNF-YC6 (Figure 5 and Figure S2). The detected interactions were consistently observed regardless of which subunit was fused to the Gal4 DNA-binding domain (DBD) or activation domain (AD) and regardless of the presence of 20 mM of 3-Aminotriazole (3-AT). Increasing 3-AT concentrations, ranging from 5 to 40 mM, did not impair yeast growth (data not shown). In addition, auto-activation was not observed (Figure 5 and Figure S2). We further analyzed the interaction between RcL1L and RcNF–YC subunits and observed that RcL1L was able to interact strongly with all six RcNF-YC subunits (Figure 5 and Figure S2). However, when RcL1L was fused to Gal4 DBD, the interactions with RcNF-YC5 and RcNF-YC6 appeared weaker than when RcL1L was fused to Gal4 AD (Figure 5 and Figure S2). Yeast cell growth was not observed in the control combinations of RcLEC1, RcL1L and RcNF-YC with the empty vectors, suggesting specific interactions of RcLEC1 and RcL1L with RcNF-YC genes (Figure 5 and Figure S2).

To confirm the observed interactions between RcLEC1 and RcL1L with the RcNF-YC subunits *in planta*, Bimolecular Fluorescence Complementation Assays (BiFC) (Walter *et al.*, 2004) were performed in tobacco protoplasts using a transient expression system (Voinnet *et al.*, 2003). BiFC was performed in duplicate and the presence or absence of interaction was confirmed testing all possible combinations of proteins fused with the *C*- or *N*-terminal end of the Yellow Fluorescence Protein (YFP; See Material and Methods).

BiFC assays confirmed that RcL1L was able to interact with all RcNF-YC subunits (Figure 6A). In addition, the BiFC assay provides the subcellular localization of interactions in living plant cells. The associations between RcL1L and RcNF-YC1 or RcNF-YC6 were predominately observed in the nucleus, while the association between RcL1L and RcNF-YC2, RcNF-YC3, RcNF-YC4 or RcNF-YC5 took place in both the nucleus and the cytoplasm (Figure 6A).

The interactions between RcLEC1 and RcNF-YC2 or RcNF-YC3 were also confirmed by BiFC and occurred in both the nucleus and the cytoplasm, while RcNF-YC4 was found in the nucleus (Figure 6B). However, under our experimental conditions, the interaction between RcLEC1 and RcNF-YC5 was not observed by the BiFC approach. This result differed from the Y2H results, which had indicated an interaction between these proteins (Figure 5 and Figure S2). The negative controls (empty vectors or RcLEC1 and RcL1L combined with a GUS protein) produced no or very low background fluorescence, while the positive control (YFP_N-GUS_N/GUS_C-YFP_C) produced fluorescence in both the nucleus and the cytoplasm (Figure 6C).

DISCUSSION

We presented Y2H and BiFC data demonstrating that RcNF-YC members physically interact with the RcLEC1 and RcL1L proteins (Figure 5, 6 and Figure S2). The BiFC assay also demonstrated that the interactions of RcLEC1 and RcL1L with the RcNF-YC proteins could be restricted to the nucleus or observed in both the nucleus and the cytoplasm (Figure 6). However, sub-cellular localization of the NF-Y subunits may differ depending on the cell type and/or the physiological status of the examined cells (Frontini *et al.*, 2004; Yu *et al.*, 2011). In mammals, NF-YA and NF-YB are nuclear, whereas NF-YC localizes to both the cytoplasmic and nuclear compartments. Additionally, the localization of NF-YC in the nucleus was determined by its interaction with its NF-YB partner (Frontini *et al.*, 2004) which is in agreement with our results where all interactions among RcLEC1 or RcL1L with RcNF-YC subunits occurred in the nucleus, although in some cases the interaction it was not restricted to this compartment.

In Arabidopsis it has been demonstrated that whereas NF-YA2 and NF-YC9 are nuclear-located, NF-YB10 is distributed in the cytoplasm and only translocated to the nucleus when co-expressed with NF-YC2 (Hackenberg *et al.*, 2011), indicating the importance of the dimer association to the nuclear function of NF-YB proteins as transcription factors. Therefore, the identification of the dimeric assembly among RcNF-YC and RcLEC1 or RcL1L could represent the first step to the functional elucidation of the role played by these proteins throughout the castor seed development, especially during the phase of lipid production and accumulation.

Both Arabidopsis LEC1 and L1L are expressed during seed development, and L1L can complement the *lec1* mutation when expressed under the control of the LEC1 promoter; therefore, the molecular functions of LEC1 and L1L are considered functionally redundant (Kwong *et al.*, 2003; Watanabe and Yamamoto, 2009). Although the roles of LEC1 in embryo development have been investigated by genetic approaches, little is known about the nature of LEC1 and the redundant role of L1L as a transcription factor (Watanabe and Yamamoto, 2009).

The specific expression pattern of an individual NF-Y gene and its interactions with the other subunits are determinant for the specific function of NF-Y complexes in transcriptional control (Hackenberg *et al.*, 2011). In this context, one possible mechanism to distinguish the molecular function of the NF-Y subunit members and to understand the role played by these events of gene duplication depend on the identification of specific interacting partners (Watanabe and Yamamoto, 2009). In castor bean, RcLEC1 and RcL1L do not share all RcNF-YC partners. While RcL1L was found to interact with all RcNF-YC subunits by Y2H and BiFC, RcLEC1 was unable to interact with RcNF-YC1 and RcNF-YC6. Although the HFM domains of RcNF-YC1 and RcNF-YC6 are highly conserved as the other RcNF-YC subunits, the non-interaction with RcLEC1 might be due to the lack of important *cis*-elements in different protein regions or could be dependent of additional unknown nuclear factors not necessary for the direct interaction observed with the other RcNF-YC subunits.

Therefore, at the protein-interaction level and using two different methodologies, we did not identify total RcNF-YC partner redundancy in the castor bean, which is in agreement with the fact that suppression of L1L gene expression in Arabidopsis resulted in defects in embryo development that differed from the observed in *lec1* mutants, suggesting that LEC1 and L1L play different roles in embryogenesis (Kwong *et al.*, 2003). Our results suggest that RcL1L could be involved in the activation of different sets of target genes through the recruitment of different RcNF-YC transcription factors (RcNF-YC1 and RcNF-YC6), conferring a unique transcriptional regulatory activity to these RcNF-YB/NF-YC dimer complexes.

In addition, it was also demonstrated that plant NF-Y complexes can extend beyond the canonical NF-YA/NF-YB/NF-YC trimer observed in animal systems (Kumimoto *et al.*, 2010). In Arabidopsis, it was reported that the dimer composed of NF-YC2 and LEC1 or L1L was able to interact with bZIP67 to activate the expression of ABA-responsive genes (Watanabe and Yamamoto, 2009). Moreover, the same authors demonstrated that NF–YA9 interfered negatively with gene activation, Based on this fact, it has been hypothesized that NF-YA proteins may act as negative regulatory factors. Thus, the results presented here may indicate that the dimeric interactions observed between

RcLEC1/RcL1L with RcNF-YC subunits may control important aspects of seed development by the specific interaction between them with different classes of seed transcription factors in addition to forming the canonical NF-Y trimer with NF-YA (Thirumurugan *et al.*, 2008). Therefore, the identification of the RcNF-YC subunit protein complexes formed with RcLEC1 and RcL1L provide interesting leads to further understand the transcriptional regulatory networks that control castor seed development.

In conclusion, our results point to a scenario in which a functional divergence between the RcLEC1 and RcL1L genes permitted their proteins to control different sets of genes through their specific assembly with distinct RcNF-YC partners. Further studies, especially in a model plant such as Arabidopsis, will be important to determine the specific target genes of LEC1-type and NF-YC complexes throughout seed development.

MATERIAL AND METHODS

Plant material

Castor seeds (Al Guarany commercial variety - EMBRAPA-CPACT, Pelotas, Rio Grande do Sul, Brazil) were dissected, frozen immediately in liquid nitrogen and stored at -80°C. Expression analyses were performed at five stages of seed development, as previously described (Cagliari *et al.*, 2010): the first stage (S1) corresponds to seeds with less than 1 cm in length that were not fully expanded; stage 2 (S2) includes fully expanded and non-pigmented seeds; stage 3 (S3) seeds present only slight pigmentation and less than 20% of the final pigmentation; Stage 4 (S4) includes totally pigmented seeds from immature and green fruits; and stage 5 (S5) is represented by seeds collected from mature dehiscent fruits.

Quantitative polymerase chain reaction (RT-qPCR)

Total RNA extraction, cDNA synthesis, primer design and RT-qPCR reactions were performed as previously described (Cagliari *et al.*, 2010). RT-qPCR endogenous reference genes were EF1B/elongation factor 1-beta (Castor ID, 29785.m000934; forward primer 5'GCAGTTCGGAGCATTGAGAT-3' and reverse primer 5'GTCGTCCACAATGGTCATCA-3') and Ubi9/ubiquitin (Castor ID, 30169.m006323; forward primer 5'-ATCGATCGAATCAAGGAACG-3' and reverse primer 5'CACCCTCAATGTTGTAGTCACG-3') (Cagliari *et al.*, 2010). Two independent experiments were performed with eight biological samples for each stage of castor seed development. Each biological sample consisted of a pool of three castor seeds. Quadruplicate experimental sets of qPCR reaction samples, including the reference control genes and quadruplicate negative controls, were prepared and run in 96-well plates. SYBR fluorescence was analyzed by StepOne Software version 2.0 (Applied Biosystems), and the CT value for each sample was calculated and reported using the $2^{\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

Sequence search

Systematic data mining searches were performed using Arabidopsis LEC1 and L1L coding sequences as queries (blastp, blastx and tblastx) against Protein and Genome databases with the default parameters and an e-value threshold of 1.0 E-50 at NCBI (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov), the completed genome projects database at the JGI (Joint Genome Institute http://www.jgi.doe.gov) and the phytozome (http://www.phytozome.net/) to identify LEC1-type genes in fully sequenced genomes, including representatives of amoebozoa, yeast, fungi, algae, mosses, plants, and vertebrate and invertebrate species. Due to the lack of complete sequenced genomes for conifers, we performed an additional search on the incomplete genome of conifers that were represented in NCBI. A gene was classified as a LEC1-type when it presented the majority of LEC1-type specific residues in its conserved histone-fold motif. RcNF-YC genes were identified using the same blast criteria and considering the presence of their respective NF-YC histone-fold motif.

Sequence alignment and phylogenetic analysis

Nucleotide and protein sequences were aligned and inspected using MUSCLE (Edgar, 2004) implemented in Molecular Evolutionary Genetics Analysis (MEGA version 5.0) (Tamura *et al.*, 2011). Taxa terminologies are abbreviated using the first letter of the genus and two letters of the species name (e.g., Rco corresponds to *Ricinus communis*).

The phylogenetic analyses were reconstructed with either complete protein sequences or with the conserved central B domain of NF-YB genes. Bayesian analysis was conducted in MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). The mixed amino acid substitution model plus gamma and invariant sites was used in two independent runs of 20,000,000 generations each with two Metropolis-coupled Monte Carlo Markov chains (MCMC) that were run in parallel (starting each from a random tree). Markov chains were sampled every 100 generations, and the first 25% of the trees were discarded

as burn-in. The remaining trees were used to compute the majority rule consensus tree, the posterior probability of clades and branch lengths.

The pairwise synonymous (Ks) and non-synonymous (Ka) numbers of substitutions corrected for multiple hits were calculated using the DnaSP software (DNA polymorphism analysis) (Librado and Rozas, 2009).

Promoter sequence data sets

The 2-kb sequence upstream from the transcription start site was extracted as a putative promoter region from the castor bean database. The sequences were compared against known *cis*-regulatory elements in the collection of the PlantCARE database using the Search for Care program (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/). The *cis*-regulatory elements were listed, and the occurrence number on each promoter was recorded.

Yeast two-hybrid assays

All interaction studies were performed with the yeast strain AH 109 with two expression vectors: pDEST22 (tryptophan marker) and pDEST32 (leucine marker). Full-length cDNAs of RcLEC1, RcL1L and RcNF-YC were introduced into the appropriate plasmids using the Gateway technology. For auto-activation assays, transformants were plated on minimal synthetic defined (SD)-glucose medium supplemented with Ura/Met/His and lacking Leu and Trp (-LT). The ability to activate transcription in yeast was evaluated by monitoring growth after 7 days on selective SD medium lacking Leu, Trp and His (-LTH) supplemented with increasing 3-AT concentrations ranging from 0 to 40 mM.

Interaction assays were performed by the co-transformation of bait and prey plasmids into yeast strain and plated on SD-LT selective medium. As negative control, empty pDEST22 and pDEST32 were used. Transformants were allowed to grow for 4-5 days at 28 °C. Isolated transformants were diluted in 50 µL of distillated water, and 5 µL this dilution was spotted on solid SD-LTH

medium supplemented with increasing 3-AT concentrations. Yeast cells were allowed to grow for 7 days at 28 °C.

Bimolecular fluorescence complementation (BiFC) system

Taking advantage of the plasmids pE3136, pE3134, pE3130, and pE3132 (Arabidopsis Biological Resource Center - ABRC) we generated the split-YFP constructs pX-*N*YFP, pY-*C*YFP, pGUS-*N*YFP and pGUS-*C*YFP, where X corresponded to RcLEC1 or RcL1L and Y corresponded to the tested RcNF-YC proteins (we also generated the plasmids with opposite *N*- and *C*-termini of the YFP protein). pX-*N*YFP contained an *N*-terminal YFP (amino acids 1–155), and pY-*C*YFP contained a *C*-terminal YFP (amino acids 156–239) coding sequence driven by cauliflower mosaic virus (CaMV) 35S promoter. The amplified cDNAs were introduced into the appropriate plasmids using the Gateway technology.

As positive controls, we used p35S-GUS::*N*YFP and 35S-GUS::*C*YFP, which contain in-frame fusions of GUS with *N*-terminal YFP and *C*-terminal YFP, respectively.

Protoplast isolation and transformation

Protoplasts were isolated from *Nicotiana tobacco* mesophyll cells as previously described (Chen *et al.*, 2002), with minor modifications. Plasmid DNA was introduced by polyethylene glycol (PEG)-mediated transfection (Schirawski *et al.*, 2000). Tobacco mesophyll cells were transformed in pairwise combinations with 10 µg of each plasmid construction. Protoplasts were incubated in the dark for 16 hours at 27 °C before microscopy analysis under an Olympus FluoView 1000 confocal laser-scanning microscope.

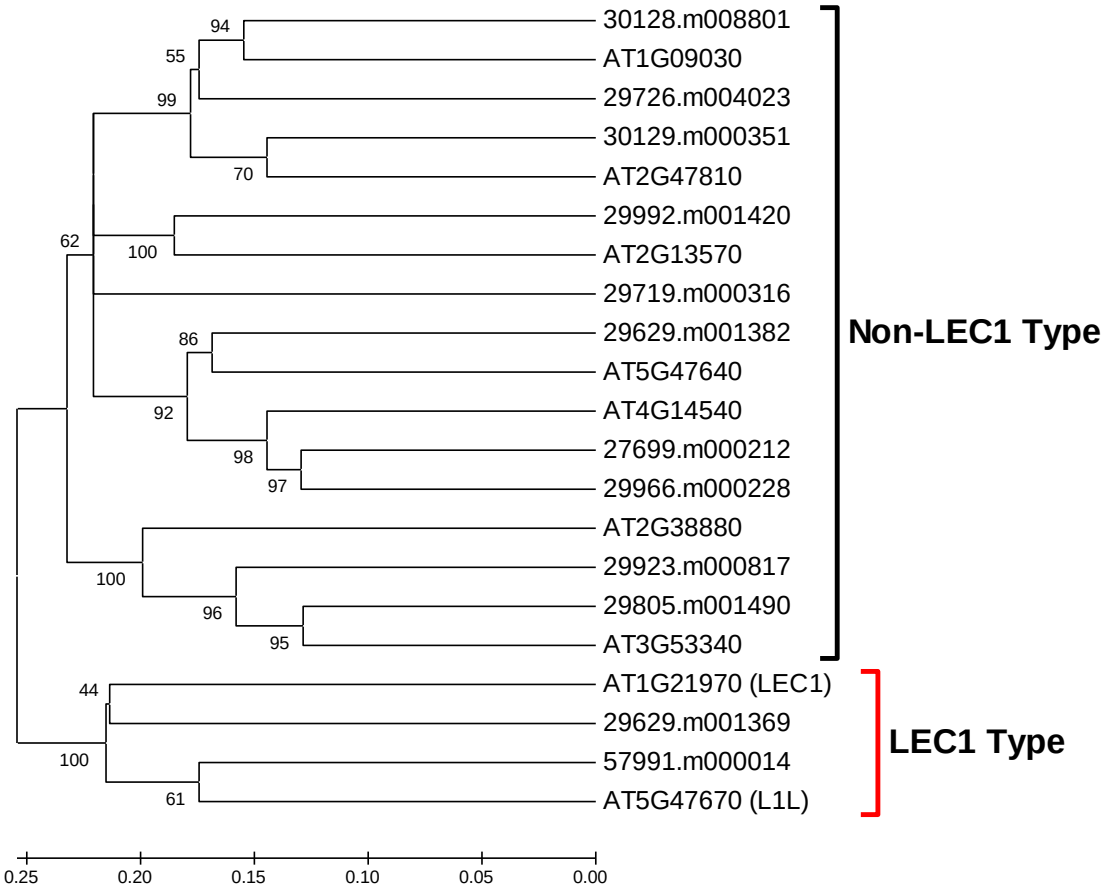


Figure 1. Arabidopsis and castor bean LEC1 and non-LEC1 type gene cluster analysis. The cladogram was constructed by the neighbor-joining method, using the MEGA 5.2 program, after 1000 bootstrap replications, with pairwise deletion and nucleotide p-distance as the main parameters. Bootstrap values higher than 50% are given at the respective nodes. The scale bar indicates genetic distance. Arabidopsis (AT5g47670, e.g.) and castor bean (29966.m000228, e.g.) genes are designed by their respective accession numbers.

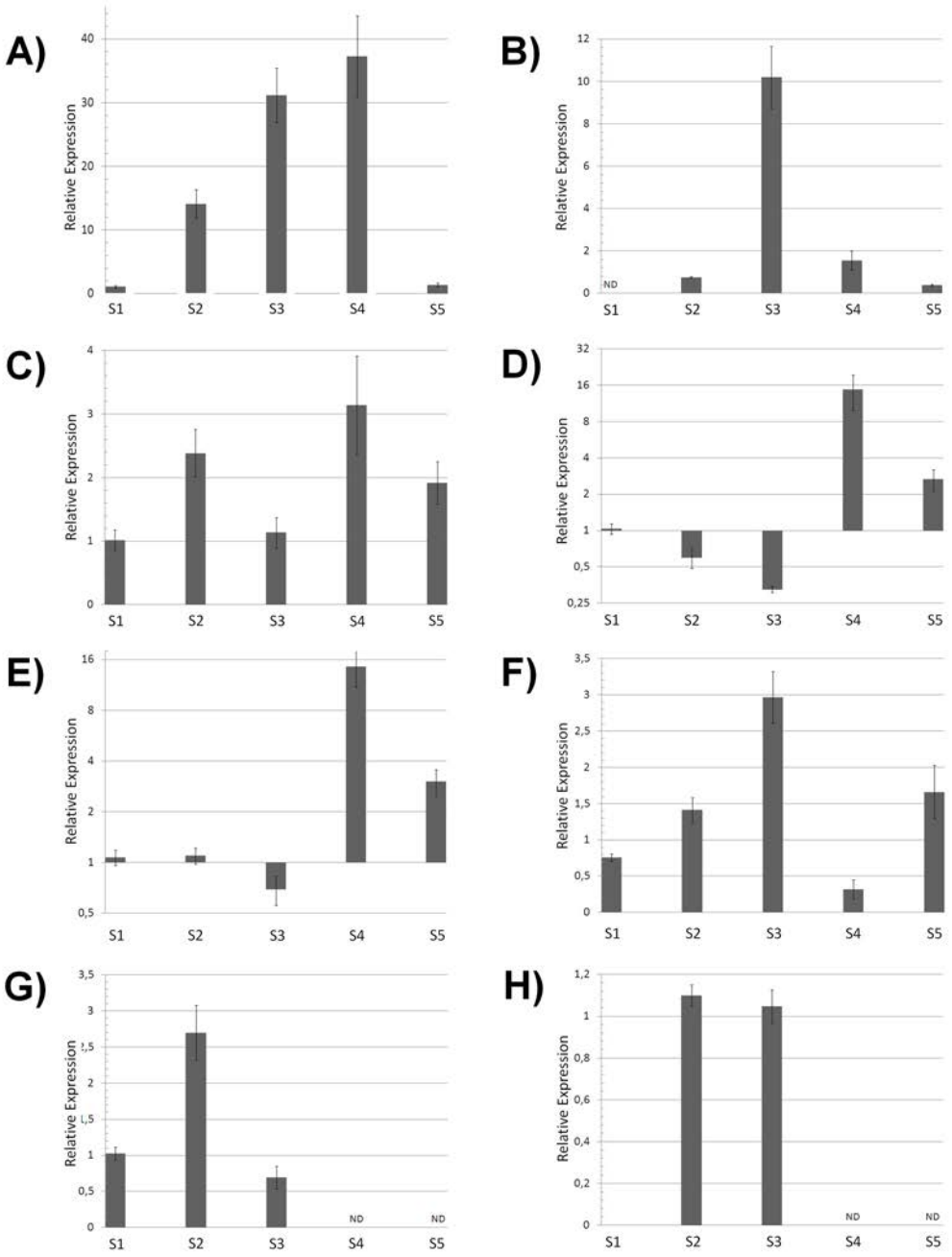


Figure 2. Relative expression pattern of NF-YB (LEC1 and L1L) and NF-YC castor genes during castor seed development (stages S1 to S5). Expression levels were normalized with respect to two housekeeping control genes. Each data point represents the mean ± standard deviation of two independent experiments, with eight biological and four experimental replicates. **(A)** LEC1; **(B)** L1L; **(C)** NF-YC1; **(D)** NF-YC2; **(E)** NF-YC3; **(F)** NF-YC4; **(G)** NF-YC5; **(H)** NF-YC6. ND: non-detected expression.

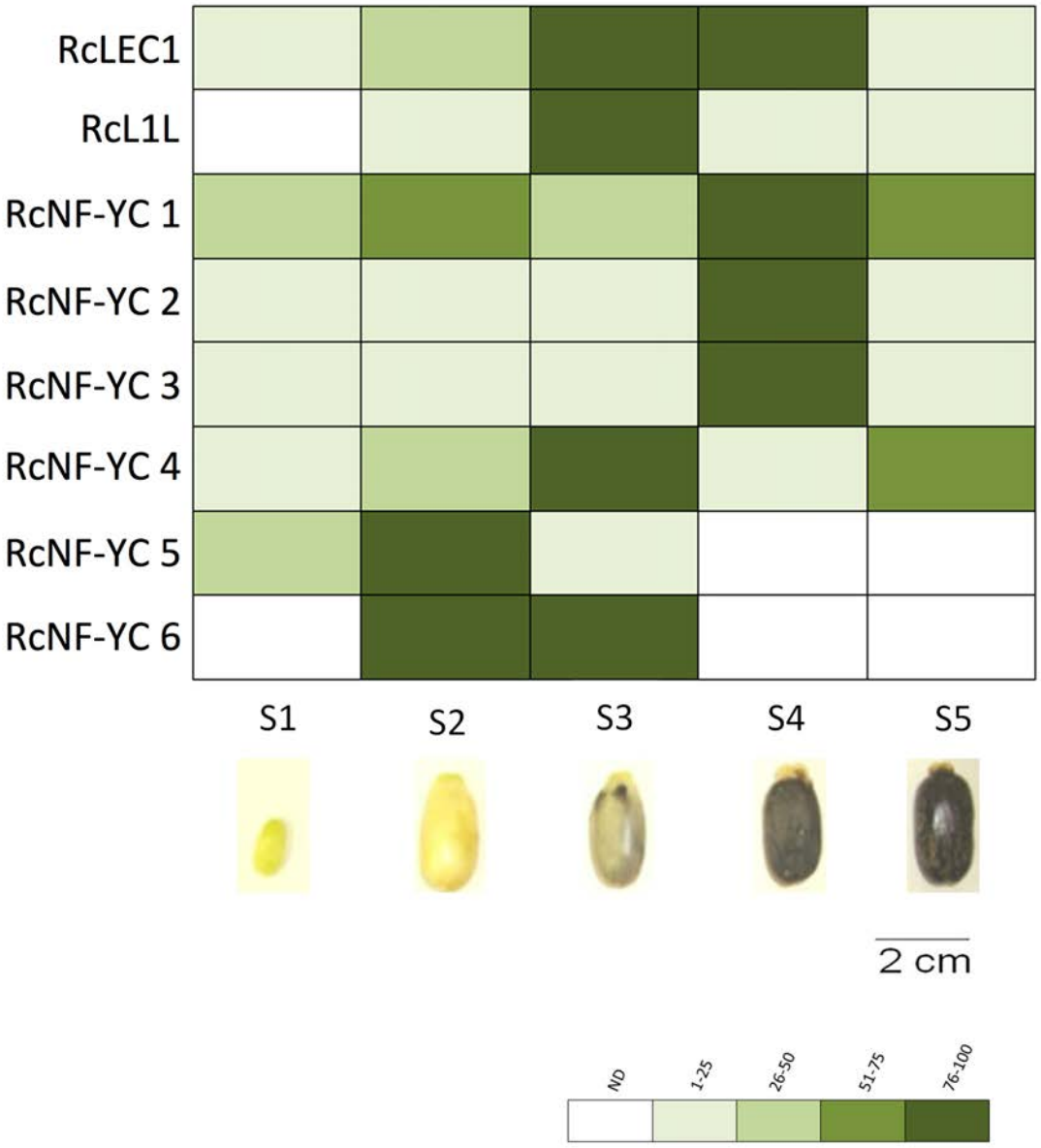


Figure 3. Partial overlapping expression of NF-Y B (LEC1 and L1L) and NF-YC genes throughout castor seed development. Colored boxes indicate the relative mRNA expression level of genes (%) at each stage of seed development (S1 to S5) compared to the maximal expression level observed. ND indicates seed stages where gene expression was not detected under our experimental conditions.

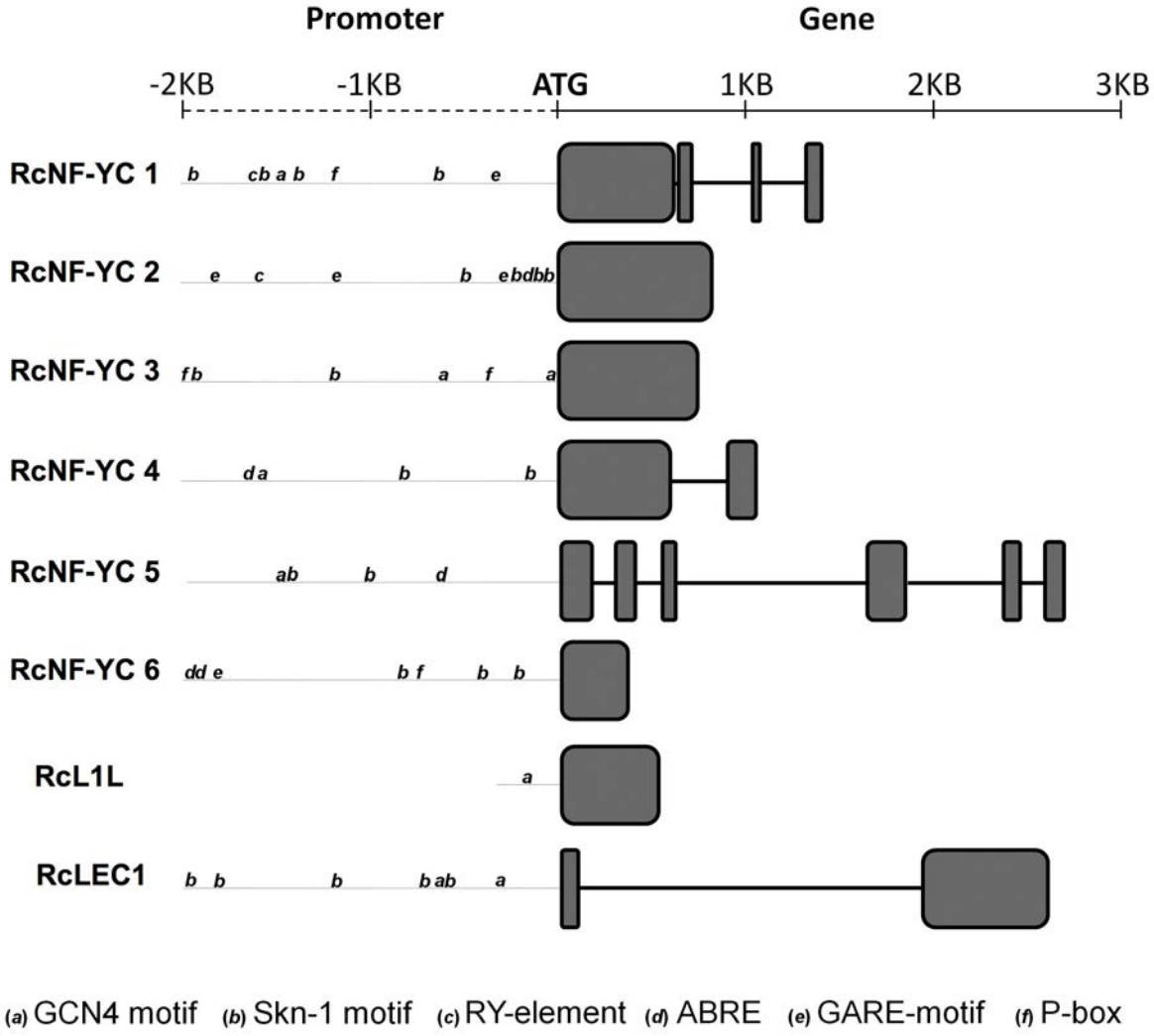


Figure 4. Structural organization and promoter analysis of the castor NF-YB (LEC1 and L1L) and NF-YC genes. Exon sequences are represented as gray boxes, and bars represent introns (on scale). Dotted bars indicate the promoter sequence of castor genes, and the letters (a-f) represent *cis*-elements involved with seed development and maturation and hormone responsiveness: a) GCN4 motif; b) Skn-1 motif; c) RY-element; d) ABRE; e) GARE-motif and f) P-box.

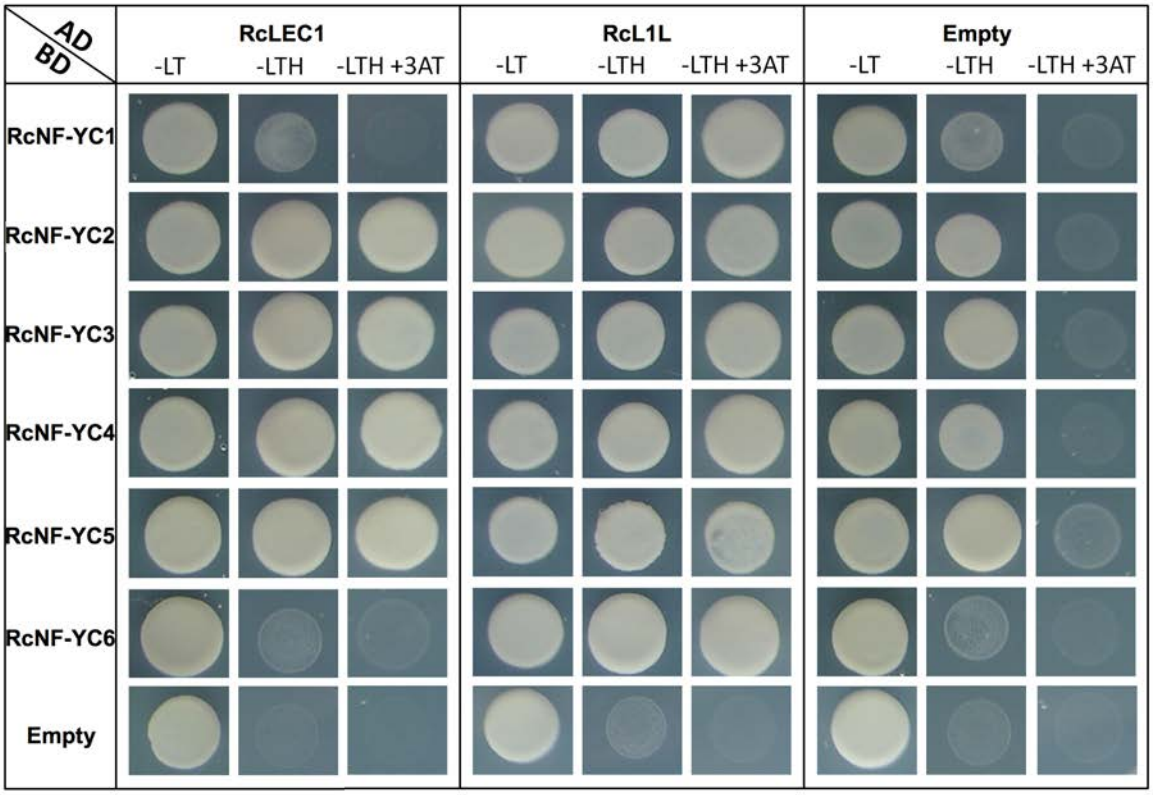


Figure 5. LEC1 and L1L interact with NF-YC proteins in yeast. Yeast cells expressing LEC1 or L1L fused to GAL4AD and NF-YC fused to GAL4DBD were spotted in minimal selective medium lacking Leu, Trp (-LT). The ability to activate transcription in yeast was evaluated by monitoring growth after 7 days on selective SD medium lacking Leu, Trp and His (-LTH) supplemented with 20 mM 3-AT (-LTH +3AT). Yeast transformed with empty plasmids pDEST22 and pDEST32, expressing the activation domain (AD) and DNA binding domain (DBD) of GAL4, respectively, were used as negative controls.

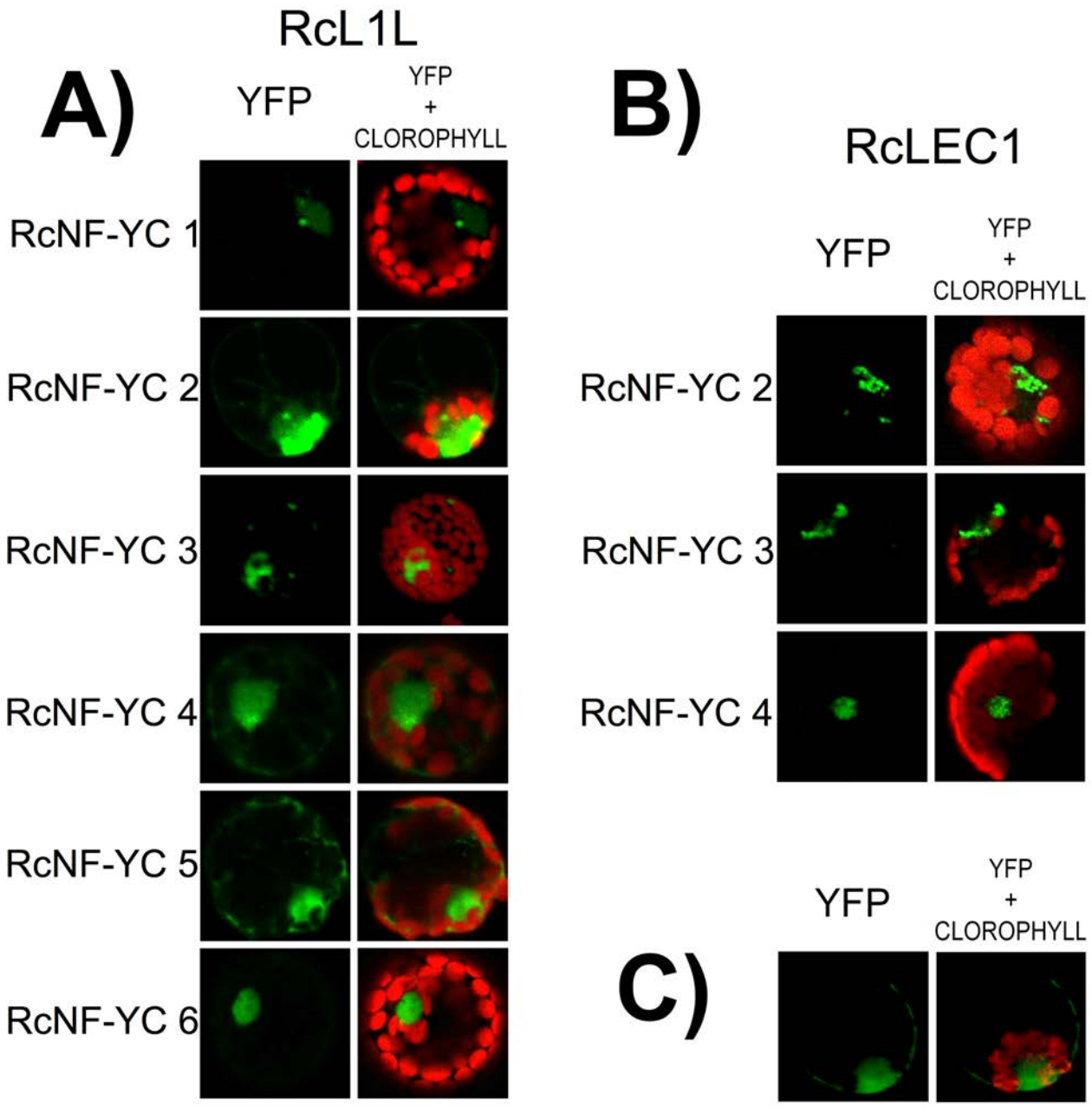
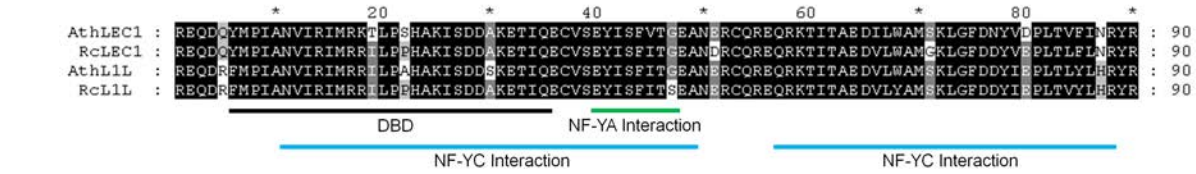


Figure 6. LEC1 and L1L interact with NF-YC proteins in tobacco mesophyll cells. Bimolecular fluorescence complementation (BiFC) assay in tobacco protoplasts transfected with *N*YFP-LEC1 and L1L and *C*YFP-NF-YC1 to 6 constructs. Green signals indicate reconstituted YFP fluorescence, and red signals indicate chlorophyll autofluorescence. **A)** Restoration of YFP fluorescence through interaction between L1L and NF-YC. **B)** Restoration of YFP fluorescence through interaction between LEC1 and NF-YC. **C)** Positive control for interaction using 35S-GUS::*N*YFP and 35S-GUS::*C*YFP cassettes.

SUPPLEMENTARY DATA

A)



B)

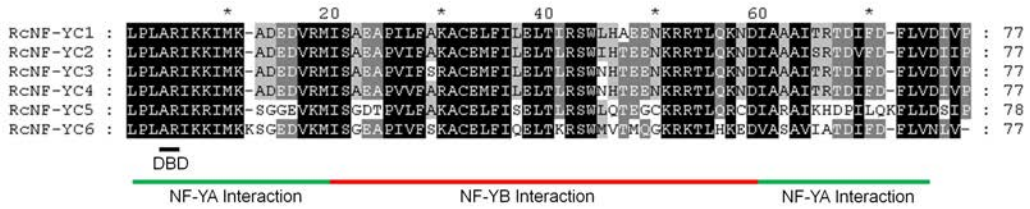


Figure S1. NF-YB and NF-YC conserved histone-fold motif. A) Arabidopsis (Ath) and Castor Bean (Rco) LEC1-type genes alignment. B) Castor NF-YC genes alignment. Black lines under the alignments indicate regions involved in contacting DNA (DNA Binding Domain-DBD), red lines under the alignments indicate NF-YB interaction regions, green lines under the alignments indicate regions involved in interaction with the NF-YA subunit and blue lines under the alignments indicate NF-YC interaction regions.

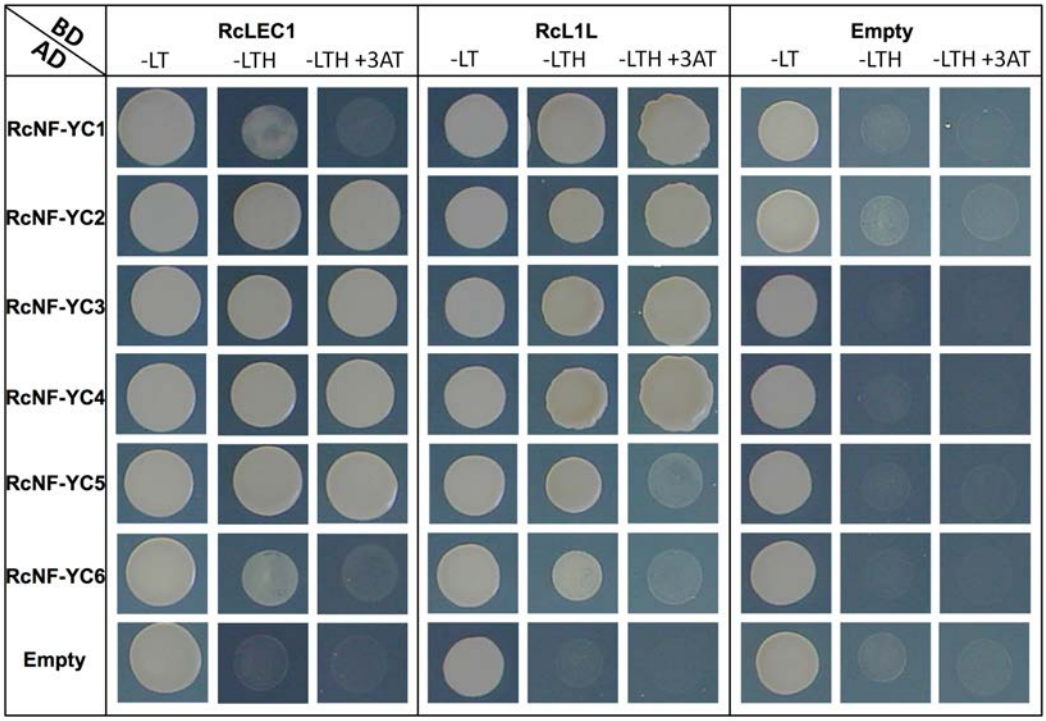


Figure S2. LEC1 and L1L interact with NF-YC proteins in yeast. Yeast cells expressing LEC1 or L1L fused to GAL4DBD and NF-YC fused to GAL4AD were spotted in minimal selective medium lacking Leu, Trp (-LT). The ability to activate transcription in yeast was evaluated by monitoring growth after 7 days on selective SD medium lacking Leu, Trp and His (-LTH) supplemented with 20 mM 3-AT (-LTH +3AT). Yeast transformed with empty plasmids pDEST22 and pDEST32, expressing the DNA binding domain (DBD) and the activation domain (AD) of GAL4, respectively, were used as negative controls.

Table S1. Summary of RcLEC1, RcL1L and RcNF-YC genes, RT qPCR primer sequences and size of amplification products.

Acronym	Castor Bean ID	Forward primer (5' 3') Reverse primer (5' 3')	Amplicon Size (pb)
LEC1	29629.m001369	AACATCAATCTCCCCAGCAC ATGGTCTCCTTGGCATCATC	197
L1L	57991.m000014	ATGCAATGAGCAAGCTAGGG CGTACCGAATTCCACACTCC	137
NF-YC 1	30147.m014493	CACCAATGGGTCAGCCTACT TTGCATAAGACCCATCGTCA	141
NF-YC 2	30169.m006305	AGAGGAGGGATTAGGGGTCA GCTTGATCAACTGGCTTTCC	135
NF-YC 3	29844.m003332	AGCCTCAGTTTCCACCACAG GCAACATAAGGGCGAGACTG	100
NF-YC 4	30128.m008750	TGAAGTACTGTCGTCAATATCAAGG AGTCTTGCTAGATTTCAGCTCCA	112
NF-YC 5	29586.m000615	TGAGCCTCTTCCTCCAGTTG AATTGAATTCAGCAGAAGAAGACA	134
NF-YC 6	30190.m011186	AAAGGTCTTGGATGGTGACG CCAAGGAGTTGCTGCAATTAT	132

Table S2. Putative regulatory *cis*-element sequences in 5'-upstream regions of NF-YB (LEC1 and L1L) and NF-YC genes.

Name of *cis* element	Organism where it was identified	Sequence of *cis* element	Function	NF-YC						NFY-B	
				1	2	3	4	5	6	LEC1	L1L
3-AF1 binding site	*Solanum tuberosum*	TAAGAGAGGAA	light responsive element					**1**			
4cl-CMA2b	*Petroselinum crispum*	TCTCACCAACC	light responsive element		1						
5UTR Py-rich stretch	*Lycopersicon esculentum*	TTTCTTCTCT	*cis*-acting element conferring high transcription levels	1	5	3	2	2		1	
AAAC-motif	*Spinacia oleracea*	CAACAAAAACCT	light responsive element				1		1		
ABRE	*Arabidopsis thaliana/Triticum aestivum*	TACGTG/GACACGTGGC	*cis*-acting element involved in the abscisic acid responsiveness		1		1	1	2		
ACE	*Petroselinum crispum*	GACACGTATG/ ACGTGGA	*cis*-acting element involved in light responsiveness		3		3		1		
AC-I	*Phaseolus vulgaris/ Populus tremuloides*	(T/C)C(T/C)(C/T)ACC(T/C)ACC	element conferring enhanced xylem expression and repressed phloem expression in transgenic tobacco; enhancing petal expression	1	1						
AC-II	*Phaseolus vulgaris*	CCACCAACCCCC	element conferring enhanced xylem expression and repressed phloem expression in transgenic tobacco; enhancing petal expression	1							
AE-box	*Arabidopsis thaliana*	AGAAACAA	part of a module for light response	1		1		1			
ARE	*Zea mays*	TGGTTT	*cis*-acting regulatory element essential for the anaerobic induction	4	3	2			1	4	
As-2-box	*Nicotiana tabacum*	GATAATGATG	involved in shoot-specific expression and light responsiveness	1					2		
ATCC-motif	*Pisum sativum*	CAATCCTC	part of a conserved DNA module involved in light responsiveness					1			
ATCT-motif	*Pisum sativum/ Arabidopsis thaliana*	AATCTAATCC	part of a conserved DNA module involved in light responsiveness		2	1	3	2			
ATGCAAAT motif	*Oryza sativa*	ATACAAAT	*cis*-acting regulatory element associated to the TGAGTCA motif	1			1		1		
AT1-motif	*Solanum tuberosum*	AATTATTTTTTATT	part of a light responsive module			1			1		

AT-rich element	*Glycine max*	ATAGAAATCAA	binding site of AT-rich DNA binding protein (ATBP-1)			2		1			
AuxRR-core	*Nicotiana tabacum*	GGTCCAT	*cis*-acting regulatory element involved in auxin responsiveness						1		
Box 4	*Petroselinum crispum*	ATTAAT	part of a conserved DNA module involved in light responsiveness		4	3		4	2	2	1
Box E	*Petroselinum crispum*	ACCCATCAAG	*cis*-element for induction upon fungal elicitation		1		1				
Box I	*Pisum sativum*	TTTCAAA	light responsive element		1	1		1	1	2	
box II	*Petroselinum hortense/Pisum sativum*	TCCACGTGGC/GTGAGGTAATAT	part of a light responsive element				1	1	2		
Box III	*Pisum sativum*	CATTTACACT	part of a conserved DNA module involved in light responsiveness		1	1		1			
Box S	*Arabidopsis thaliana*	AGCCACC	elicitation; wounding and pathogen responsievness		1						
Box-W1	*Petroselinum crispum*	TTGACC	fungal elicitor responsive element	1	1	1		1	1		1
CAT-box	*Arabidopsis thaliana*	GCCACT	*cis*-acting regulatory element related to meristem expression		1	1	1	2	1		
CATT-motif	*Zea mays*	GCATTC	part of a light responsive element	3		2			1		
CCAAT-box	*Hordeum vulgare*	CAACGG	MYBHv1 binding site				1			2	
CGTCA-motif	*Hordeum vulgare*	CGTCA	*cis*-acting regulatory element involved in the MeJA-responsiveness	1	1		2	2	1	2	1
chs-CMA1a	*Daucus carota*	TTACTTAA	part of a light responsive element		1			1			
chs-CMA2a	*Hordeum vulgare*	GCAATTCC	part of a light responsive element	1		1					
circadian	*Lycopersicon esculentum*	CAANNNNATC	*cis*-acting regulatory element involved in circadian control	4	2	3	2		1	2	
E2Fb	*Nicotiana tabacum*	TTTGCCGC	E2F-binding site; G1-M transition of cell cycle		2						
ERE	*Dianthus caryophyllus*	ATTTCAAA	ethylene-responsive element		1	1					
GAG-motif	*Arabidopsis thaliana/ Spinacia oleracea/Hordeum vulgare*	AGAGAGT/ AGAGATG/ GGAGATG	part of a light responsive element	2	1		1	1			

Gap-box	*Arabidopsis thaliana*	AAATGGAGA	part of a light responsive element					1			
GA-motif	*Arabidopsis thaliana/Helianthus annuus*	ATAGATAA/AAAGATGA	part of a light responsive element		1	3	1		1		
GARE-motif	*Brassica oleracea*	TCTGTTG	gibberellin-responsive element	1	3				1		
GATA-motif	*Solanum tuberosum/ Arabidopsis thaliana/Pisum sativum*	AAGGATAAGG/GATAGGA/ GATAGGG	part of a light responsive element	5	3		1	1		1	
GATT-motif	*Triticum aestivum*	CTGCAGATTTCT	part of a light responsive element		1				1		
G-BOX	*Antirrhinum majus/ Brassica oleracea/ Zea mays/ Daucus carota/Solanum tuberosum/ Pisum sativum/ Arabidopsis thaliana*	CACGTA/ TAACACGTAG/ GACATGTGGT/CACATGG/ CACGTG/GCCACGTGGA	*cis*-acting regulatory element involved in light responsiveness		4	1	4	2	5		
GCN4_motif	*Oryza sativa*	TGTGTCA	*cis*-regulatory element involved in endosperm expression	1		2	2	1		2	1
GT1-motif	*Avena sativa/ Arabidopsis thaliana/ Solanum tuberosum*	GGTTAAT/ GGTTAA/ATGGTGGTTGG	light responsive element	3	1		1	3			
H-box	*Phaseolus vulgaris*	CCTACCNNNNNNNCTNNNNA	Control of developmentally regulated and stress induced expression binds protein factors KAP-1 and KAP-2		1						
HD-Zip 3	*Arabidopsis thaliana*	GTAAT(G/C)ATTAC	protein binding site				1				
HSE	*Brassica oleracea*	AAAAAATTTC	*cis*-acting element involved in heat stress responsiveness	2	1		3				
I-box	*Arabidopsis thaliana/ Zea mays/ Solanum tuberosum/Nicotiana plumbaginifolia /Flaveria trinervia*	CCTTATCCT/CTCTTATGCT/ GATATGG	part of a light responsive element	4	3		2			2	
LAMP-element	*Pisum sativum*	CTTTATCA	part of a light responsive element		1						
L-BOX	*Petroselinum crispum*	TCTCACCAACC	part of a light responsive element		1					1	
MBS	*Arabidopsis thaliana/ Zea mays*	TAACTG/CGGTCA	MYB binding site involved in drought-inducibility	5	5	2		1	3	3	

MBSII	*Petunia hybrida*	AAAAGTTAGTTA	MYB binding site involved in flavonoid biosynthetic genes regulation				1					
MNF1	*Zea mays*	GTGCCC(A/T)(A/T)	light responsive element		1				1	1		
MRE	*Petroselinum crispum*	AACCTAA	MYB binding site involved in light responsiveness				1	1				1
O2-site	*Zea mays*	GATGACATGG	cis-acting regulatory element involved in zein metabolism regulation	2	1		1	1		1		
P-box	*Oryza sativa*	CCTTTTG	gibberellin-responsive element	1		2			1			
Pc-CMA2a	*Spinacia oleracea*	CAACCAATGAAAA	part of a light responsive element	1								
Pc-CMA2c	*Spinacia oleracea*	GCCCACACA	part of a light responsive element	1								
RY-element	*Helianthus annuus*	CATGCATG	*cis*-acting regulatory element involved in seed-specific regulation		1							
Skn-1_motif	*Oryza sativa*	GTCAT	*cis*-acting regulatory element required for endosperm expression	4	4	2	2	2	3	5		
Sp1	*Zea mays*	CC(G/A)CCC	light responsive element	5					1			1
TCA-element	*Nicotiana tabacum/ Brassica oleracea*	CCATCTTTTT/CAGAAAAGGA	*cis*-acting element involved in salicylic acid responsiveness	1	1	6	3					
TCCC-motif	*Spinacia oleracea*	TCTCCCT	part of a light responsive element			3						
TC-rich repeats	*Nicotiana tabacum*	GTTTTCTTAC	*cis*-acting element involved in defense and stress responsiveness	1			1	2	2	2		
TCT-motif	*Arabidopsis thaliana*	TCTTAC	part of a light responsive element	1	1	2		2	1	2		
TGACG-motif	*Hordeum vulgare*	TGACG	*cis*-acting regulatory element involved in the MeJA-responsiveness	1	1		2	2	1	2		1
TGA-element	*Brassica oleracea*	AACGAC	auxin-responsive element	1	1	1						
W box	*Arabidopsis thaliana*	TTGACC	elicitation; wounding and pathogen responsiveness. Binds WRKY type transcription factors	1	1	1		1	1			1
WUN-motif	*Brassica oleracea*	AAATTTCCT	wound-responsive element		1							

REFERENCES

Albani D, Hammond-Kosack MC, Smith C, Conlan S, Colot V, Holdsworth M, Bevan MW. 1997. The wheat transcriptional activator SPA: a seed-specific bZIP protein that recognizes the GCN4-like motif in the bifactorial endosperm box of prolamin genes. *Plant Cell* **9**, 171-84.
Arents G, Moudrianakis EN. 1995. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc Natl Acad Sci U S A* **92**, 11170-4.
Baud S, Lepiniec L. 2010. Physiological and developmental regulation of seed oil production. *Prog Lipid Res* **49**, 235-49.
Baumlein H, Nagy I, Villarroel R, Inze D, Wobus U. 1992. Cis-analysis of a seed protein gene promoter: the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. *Plant J* **2**, 233-9.
Baxevanis AD, Arents G, Moudrianakis EN, Landsman D. 1995. A variety of DNA-binding and multimeric proteins contain the histone fold motif. *Nucleic Acids Res* **23**, 2685-91.
Ben-Naim O, Eshed R, Parnis A, Teper-Bamnolker P, Shalit A, Coupland G, Samach A, Lifschitz E. 2006. The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA. *Plant J* **46**, 462-76.
Bewley JD. 1997. Seed Germination and Dormancy. *Plant Cell* **9**, 1055-66.
Blackwell TK, Bowerman B, Priess JR, Weintraub H. 1994. Formation of a monomeric DNA binding domain by Skn-1 bZIP and homeodomain elements. *Science* **266**, 621-8.
Braybrook SA, Harada JJ. 2008. LECs go crazy in embryo development. *Trends Plant Sci* **13**, 624-30.
Cagliari A, Margis-Pinheiro M, Loss G, Mastroberti AA, Mariath JED, Margis R. 2010. Identification and expression analysis of castor bean (Ricinus communis) genes encoding enzymes from the triacylglycerol biosynthesis pathway. *Plant Science* **179**, 499-509.
Calvenzani V, Testoni B, Gusmaroli G, Lorenzo M, Gnesutta N, Petroni K, Mantovani R, Tonelli C. 2012. Interactions and CCAAT-binding of Arabidopsis thaliana NF-Y subunits. *Plos One* **7**, e42902.
Cao S, Kumimoto RW, Siriwardana CL, Risinger JR, Holt BF, 3rd. 2011. Identification and characterization of NF-Y transcription factor families in the monocot model plant Brachypodium distachyon. *PLoS One* **6**, e21805.
Caupin H. 1997. Products from castor oil: past, present, and future. *Lipid Technol.Appl.*, 787-95.
Ceribelli M, Dolfini D, Merico D, Gatta R, Vigano AM, Pavesi G, Mantovani R. 2008. The histone-like NF-Y is a bifunctional transcription factor. *Mol Cell Biol* **28**, 2047-58.
Chen SL, Dai SX, Li JK, Wang SS, P.; A, Aloys H. 2002. Isolation of protoplast and ion channel recording in plasma membrane of suspension cells of Populus euphratica. *For. Stud. China* **4**, 1-4.
Dorn A, Bollekens J, Staub A, Benoist C, Mathis D. 1987. A multiplicity of CCAAT box-binding proteins. *Cell* **50**, 863-72.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-7.
Edwards D, Murray JA, Smith AG. 1998. Multiple genes encoding the conserved CCAAT-box transcription factor complex are expressed in Arabidopsis. *Plant Physiol* **117**, 1015-22.
Frontini M, Imbriano C, Manni I, Mantovani R. 2004. Cell cycle regulation of NF-YC nuclear localization. *Cell Cycle* **3**, 217-22.
Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, Goodman HM. 1992. Isolation of the Arabidopsis ABI3 gene by positional cloning. *Plant Cell* **4**, 1251-61.
Gusmaroli G, Tonelli C, Mantovani R. 2002. Regulation of novel members of the Arabidopsis thaliana CCAAT-binding nuclear factor Y subunits. *Gene* **283**, 41-8.
Gutierrez L, Van Wuytswinkel O, Castelain M, Bellini C. 2007. Combined networks regulating seed maturation. *Trends Plant Sci* **12**, 294-300.

Hackenberg D, Wu Y, Voigt A, Adams R, Schramm P, Grimm B. 2011. Studies on Differential Nuclear Translocation Mechanism and Assembly of the Three Subunits of the Arabidopsis thaliana Transcription Factor NF-Y. *Mol Plant*.
Hammond-Kosack MC, Holdsworth MJ, Bevan MW. 1993. In vivo footprinting of a low molecular weight glutenin gene (LMWG-1D1) in wheat endosperm. *EMBO J* **12**, 545-54.
Harada JJ. 1997. Cellular and Molecular Biology of Plant Seed Development*Cellular and molecular* Dordrecht: Kluwer Academic, 545-92.
Holdsworth MJ, Bentsink L, Soppe WJ. 2008. Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination. *New Phytol* **179**, 33-54.
Kim IS, Sinha S, de Crombrugghe B, Maity SN. 1996. Determination of functional domains in the C subunit of the CCAAT-binding factor (CBF) necessary for formation of a CBF-DNA complex: CBF-B interacts simultaneously with both the CBF-A and CBF-C subunits to form a heterotrimeric CBF molecule. *Mol Cell Biol* **16**, 4003-13.
Kumimoto RW, Zhang Y, Siefers N, Holt BF, 3rd. 2010. NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in Arabidopsis thaliana. *Plant J*.
Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, Harada JJ. 2003. LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell* **15**, 5-18.
Lee H, Fischer RL, Goldberg RB, Harada JJ. 2003. Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. *Proc Natl Acad Sci U S A* **100**, 2152-6.
Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-2.
Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-8.
Lotan T, Ohto M, Yee KM, West MA, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ. 1998. Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* **93**, 1195-205.
Luerssen H, Kirik V, Herrmann P, Misera S. 1998. FUSCA3 encodes a protein with a conserved VP1/AB13-like B3 domain which is of functional importance for the regulation of seed maturation in Arabidopsis thaliana. *Plant J* **15**, 755-64.
Mantovani R. 1998. A survey of 178 NF-Y binding CCAAT boxes. *Nucleic Acids Res* **26**, 1135-43.
Marzabal P, Busk PK, Ludevid MD, Torrent M. 1998. The bifactorial endosperm box of gamma-zein gene: characterisation and function of the Pb3 and GZM cis-acting elements. *Plant J* **16**, 41-52.
Masiero S, Imbriano C, Ravasio F, Favaro R, Pelucchi N, Gorla MS, Mantovani R, Colombo L, Kater MM. 2002. Ternary complex formation between MADS-box transcription factors and the histone fold protein NF-YB. *J Biol Chem* **277**, 26429-35.
Meinke DW, Franzmann LH, Nickle TC, Yeung EC. 1994. Leafy Cotyledon Mutants of Arabidopsis. *Plant Cell* **6**, 1049-64.
Michel D, Salamini F, Bartels D, Dale P, Baga M, Szalay A. 1993. Analysis of a desiccation and ABA-responsive promoter isolated from the resurrection plant Craterostigma plantagineum. *Plant J* **4**, 29-40.
Mu J, Tan H, Zheng Q, Fu F, Liang Y, Zhang J, Yang X, Wang T, Chong K, Wang XJ, Zuo J. 2008. LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in Arabidopsis. *Plant Physiol* **148**, 1042-54.
Mundy J, Yamaguchi-Shinozaki K, Chua NH. 1990. Nuclear proteins bind conserved elements in the abscisic acid-responsive promoter of a rice rab gene. *Proc Natl Acad Sci U S A* **87**, 1406-10.
Parcy F, Valon C, Kohara A, Misera S, Giraudat J. 1997. The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in concert to control multiple aspects of Arabidopsis seed development. *Plant Cell* **9**, 1265-77.

Peng Y, Jahroudi N. 2002. The NFY transcription factor functions as a repressor and activator of the von Willebrand factor promoter. *Blood* **99**, 2408-17.
Peng Y, Jahroudi N. 2003. The NFY transcription factor inhibits von Willebrand factor promoter activation in non-endothelial cells through recruitment of histone deacetylases. *J Biol Chem* **278**, 8385-94.
Rogers JC, Rogers SW. 1992. Definition and functional implications of gibberellin and abscisic acid cis-acting hormone response complexes. *Plant Cell* **4**, 1443-51.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-4.
Sanjaya, Durrett TP, Weise SE, Benning C. 2011. Increasing the energy density of vegetative tissues by diverting carbon from starch to oil biosynthesis in transgenic Arabidopsis. *Plant Biotechnol J* **9**, 874-83.
Satoh S, Kamada H, Harada H, Fujii T. 1986. Auxin-controlled glycoprotein release into the medium of embryogenic carrot cells. *Plant Physiol* **81**, 931-3.
Schirawski J, Planchais S, Haenni AL. 2000. An improved protocol for the preparation of protoplasts from an established Arabidopsis thaliana cell suspension culture and infection with RNA of turnip yellow mosaic tymovirus: a simple and reliable method. *J Virol Methods* **86**, 85-94.
Shen B, Allen WB, Zheng P, Li C, Glassman K, Ranch J, Nubel D, Tarczynski MC. 2010. Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize. *Plant Physiol* **153**, 980-7.
Shiota H, Satoh R, Watabe K, Harada H, Kamada H. 1998. C-ABI3, the carrot homologue of the Arabidopsis ABI3, is expressed during both zygotic and somatic embryogenesis and functions in the regulation of embryo-specific ABA-inducible genes. *Plant Cell Physiol* **39**, 1184-93.
Siefers N, Dang KK, Kumimoto RW, Bynum WEt, Tayrose G, Holt BF, 3rd. 2009. Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. *Plant Physiol* **149**, 625-41.
Sinha S, Kim IS, Sohn KY, de Crombrugghe B, Maity SN. 1996. Three classes of mutations in the A subunit of the CCAAT-binding factor CBF delineate functional domains involved in the three-step assembly of the CBF-DNA complex. *Mol Cell Biol* **16**, 328-37.
Sinha S, Maity SN, Lu J, de Crombrugghe B. 1995. Recombinant rat CBF-C, the third subunit of CBF/NFY, allows formation of a protein-DNA complex with CBF-A and CBF-B and with yeast HAP2 and HAP3. *Proc Natl Acad Sci U S A* **92**, 1624-8.
Skriver K, Olsen FL, Rogers JC, Mundy J. 1991. cis-acting DNA elements responsive to gibberellin and its antagonist abscisic acid. *Proc Natl Acad Sci U S A* **88**, 7266-70.
Stephenson TJ, McIntyre CL, Collet C, Xue GP. 2007. Genome-wide identification and expression analysis of the NF-Y family of transcription factors in Triticum aestivum. *Plant Mol Biol* **65**, 77-92.
Stone SL, Braybrook SA, Paula SL, Kwong LW, Meuser J, Pelletier J, Hsieh TF, Fischer RL, Goldberg RB, Harada JJ. 2008. Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc Natl Acad Sci U S A* **105**, 3151-6.
Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, Goldberg RB, Harada JJ. 2001. LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proc Natl Acad Sci U S A* **98**, 11806-11.
Takaiwa F, Yamanouchi U, Yoshihara T, Washida H, Tanabe F, Kato A, Yamada K. 1996. Characterization of common cis-regulatory elements responsible for the endosperm-specific expression of members of the rice glutelin multigene family. *Plant Mol Biol* **30**, 1207-21.
Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-9.
Tan H, Yang X, Zhang F, Zheng X, Qu C, Mu J, Fu F, Li J, Guan R, Zhang H, Wang G, Zuo J. 2011. Enhanced seed oil production in canola by conditional expression of Brassica napus LEAFY COTYLEDON1 and LEC1-LIKE in developing seeds. *Plant Physiol* **156**, 1577-88.

Thirumurugan T, Ito Y, Kubo T, Serizawa A, Kurata N. 2008. Identification, characterization and interaction of HAP family genes in rice. *Mol Genet Genomics* **279**, 279-89.
Vicente-Carbajosa J, Moose SP, Parsons RL, Schmidt RJ. 1997. A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proc Natl Acad Sci U S A* **94**, 7685-90.
Voinnet O, Rivas S, Mestre P, Baulcombe D. 2003. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J* **33**, 949-56.
Wagner DB, Furnier GR, Saghai-Maroof MA, Williams SM, Dancik BP, Allard RW. 1987. Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. *Proc Natl Acad Sci U S A* **84**, 2097-100.
Walter M, Chaban C, Schutze K, Batistic O, Weckermann K, Nake C, Blazevic D, Grefen C, Schumacher K, Oecking C, Harter K, Kudla J. 2004. Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant J* **40**, 428-38.
Watanabe Y, Yamamoto S. 2009. Hōjinzei kihon tsūtatsu no gimonten. Tōkyō: Gyōsei.
Wobus U, Weber H. 1999. Seed maturation: genetic programmes and control signals. *Curr Opin Plant Biol* **2**, 33-8.
Wu CY, Suzuki A, Washida H, Takaiwa F. 1998. The GCN4 motif in a rice glutelin gene is essential for endosperm-specific gene expression and is activated by Opaque-2 in transgenic rice plants. *Plant J* **14**, 673-83.
Xi DM, Zheng CC. 2011. Transcriptional regulation of seed storage protein genes in Arabidopsis and cereals. *Seed Science Research* **21**, 247–54.
Xing Y, Fikes JD, Guarente L. 1993. Mutations in yeast HAP2/HAP3 define a hybrid CCAAT box binding domain. *EMBO J* **12**, 4647-55.
Yamaguchi-Shinozaki K, Mundy J, Chua NH. 1990. Four tightly linked rab genes are differentially expressed in rice. *Plant Mol Biol* **14**, 29-39.
Yamamoto A, Kagaya Y, Toyoshima R, Kagaya M, Takeda S, Hattori T. 2009. Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. *Plant J* **58**, 843-56.
Yazawa K, Kamada H. 2007. Identification and characterization of carrot HAP factors that form a complex with the embryo-specific transcription factor C-LEC1. *J Exp Bot* **58**, 3819-28.
Yazawa K, Takahata K, Kamada H. 2004. Isolation of the gene encoding Carrot leafy cotyledon1 and expression analysis during somatic and zygotic embryogenesis. *Plant Physiol Biochem* **42**, 215-23.
Yu Y, Li Y, Huang G, Meng Z, Zhang D, Wei J, Yan K, Zheng C, Zhang L. 2011. PwHAP5, a CCAAT-binding transcription factor, interacts with PwFKBP12 and plays a role in pollen tube growth orientation in Picea wilsonii. *J Exp Bot* **62**, 4805-17.

4.5 Capítulo 5

Identificação dos genes alvo diretos de LEC1 *(*LEAFY COTYLEDON1*)* durante os estágios cotiledonar e de início da maturação em soja (*Glycine Max* (L.) Merrill)

Alexandro Cagliari[1,2 §], Julie Pelletier[2], Russell Baden[2] , Marcia Margis-Pinheiro[1] e John Harada[2]

[1]Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Brasil.

[2]College of Biological Sciences, Universidade da Califórnia – Davis – EUA.

[§]Autor para correspondência:

Alexandro Cagliari. Departamento de Genética. Universidade Federal do Rio Grande do Sul. Avenida Bento Gonçalves 9500. Prédio 43.312. CEP 91501-970. Porto Alegre, RS, Brasil. Phone: +55 (51) 33089830. E-mail: alexandrocagliari@yahoo.com.br. Homepage: http://www.ufrgs.br/rnai/ngfp/

Resumo

LEC1 (LEAFY COTYLEDON1) codifica para uma das subunidades (NF-YB) que constituem o fator de transcrição eucariótico NF-Y (*Nuclear Factor of the Y box*) e apresenta importante função tanto durante a embriogênese quanto na maturação da semente. Porém, até o momento não existem informações sobre os genes alvo de LEC1 ao longo do desenvolvimento da semente. O objetivo deste trabalho foi identificar os genes alvo diretos de LEC1 no início do desenvolvimento embriogênico (estágio cotiledonar) e durante o início da maturação, de forma comparar os genes alvo de LEC1 das fases iniciais e finais do desenvolvimento da semente de soja. Os resultados demonstram que o estágio cotiledonar e de início da maturação, apesar de apresentarem muitos genes alvo diretos induzidos e reprimidos compartilhados, também apresentam características distintas quanto ao genes alvo diretos induzidos por LEC1: enquanto no estágio cotiledonar os genes alvo diretos são representados por genes envolvidos em ciclo celular, remodelação da cromatina e ativação transcricional, durante o início da maturação observamos um enriquecimento em processos como metabolismo e acúmulo de substâncias de reserva. Além disso, observamos que LEC1 se liga não somente a promotores enriquecidos na sequencia canônica CCAAT como também na sequencia típica de ligação de fatores de transcrição do tipo bZIP, fato que também foi observado para Arabidopsis e, parece indicar que fatores de transcrição do tipo bZIP podem estar diretamente relacionados com a regulação do desenvolvimento e maturação da semente exercida por LEC1.

PALAVRAS-CHAVE: Leafy cotyledon1 (LEC1), Maturação da semente, soja (*Glycine max*), regulação da expressão gênica.

INTRODUÇÃO

O desenvolvimento da semente é um processo crucial no ciclo de vida das plantas superiores, possibilitando a ligação entre duas gerações esporofíticas distintas e, portanto, promovendo a manutenção da espécie (Gutierrez *et al.*, 2007).

O desenvolvimento de sementes pode ser dividido em morfogênese e maturação (West and Harada, 1993). Nos estágios iniciais da embriogênese, o embrião passa por uma fase de intensa divisão celular e morfogênese, seguida por uma fase de maturação, a qual inclui o acúmulo de substâncias de reserva, a supressão da germinação, a aquisição de tolerância à dessecação, perda de água e, geralmente, passa por um processo de indução de dormência (Wobus and Weber, 1999). As proteínas, bem como lipídeos, amidos e açúcares acumulados durante a fase maturação têm como finalidade a proteção do embrião contra a dessecação bem como o fornecimento dos nutrientes necessários para o estabelecimento da plântula após o início da germinação (Braybrook and Harada, 2008). Uma complexa rede regulatória desencadeia o início da maturação e o acúmulo de substâncias de reserva. Esta inclui uma reprogramação transcricional e fisiológica por meio de rotas responsivas, principalmente, açúcares e hormônios (Gibson, 2004).

Inúmeros fatores de transcrição estão envolvidos com o controle da fase de maturação da semente, dentre os quais podemos citar LEAFY COTYLEDON1 (LEC1), LEC1-LIKE (L1L), LEAFY COTYLEDON2 (LEC2), FUSCA3 (FUS3), WRINKLED1 (WRI1), ABSCICSIC ACID INSENSITIVE3 (ABI3) (Giraudat *et al.*, 1992; Gusmaroli *et al.*, 2002; Lotan *et al.*, 1998; Meinke *et al.*, 1994; Parcy *et al.*, 1997; Sanjaya *et al.*, 2011; Stone *et al.*, 2008; Wobus and Weber, 1999; Yazawa *et al.*, 2004).

LEC1 codifica para uma das subunidades (NF-YB) que constituem o fator de transcrição eucariótico NF-Y (*Nuclear Factor of the Y box*) (Dorn *et al.*, 1987; Stephenson *et al.*, 2007). Em mamíferos, as três subunidades (NF-YA, NF-YB e NF-YC) do fator de transcrição NF-Y formam um oligotrímero e são necessárias para a ligação à sequencia CCAAT localizada nos promotores de seus genes alvo (Sinha *et al.*, 1995).

LEC1 apresenta importante função tanto durante a embriogênese quanto na maturação da semente (Kwong *et al.*, 2003; Lee *et al.*, 2003; Lotan *et al.*, 1998; Watanabe and Yamamoto, 2009; Yu *et al.*, 2011). Durante a fase de maturação da semente, inúmeras evidências apontam para um importante papel direto ou indireto de LEC1 na ativação de genes de armazenagem, responsáveis pelo acúmulo de macromoléculas de reserva (Braybrook and Harada, 2008) e também pela biossíntese de ácidos graxos e triacilglicerídeos (Mu *et al.*, 2008; Shen *et al.*, 2010; Tan *et al.*, 2011).

A soja (Glycine max) representa uma cultura de grande importância econômica no cenário mundial, sendo utilizada para prover alimentos tanto para humanos quanto para animais (Duranti, 2006). O aumento no número de estudos sobre o desenvolvimento da semente de soja utilizando tecnologias tais como microarranjo (Le *et al.*, 2007) e sequenciamento de cDNA (Sha *et al.*, 2012) enfatizam a importância do conhecimento dos mecanismos genéticos atuantes para o desenvolvimento da semente dessa importante oleaginosa (Jones and Vodkin, 2013).

Dados de análise de expressão por microarranjo obtidos anteriormente pelo grupo do Dr. Harada (Universidade da California – UC – Davis) e do Dr. Goldberg (UC – Los Angeles) ao longo do desenvolvimento da semente de soja demonstram que LEC1 é expresso desde o estágio globular até as etapas finais da maturação da

semente (Le *et al.*, 2007) (Figura 1). Por outro lado, proteínas de reserva em sementes tem sua expressão detectada somente a partir do início da maturação da semente de soja (Le *et al.*, 2007) (Figura 1).

Muito já se sabe quanto à importância de LEC1 para a fase de maturação da semente. Porém, até o momento não existem informações sobre os genes alvo de LEC1 durante esta fase de acúmulo de substâncias de reserva bem como em fases anteriores ao processo de maturação da semente (Figura 1).

Portanto, o objetivo deste capítulo foi identificar os genes alvo diretos de LEC1 no início do desenvolvimento embriogênico (estágio cotiledonar) e durante o início da maturação, de forma comparar os genes alvo de LEC1 de fases iniciais e finais ao longo do desenvolvimento da semente de soja.

RESULTADOS E DISCUSSÃO

Genes modulados por LEC1 antes e durante o processo de maturação da semente de soja

O fato de que em Arabidopsis a superexpressão constitutiva de LEC1 apresenta caráter letal, associado à inexistência de mutantes LEC1 em soja, torna difícil a identificação direta dos genes modulados por LEC1. Devido a estas limitações, nós delineamos uma estratégia visando a identificação dos genes cuja expressão é modulada por LEC1 antes e durante o processo de maturação baseada em análise de co-expressão em regiões da semente onde LEC1 apresenta ou não expressão detectável.

Para tanto, os dados de expressão por RNA-seq foram obtidos por microdissecção e captura a lazer (LCM) de diferentes compartimentos de sementes de soja (contrastantes quanto à presença ou ausência de LEC1) nos estágios cotiledonar (embrião (EP), tegumento interno (II) e tegumento externo (OI)) e de início da maturação (eixo do parênquima (AXPY), parênquima abaxial cotiledonar (ABAPY), compartimento adaxial cotiledonar (ADAPY), tegumento do parênquima (SCPY) e tegumento palisádico (SCPS)) (Figura 2).

No estágio cotiledonar, para que um gene fosse considerado induzido na presença de LEC1, este deveria apresentar sua expressão induzida no EP (onde LEC1 é expresso) quando comparado a II e OI (onde LEC1 não é expresso). Por outro lado, para que um gene fosse considerado reprimido na presença de LEC1, este deveria apresentar sua expressão reprimida em EP quando comparado a II e OI (Figura 2).

No estágio inicial da maturação, para que um gene fosse considerado induzido por LEC1 este deveria apresentar sua expressão induzida em AXPY, ABPY e ADPY (onde LEC1 é expresso) quando comparado a SCPY e SCPS (onde LEC1 não é expresso). Por outro lado, para que um gene fosse considerado reprimido por LEC1 este deveria apresentar sua expressão reprimida no AXPY, ABPY e ADPY quando comparado a SCPY e SCPS (Figura 2).

Considerando tais critérios de modulação da expressão, análises de RNA-seq mostraram que, do total de genes induzidos por LEC1, 1558 foram induzidos tanto no estágio cotiledonar quanto no início da maturação, 1536 foram induzidos de forma exclusiva durante o estágio cotiledonar e 1881 foram induzidos somente durante o início da maturação (Figura 3A). Por outro lado, 1307 genes foram reprimidos no estágio cotiledonar, 2196 durante o início da maturação e 1207 genes foram reprimidos pela presença de LEC1 em ambos os estágios analisados (Figura 3B).

Dados de ChIP-seq utilizando anticorpo anti-LEC1 demonstraram que o fator de transcrição LEC1 se liga a 24.507 genes distintos durante o estágio cotiledonar e 24.665 genes durante o início da maturação (Figura 3C), em ambos os casos com preferência de ligação em regiões próximas ao códon de início do gene (ATG) (Figura S1). Deste total, 20.382 genes são ligados por LEC1 tanto no estágio cotiledonar quanto de início da maturação. Por outro lado, 4.125 genes são alvos de ligação de LEC1 exclusivamente durante o estágio cotiledonar e 4.203 apenas no início da maturação (Figura 3C).

Genes Alvo diretos de LEC1 antes e durante a maturação da semente

Os dados relacionados à sobreposição entre os genes modulados por LEC1 (RNA-seq) e os genes ligados por LEC1 (ChIP-seq) nos permitem identificar os genes alvo diretos de LEC1, ou seja, os genes tanto modulados quanto ligados por este fator de transcrição (Figura 4).

Durante o estágio cotiledonar, LEC1 apresentou 3.094 alvos diretos induzidos e 2.514 alvos diretos reprimidos (Figuras 4A e 4B). Quanto ao estágio de início da maturação, 3.439 genes foram caracterizados como alvos diretamente induzidos, e 3.404 como alvos diretamente reprimidos pelo fator de transcrição LEC1 (Figuras 4C e 4D).

Considerando o total de genes alvo diretamente induzidos por LEC1, 1.538 são específicos do estágio cotiledonar, 1.883 específicos do início da maturação e 1.556 compartilhados em ambos os estágios (Figura 5A).

Dentre os genes alvo diretamente induzidos e reprimidos por LEC1, seja de forma estágio-exclusiva (cotiledonar ou início da maturação) ou de forma sobreposta em ambos os estágios de desenvolvimento, observamos a presença de inúmeros genes codificantes para as três subunidades que compõem o fator de transcrição NF-Y (NF-YA, NF-YB e NF-YC) (dados não mostrados). Isso demonstra que LEC1 não só tem a capacidade de interagir com os outros membros das outras subunidades do fator de transcrição NF-Y (Calvenzani *et al.*, 2012; Kumimoto *et al.*, 2010) como também, no caso de soja, modula a expressão de tais genes através da ligação a seus promotores.

Quanto aos genes alvo estágio-específicos, análises de ontologia gênica demonstraram que os alvos diretos induzidos e específicos do estágio cotiledonar

estão envolvidos em processos de montagem e organização de cromatina e cromossomos, morfogênese e ciclo celular, interação proteína-DNA e ativação da transcrição, dentre outros (Figura 6).

Por outro lado, os alvos diretos induzidos e específicos de LEC1 durante o início da maturação estão enriquecidos em genes envolvidos em processos metabólicos característicos da fase de maturação da semente (figura 7) tais como morfogênese pós-embriônica, metabolismo de carboidratos e lipídeos (ácido graxo desaturases - Glyma19g32940, Glyma02g39230, Glyma14g27990, Glyma19g32940; tioesterases - Glyma08g46360; ácido graxo elongases - Glyma04g06110; ácido graxo sintetases - Glyma04g36950, Glyma05g27260; aciltransferases - Glyma01g01350, Glyma16g25690, Glyma09g07520, Glyma14g09680; ácido graxo hidroxilases - Glyma01g0417, Glyma07g2967, Glyma08g0858, Glyma08g46360, Glyma09g24530, Glyma09g40550), resposta mediada por açúcares e reserva de nutrientes (proteínas abundantes no fim da embriogênese - Glyma19g32920, Glyma03g30040; glicininas - Glyma10g04280, Glyma03g32030, Glyma13g18450; albuminas - Glyma06g07070, Glyma08g43375,Glyma08g45820, Glyma11g12810, Glyma11g36460, Glyma17g14930, Glyma18g26120, Glyma18g41590; cruciferina - Glyma19g34770), dentre outros.

Portanto, observamos que o estágio cotiledonar e de início da transcrição, embora apresentem muitos genes alvo de LEC1 compartilhados, também apresentam características distintas quanto ao genes alvo diretos induzidos por LEC1: enquanto no estágio cotiledonar os genes alvo diretos são representados por genes envolvidos em ciclo celular, remodelação da cromatina e ativação transcricional (Figura 8A), durante o início da maturação observamos um enriquecimento em processos como metabolismo e acúmulo de substâncias de reserva (Figura 8C).

Por outro lado, os genes alvo diretos de LEC1 compartilhados nos estágios cotiledonar e de início da maturação encontram-se enriquecidos em categorias ontológicas tais como fotossíntese e processos reprodutivos celulares (Figura 8B). Além disso, genes codificantes de oleosinas (Glyma17g13120, Glyma06g23340, Glyma20g33850, Glyma14g15015), envolvidos na formação dos oleossomos (principal estrutura de armazenagem de óleo em sementes), também são alvos diretos induzidos por LEC1, tanto durante o estágio cotiledonar quanto durante o início da maturação da semente de soja (Figura 8B).

Corroborando com a importância de LEC1 tanto na fase de desenvolvimento do embrião quanto na fase de maturação, nós observamos que, dentre os genes alvo diretos induzidos por LEC1 e comuns aos estágios cotiledonar e de início da maturação, podemos citar importantes fatores de transcrição envolvidos com desenvolvimento e maturação da semente tais como FUS3 (Glyma16g05480), ABI3 (Glyma08g47240 e Glyma18g38490), LEC1 (Glyma07g39820), WR1 (Glyma08g24420 e Glyma15g34770) e bZIP67 (Glyma13g39340). Análises de ChIP seguido de PCR quantitativo (ChiP-qPCR) foram realizadas para validação da ligação de LEC1 a alguns desses fatores de transcrição (FUS3, ABI3 e LEC1) tanto durante o estágio cotiledonar quanto de início da maturação (Figura 9).

Quanto aos genes alvo diretos reprimidos por LEC1, 1.306 genes são exclusivos do estágio cotiledonar, 2.195 do início da maturação e 1.208 são alvos diretos reprimidos em ambos os estágios (Figura 5B). Muitas das categorias ontológicas nas quais os genes alvo diretos reprimidos por LEC1 foram classificados são comuns tanto a genes estágio-específicos quanto a genes comuns a ambos os estágios (Figuras 10, 11 e 12), mostrando mais uma vez que os dois estágios de desenvolvimento da semente de soja compartilham muitos dos genes alvo de LEC1.

Quanto às categorias ontológicas exclusivas, dentre os genes alvo diretos reprimidos em ambos os estágios podemos citar genes envolvidos com ritmo circadiano, transcrição e resposta a giberilina (Figura 10). Quanto aos genes alvo diretos reprimidos no estágio cotiledonar podemos citar os envolvidos em transporte transmembrana, retículo endoplasmático, peroxissomos, endossomos, vacúolo enquanto que os genes alvo diretos reprimidos no estágio de início da maturação incluem os genes envolvidos em processos tais como fosforilação, diferenciação celular, resposta a estímulo por luz, resposta a água, resposta a diferentes hormônios (ácido abscísico, ácido jasmônico, auxina), desenvolvimento de pólen e órgão florais, desenvolvimento de folhas raiz e parede celular, etc (Figura 12).

LEC1 se liga preferencialmente a sequencias de DNA ricas em motivos de ligação para fatores de transcrição do tipo bZIP

Os principais motivos de DNA identificados nos promotores dos genes aos quais a proteína LEC1 se liga estão apresentados nas figuras 13 e 14. Em mamíferos, é sabido que o fator de transcrição NF-Y, do qual LEC1 representa uma das subunidades, liga-se a sequencia nucleotídica CCAAT dos promotores de seus genes alvo (Dorn *et al.*, 1987). Em concordância com esse fato, observamos que o sequencia CCAAT encontra-se enriquecida nos promotores dos genes alvo diretos induzidos e reprimidos por LEC1 durante o estágio cotiledonar (Figura 13). Quanto ao estágio inicial da maturação, observamos a presença do motivo CCAAT apenas nos promotores dos genes alvo diretos reprimidos por LEC1 (Figura 14).

Considerando o enriquecimento da sequencia CCAAT em genes alvo de LEC1 e que em mamíferos as três subunidades do fator de transcrição NF-Y são

necessárias para a ligação à sequencia CCAAT (Sinha *et al.*, 1995), uma questão interessante é se LEC1 está se ligando diretamente ou por meio da interação com outras subunidades NF-Y à sequencia CCAAT de seus genes alvo. Neste contexto, experimentos de interação proteína-proteína podem ser úteis para identificar quais as proteínas que interagem com LEC1 nesses estágios de desenvolvimento de soja e experimentos de Re-ChIP (Truax and Greer, 2012) podem auxiliar na identificação de múltiplas proteínas possivelmente ligadas à sequencia CCAAT.

Interessantemente, a presença do motivo de ligação de fatores de transcrição do tipo bZIP (AGCG) (Izawa *et al.*, 1993) foi a mais frequente nos promotores dos genes alvo diretos, tanto induzidos quanto reprimidos por LEC1, e em ambos os estágios cotiledonar e início da maturação (Figuras 13 e 14). Esses resultados estão de acordo com resultados prévios e ainda não publicados pelo grupo do Dr. Harada em Arabidopsis onde, experimentos de ChIP-seq demonstraram que LEC1 se liga a promotores ricos em motivos de ligação para fatores de transcrição do tipo bZIP e não em motivos CCAAT.

Experimentos de transativação utilizando protoplastos de Arabidopsis haviam previamente demonstrado que LEC1 ativa a expressão de genes responsivos a ABA (ácido abscísico) através de interação com bZIP67 (Watanabe and Yamamoto, 2009; Yamamoto *et al.*, 2009). Portanto, além da sinalização mediada por ABA estar intimamente relacionada com a expressão de LEC1 (Finkelstein *et al.*, 2002), nós também observamos que muitos dos alvos diretos de LEC1 em soja e em Arabidopsis (dados não publicados) apresentam motivos conservados de ligação de fatores de transcrição do tipo bZIP. Isso demonstra que a relação entre ABA, fatores de transcrição bZIP e LEC1 parece ser muito importante no controle das fases de desenvolvimento e da maturação de sementes.

Em conclusão, nós observamos que LEC1 apresenta uma grande quantidade de genes alvo induzidos e principalmente reprimidos comuns às duas fases de desenvolvimento da semente analisadas. Porém, sobretudo quanto aos genes alvo diretos induzidos, LEC1 também apresenta enriquecimento em diferentes processos metabólicos durante os estágios cotiledonar e de início da maturação da semente de soja: enquanto no estágio cotiledonar os genes alvo diretos são representados por genes envolvidos em ciclo celular, remodelação da cromatina e ativação transcricional durante o início da maturação ocorre enriquecimento em processos como metabolismo e acúmulo de substâncias de reserva, típicos da fase de maturação de sementes. Além disso, LEC1 se ligou não somente a promotores enriquecidos na sequencia CCAAT como também na sequencia típica de ligação de fatores de transcrição do tipo bZIP.

Com base nas informações acima citadas, especula-se que: i) LEC1 estaria se ligando diretamente ao motivo bZIP, em adição ao motivo canônico CCAAT, ii) fatores de transcrição bZIP estariam se ligando ao DNA e recrutando LEC1, iii) LEC1 estaria interagindo com fatores de transcrição bZIP e esta interação, então, permitiria a ligação de LEC1 ao promotor de seus genes alvo. Neste cenário, ensaios de ligação ao DNA e de interação proteína-proteína visando elucidar a relação de LEC1 com fatores de transcrição bZIP (especialmente bZIP67) e com as outras subunidades que compõem o fator de transcrição NF-Y (NF-YA e NF-YC), utilizando soja e Arabidopsis como modelos de estudo, podem contribuir para elucidar os mecanismos regulatórios envolvidos no processo de controle do desenvolvimento e maturação de sementes.

MATERIAL E MÉTODOS

Analise do transcritoma por RNA-seq

Os dados de expressão por RNA-seq foram obtidos por microdissecção e captura a lazer (LCM) de diferentes tecidos de sementes de soja no estágio cotiledonar [embrião (EP), tegumento interno (II) e tegumento externo (OI)] e no início da maturação [eixo do parênquima (AXPY), parênquima abaxial cotiledonar (ABAPY), compartimento adaxial cotiledonar (ADAPY), tegumento do parênquima (SCPY) e tegumento palisádico (SCPS)]. A coleta e fixação das amostras foi adaptada para soja a partir de protocolo previamente estabelecido para Arabidopsis pelo laboratório do Dr. Harada (Belmonte *et al.*, 2013). RNA total (10µg) foi utilizado para síntese de cDNA e preparação das bibliotecas para sequenciamento utilizando Illumina Hiseq 2000 (Illumina CO).

Quanto ao mapeamento, as sequencias foram alinhados utilizando o programa Bowtie v 0.12.7 permitindo dois erros (mismatches) entre as sequencias analisadas. Como sequencia de refência foi utilizado o genoma de soja disponível no banco de dados Phytozome (http://www.phytozome.net).

As sequencias obtidas foram analisadas utilizando o pacote de programa EdgeR (Robinson *et al.*, 2010). Foram considerados diferencialmente expressos os genes que apresentaram uma expressão duas vezes maior ou menor em diferentes situações (fold *change* > 2), considerando um FDR (*False Discovery Rate*) <0.01.

ChIP-seq

Os experimentos de imunoprecipitação da cromatina seguida por sequenciamento (ChIP-seq) foram executados de acordo com protocolo previamente descrito (Jeong *et al.*, 2011). Foram utilizadas duplicatas experimentais para os estágios cotiledonar e de início da maturação do embrião de soja. Para cada amostra biológica foram utilizados 130 embriões no estágio cotiledonar e 10 embriões do estágio de inicio da maturação, os quais permitiram o isolamento de cerca de 20-30 µg de cromatina total de cada replicata, suficientes para o realização das etapas posteriores de ChIP.

Para cada estágio analisado (cotiledonar e início da maturação), sequencias com 50 pb provenientes de duas replicatas biológicas do Input (cromatina total) e de duas replicatas biológicas de amostras de ChIP utilizando anticorpo anti-LEC1 foram alinhadas ao genoma de soja (v1.0) usando o programa bowtie versão 0.12.7, e permitindo dois erros (*mismatches*) entre as sequencias alinhadas. Apenas sequencias que mapearam de forma única no genoma foram utilizadas para análises futuras. Tais sequencias únicas foram filtradas para remover sequencias duplicadas (cópias duplicadas por PCR) usando o programa SAmtools. As regiões estatisticamente significativas de ligação de LEC1 foram identificadas usando o programa CisGenome (versão 2), com um ponto de corte de $p<1e-05$.

Os genes aos quais LEC1 se liga foram identificados usando os programas para anotação gênica Bedtools e Phytozome (versão 9). Para tanto, genes com uma região de ligação dentro de 2Kb à montante da anotação para o início do gene, dentro da região codificante do genes ou a menos de 0.5 kb à jusante do término da porção codificante do gene foram considerados genes ao qual LEC1 se liga.

CHIP-qPCR

Os experimentos de ChIP foram realizados utilizando as mesmas condições descritas anteriormente e em duplicatas biológicas. Os produtos de ChIP foram analisados por PCR quantitativo em tempo real (qPCR) utilizando os oligonucleotídeos iniciadores listados na tabela S1.

As análises por qPCR foram realizadas em um equipamento *Applied Biosystem* (*StepOne Plus*). As reações foram compostas de uma desnaturação inicial a 94°C por 5 min, seguido por 40 ciclos de 10 s a 94°C, 15 s a 60°C e 15 s a 72°C. Após, as amostras foram aquecidas de 55 para 99°C com um aumento de 0,1°C/s para adquirir os dados produzidos pela curva de desnaturação dos produtos amplificados. qPCRs foram realizados em um volume final de 20 µL composto de 10 µL de cada amostra de cDNA diluída de 50 a 100 vezes em 2 µL de *Platinum Taq* 10x PCR *buffer*, 1,2 µL $MgCl_2$ 50 mM, 0,4 µL dNTPs 5 mM, 0,4 µL do par de oligonucleotídeos a 10 µM, 3,95 µL H_2O, 2,0 µL *SYBR green* (1:100,000 - *Molecular Probe*, OR, USA), e 0.05 µL *Platinum Taq* DNA *polymerase* (5 U/µl) (Invitrogen CA, USA). A temperatura de anelamento de todos os oligonucleotídeos foi ajustada para 60°C.

A quantificação da expressão dos genes de interesse foi realizada através do método de quantificação relativa 2-ΔΔCt (Livak and Schmittgen, 2001; Schmittgen and Livak, 2008), com a utilização genes constitutivos (Libault *et al.*, 2008).

Identificação genes alvo diretos de LEC1

A intersecção entre os genes regulados e ligados por LEC1 (alvos diretos) foi feita utilizando o programa Venny (http://bioinfogp.cnb.csic.es/tools/venny/index.html).

A estatística da intersecção (fator de representação e *p-value* hipergeométrico) dos genes foi feita utilizando ferramentas disponíveis no site (http://nemates.org).

Identificação de motivos de ligação de LEC1

As sequencias de DNA associadas com alvos diretos de LEC1 foram extraídas e analisadas usando o programa MEME-ChIP (http://meme.nbcr.net/meme/cgi-bin/meme-chip.cgi) para a identificação de motivos enriquecidos.

Ontologia gênica

As análises de ontologia gênica e de PAGE (*Parametric Analysis of Gene Set Enrichment*) foram realizadas com o software on-line AgriGO (Du *et al.*, 2010) utilizando o teste estatístico hipergeométrico, multi-teste de ajustamento de Hochberg (FDR) e com seleção utilizando um nível de significância de < 0.01.

FIGURAS

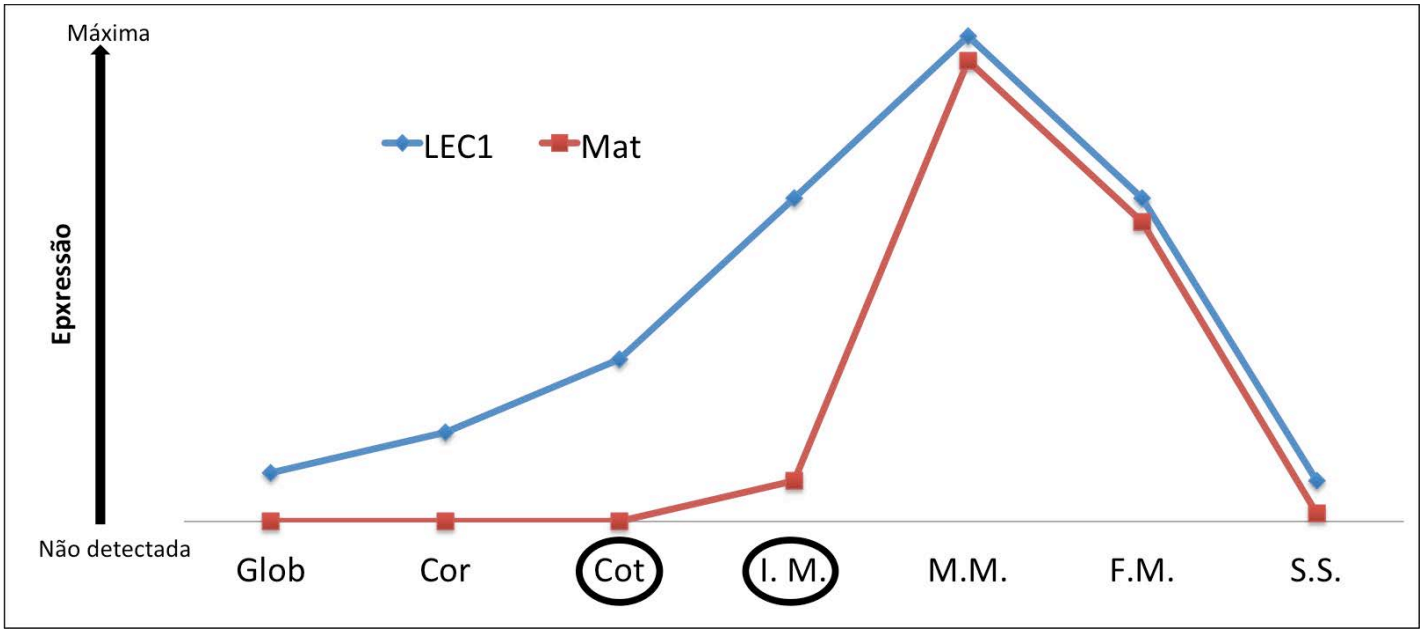


Figura 1. Representação esquemática (baseada em dados de RNA-seq) do perfil de expressão de LEC1 e de proteínas de reserva (Mat) ao longo do desenvolvimento da semente de soja. Enquanto a expressão de LEC1 começa e ser detectada no estágio globular (Glob) a expressão de proteínas de armazenamento (Mat) começa a ser detectada durante o início da maturação (I.M.). Círculos pretos indicam fases escolhidas para o estudo de genes alvo de LEC1. (Glob: globular; Cor: coração; Cot: cotiledonar; I.M.: início da maturação; M.M: meio da maturação; F.M.: final da maturação; S.S.: semente seca.

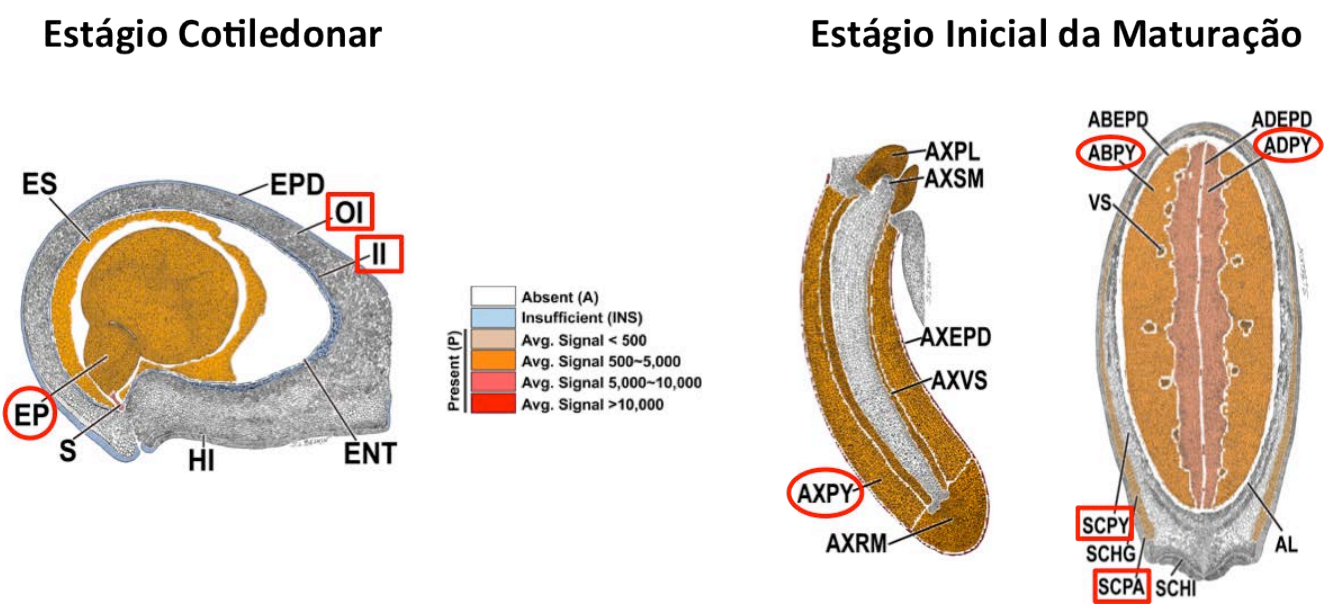


Figura 2. Perfil de expressão de LEC1 em diferentes compartimentos de sementes de soja nos estágios cotiledonar e de início da maturação utilizando dados de microarranjo depositados no banco *Gene Networks in Seed development* (http://seedgenenetwork.net). Círculos e retângulos vermelhos indicam compartimentos isolados por microdissecção e captura a lazer (LCM) para as análises de expressão por RNA-seq onde LEC1 apresenta e não apresenta expressão, respectivamente. Estágio cotiledonar: embrião (EP), tegumento interno (II) e tegumento externo (OI); início da maturação: eixo do parênquima (AXPY), parênquima abaxial cotiledonar (ABAPY), compartimento adaxial cotiledonar (ADAPY), tegumento do parênquima (SCPY) e tegumento palisádico (SCPS). A legenda traz a média da intensidade do sinal detectado pela análise de microarranjo, que reflete o nível de expressão.

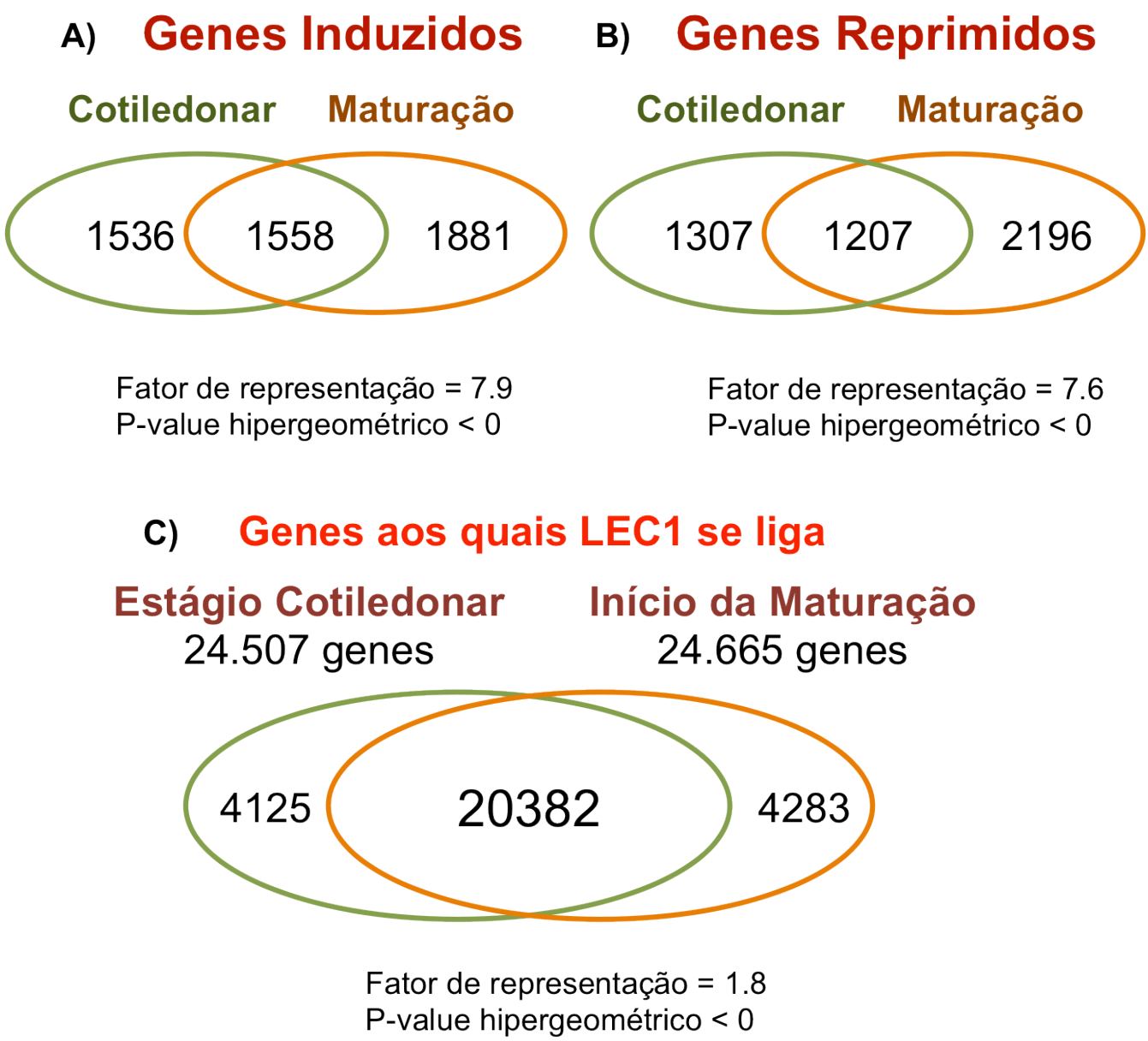


Figura 3. Genes modulados pela expressão de LEC1. A) Genes induzidos pela expressão de LEC1 em cada estágio (cotiledonar e início da maturação) e sua sobreposição. **B)** Genes reprimidos pela expressão de LEC1 em cada estágio (cotiledonar e início da maturação) e sua sobreposição. **C)** Genes aos quais LEC1 se liga em cada estágio (cotiledonar e início da maturação) e sua sobreposição. Um fator de representação >1 indica que a intersecção observada (genes comuns ao estágio cotiledonar e de início da maturação) é maior do que a esperada para dois grupos aleatórios de genes, considerando o tamanho do genoma.

Figura 4. Genes alvos diretos de LEC1. A) Genes alvos diretos induzidos pela expressão de LEC1 no estágio cotiledonar. **B)** Genes alvos diretos reprimidos pela expressão de LEC1 no estágio cotiledonar. **C)** Genes alvos diretos induzidos pela expressão de LEC1 no estágio de início da maturação. **D)** Genes alvos diretos reprimidos pela expressão de LEC1 no estágio de início da maturação. Um fator de representação >1 indica que a intersecção observada (genes induzidos e ligados por LEC1) é maior do que a esperada para dois grupos aleatórios de genes, considerando o tamanho do genoma.

Figura 5. Alvos diretos de LEC1 em ambos os estágio cotiledonar e de início da maturação. A) Genes alvos diretos induzidos pela expressão de LEC1 em ambos os estágios cotiledonar e de início da maturação. **B)** Genes alvos diretos reprimidos pela expressão de LEC1 em ambos os estágios cotiledonar e de início da maturação. Um fator de representação >1 indica que a intersecção observada (genes comuns ao estágio cotiledonar e de início da maturação) é maior do que a esperada para dois grupos aleatórios de genes, considerando o tamanho do genoma.

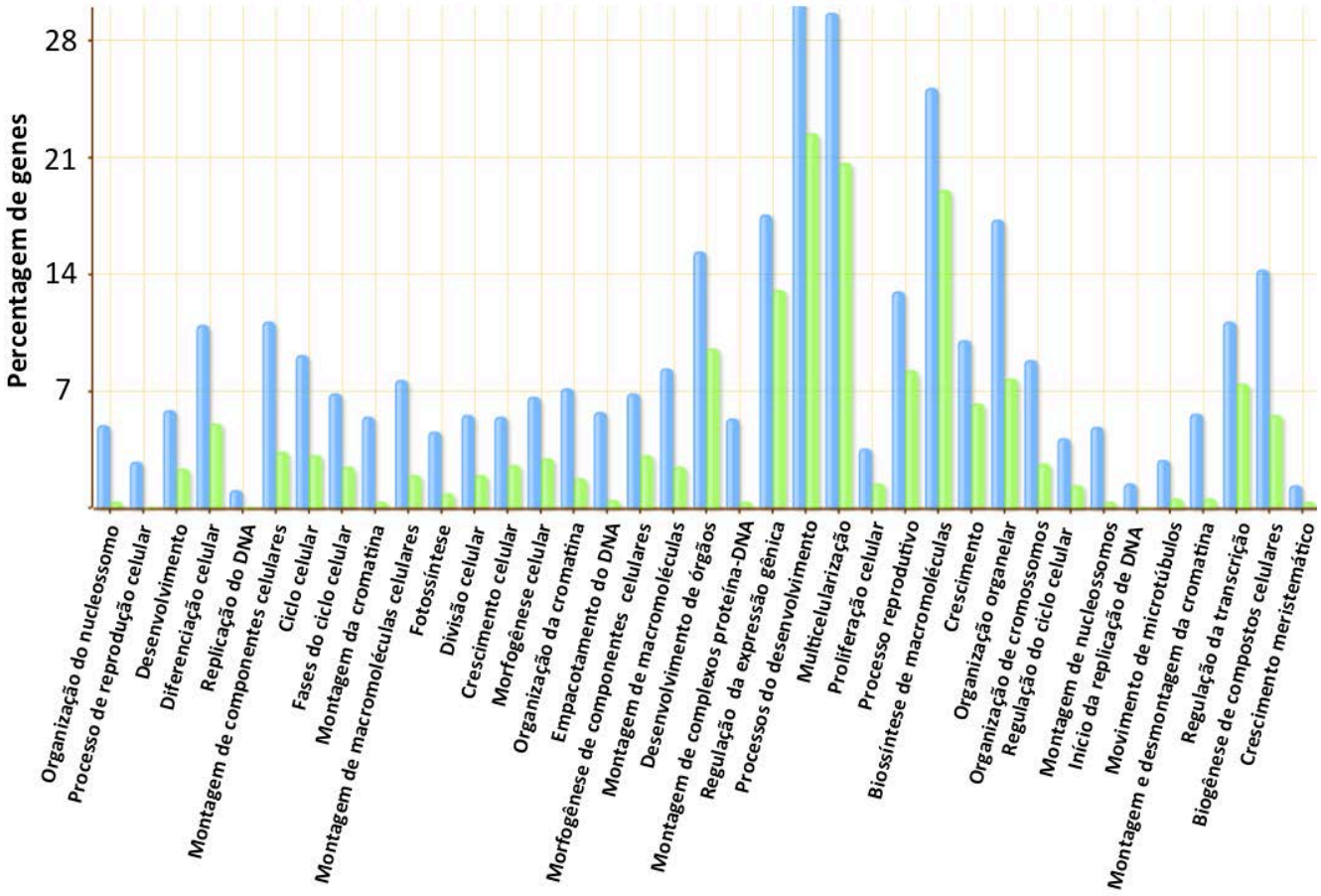


Figura 6. Análise ontológica dos genes alvos diretos induzidos durante o estágio cotiledonar. As barras azuis representam a percentagem de genes alvo diretos induzidos por LEC1 e envolvidos nos diferentes processos metabólicos (estatisticamente significativos, ver material e métodos) durante o estágio cotiledonar. As barras verdes representam a percentagem de genes descritos em soja para cada processo metabólico quando comparado ao número total de genes do genoma da espécie (*background*). A comparação entre barras azuis e verdes permite a visualização do enriquecimento de genes alvos de LEC1 em determinada categoria metabólica.

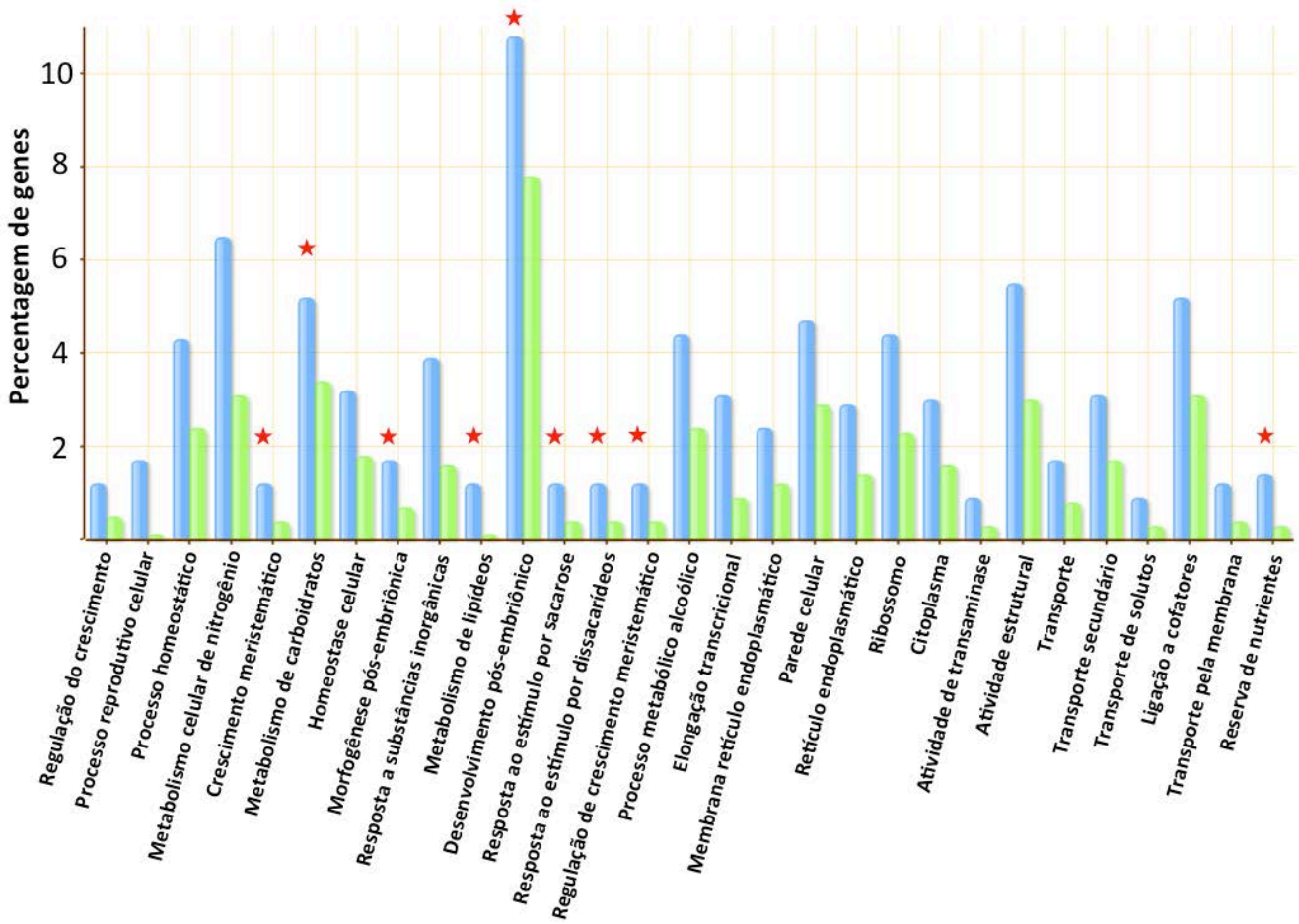


Figura 7. Análise ontológica dos genes alvos diretos induzidos durante o estágio de início da maturação. As barras azuis representam a percentagem de genes alvo diretos induzidos por LEC1 e envolvidos nos diferentes processos metabólicos (estatisticamente significativos, ver material e métodos) durante o início da maturação. As barras verdes representam a percentagem de genes descritos em soja para cada processo metabólico quando comparado ao número total de genes do genoma da espécie (*background*). A comparação entre barras azuis e verdes permite a visualização do enriquecimento de genes alvos de LEC1 em determinada categoria metabólica. As estrelas vermelhas chamam a atenção para processos metabólicos característicos na fase de maturação da semente.

A)

Ontologia	1	2	1- Cotiledonar P-value	1- Cotiledonar No de Genes	2- Maturação P-value	2- Maturação No de Genes
Regulação da transcrição			1.5e-14	442	0.0026	573
Transcrição			5.8e-13	463	0.0079	611
Crescimento embriogênico			2e-11	135	1.2-06	179
Padrão de formação abaxial/adaxial			4.9e-11	32	9.6-06	33
Regulação da transcrição			5.4e-09	277	---	---
Desenvolvimento meristemático			1.4e-08	53	0.00059	61
Crescimento meristemático			8.7e-07	34	0.0015	39

B)

Ontologia	1	2	1- Cotiledonar P-value	1- Cotiledonar No de Genes	2- Maturação P-value	2- Maturação No de Genes
Processos reprodutivos celulares			3e-54	57	7.1e-70	74
Fotossíntese			8.8e-34	102	1.2e-49	156
Fotossíntese/ captação de luz			3.5e-19	31	2.3e-18	37
Fotossíntese/ Fotossistema I			2.9e-13	16	1.5e-10	17
Biogênese de oleossomos			8.4e-08	10	1.1e-08	13

C)

Ontologia	1	2	1- Cotiledonar P-value	1- Cotiledonar No de Genes	2- Maturação P-value	2- Maturação No de Genes
Metabolismo de Carboidratos			2.1e-06	68	2.7e-32	89
Metabolismo de lipídeos			0.0024	98	9.2e-79	102
Reserva de nutrientes			---	---	6.2e-54	154
Biogênese de macromoléculas			8.6e-06	76	2.2e-17	86

Figura 8. Análise de enriquecimento de ontologia gênica (PAGE – *Parametric Analysis of Gene Set Enrichment*) dos principais grupos de genes alvos diretos induzidos por LEC1 nos estágios cotiledonar e de início da maturação. A) Processos metabólicos enriquecidos em genes alvos diretos induzidos por LEC1 durante o estágio cotiledonar quando comparado ao início da maturação. **B)** Processos metabólicos enriquecidos em genes alvos diretos induzidos por LEC1 em ambos os estágios cotiledonar e de início da maturação. **C)** Processos metabólicos enriquecidos em genes alvos diretos induzidos por LEC1 durante o início da maturação quando comparado ao estágio cotiledonar. As cores indicam enriquecimento forte (vermelho), fraco (amarelo) ou ausente (cinza) na comparação entre os estágios analisados.

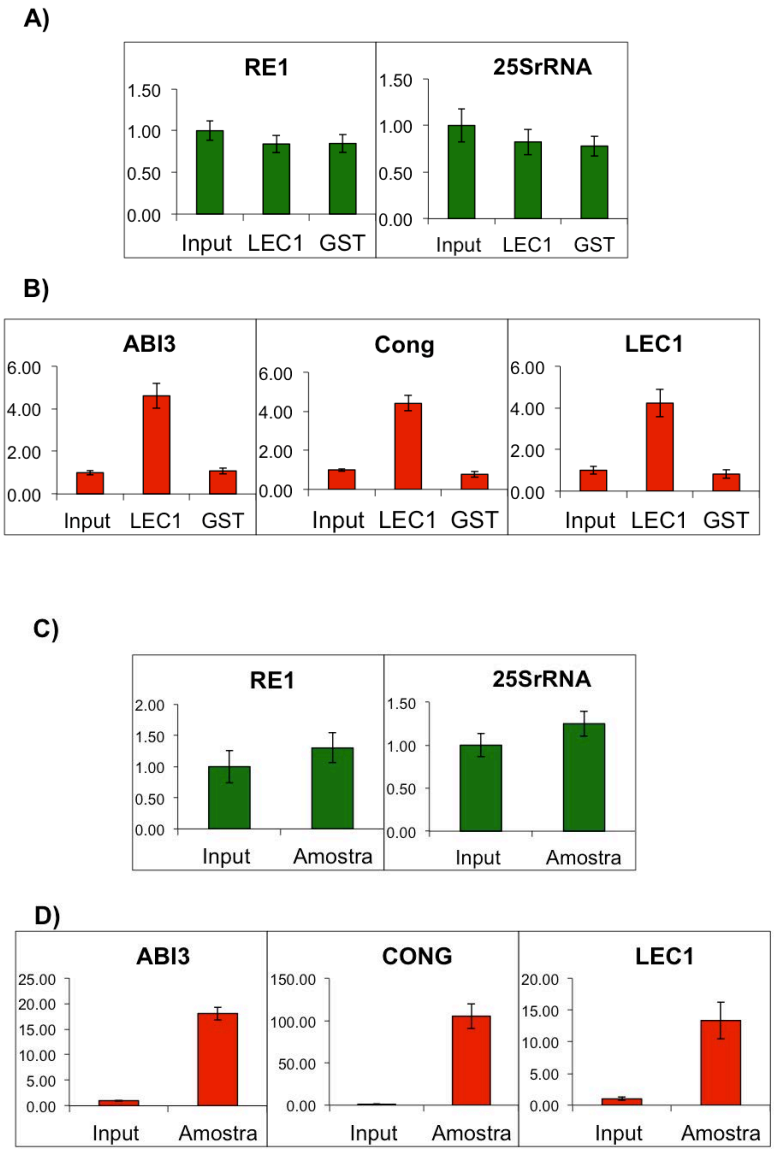


Figura 9. CHIP qPCR de alvos diretos de LEC1 em ambos estágios cotiledonar e de início da maturação. A) Expressão de genes não alvos de LEC1 (controles negativos) de ChIP no estágio cotiledonar. **B)** ChIP qPCR de ABI3, Conglicinina (CONG) e LEC1 durante o estágio cotiledonar **C)** Expressão de genes não alvos de LEC1 (controles negativos) de ChIP no estágio de início da maturação. **D)** ChIP qPCR de ABI3, β-conglicinina (CONG) e LEC1 durante o estágio de início da maturação. Como controle negativo para ChIP utilizamos anticorpo anti GST (**A** e **B).** Valores plotados representam a média de duas replicatas biológicas tanto para o Input (cromatina não imunoprecipitada) quanto para as amostras obtidas pela imunoprecipitação com o anticorpo anti LEC1 e/ou anti GST.

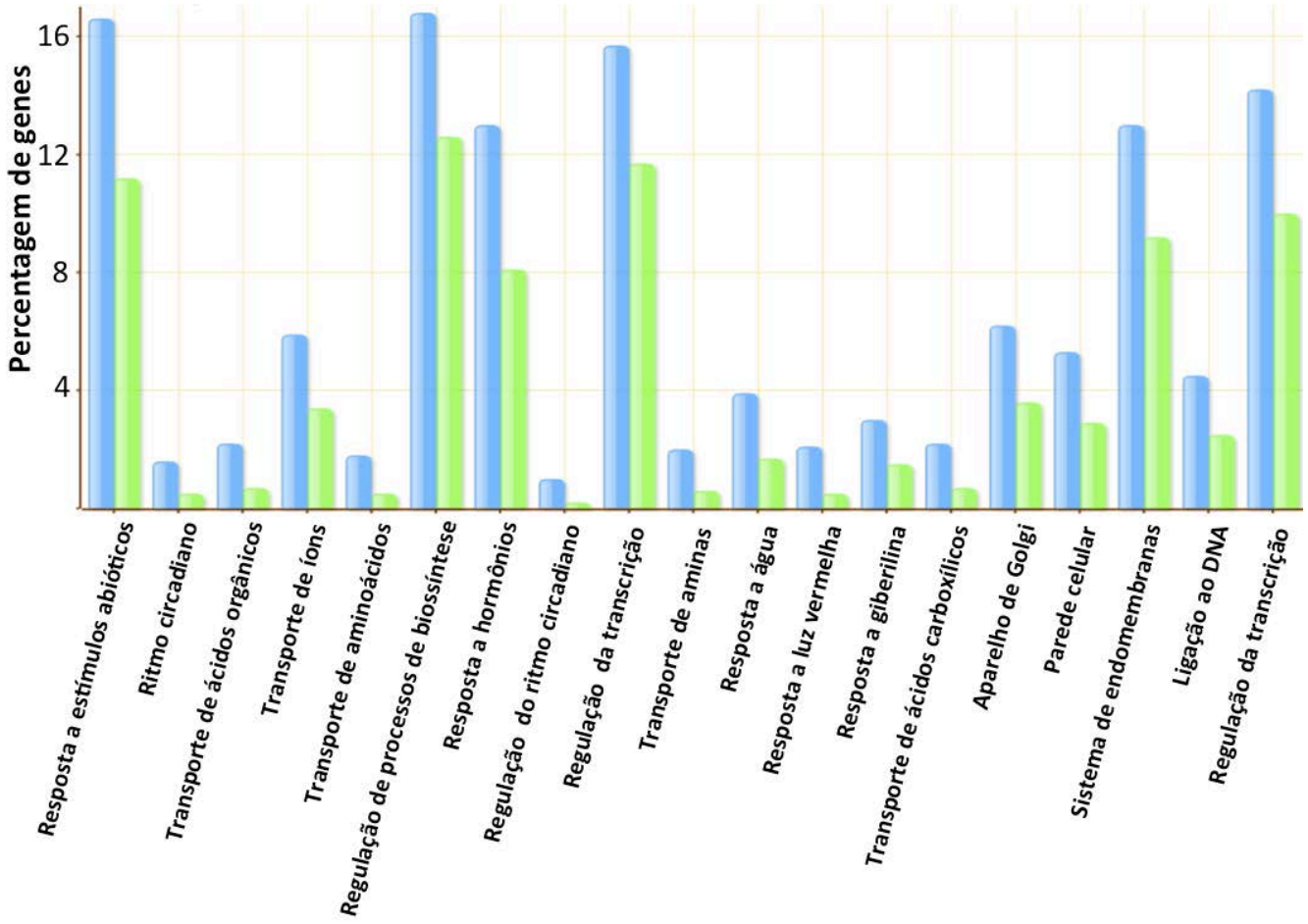


Figura 10. Análise ontológica dos genes alvos diretos reprimidos em ambos os estágios cotiledonar e de início da maturação. As barras azuis representam a percentagem de genes alvo diretos reprimidos por LEC1 e envolvidos nos diferentes processos metabólicos (estatisticamente significativos, ver material e métodos) em ambos os estágios cotiledonar e de início da maturação. As barras verdes representam a percentagem de genes descritos em soja para cada processo metabólico quando comparado ao número total de genes do genoma da espécie (*background*). A comparação entre barras azuis e verdes permite a visualização do enriquecimento de genes alvo de LEC1 em determinada categoria metabólica.

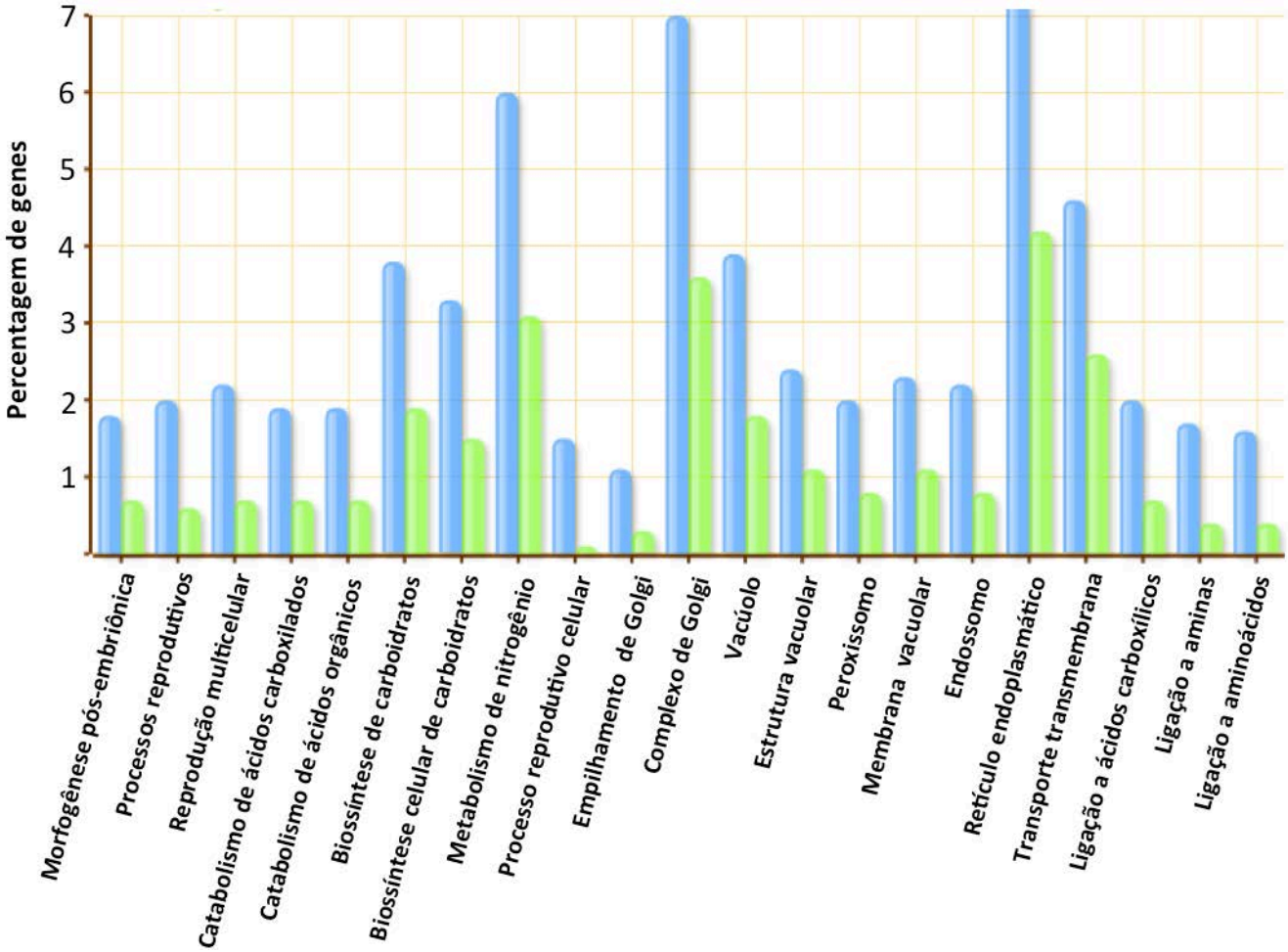


Figura 11. Análise ontológica dos genes alvos diretos reprimidos no estágio cotiledonar. As barras azuis representam a percentagem de genes alvo diretos reprimidos por LEC1 e envolvidos nos diferentes processos metabólicos (estatisticamente significativos, ver material e métodos) durante o estágio cotiledonar. As barras verdes representam a percentagem de genes descritos em soja para cada processo metabólico quando comparado ao número total de genes do genoma da espécie (*background*). A comparação entre barras azuis e verdes permite a visualização do enriquecimento de genes alvos de LEC1 em determinada categoria metabólica.

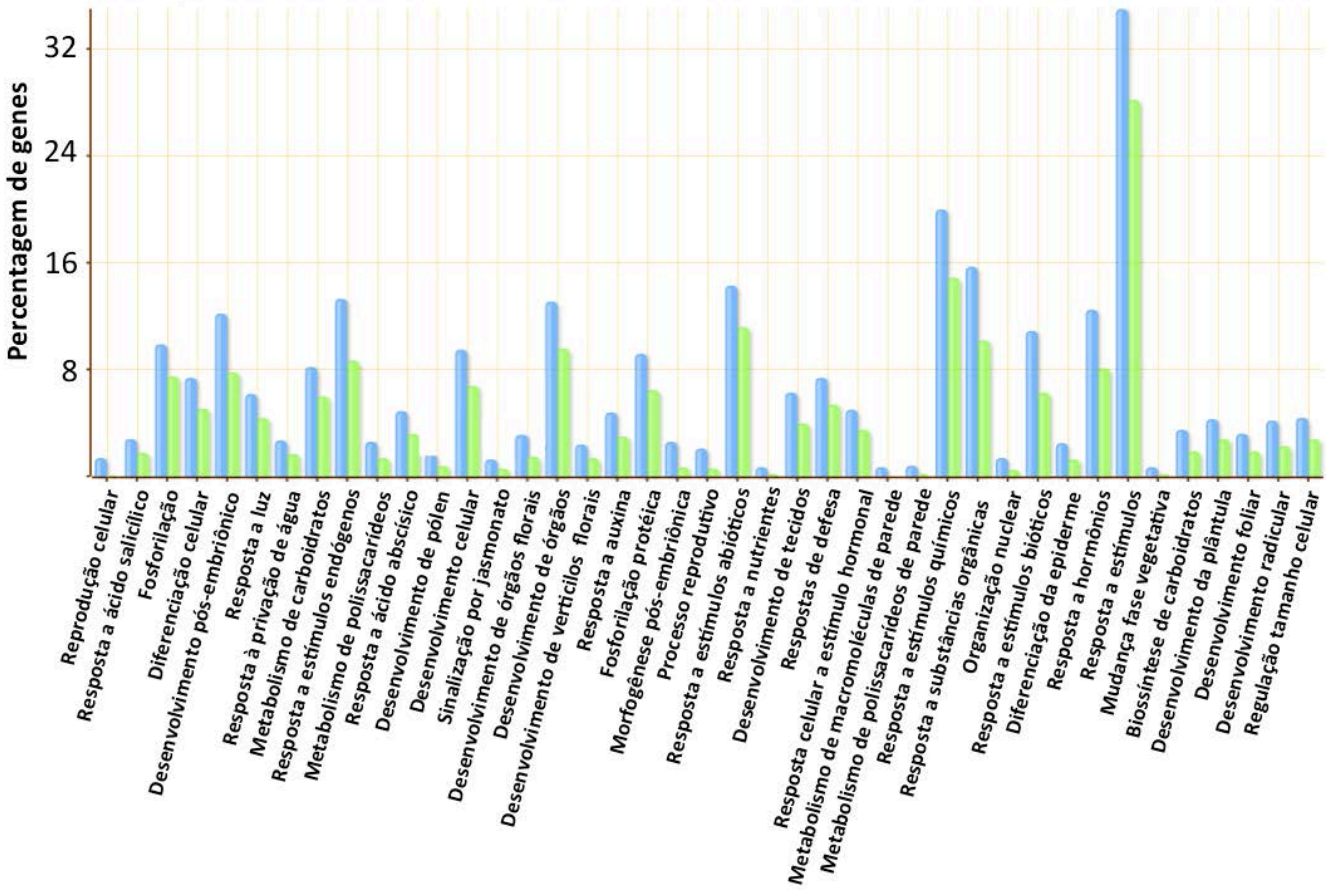
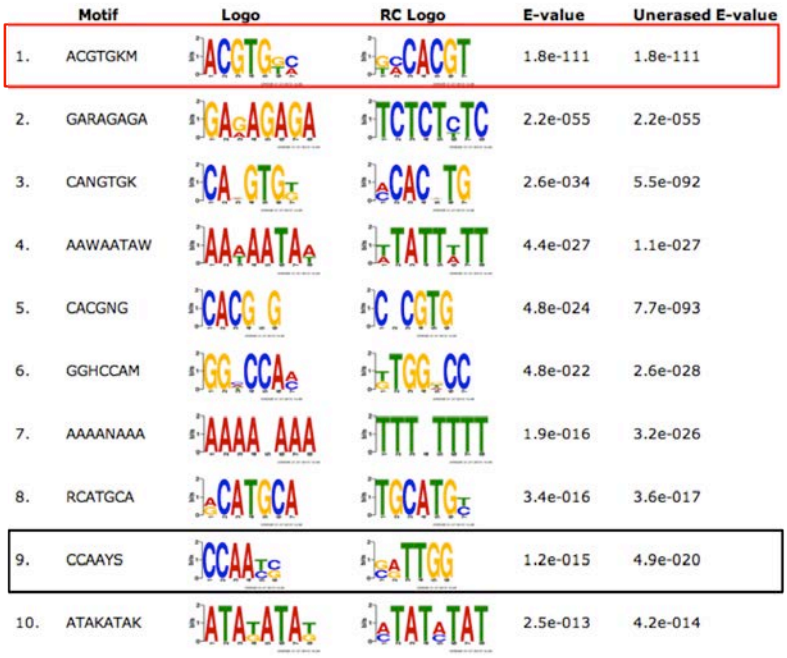


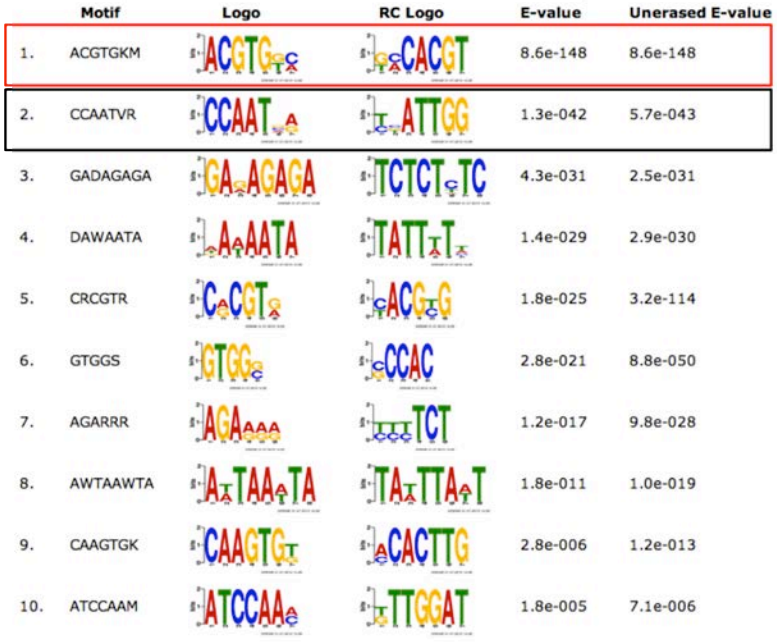
Figura 12. Análise ontológica dos genes alvos diretos reprimidos durante o estágio de início da maturação. As barras azuis representam a percentagem de genes alvo diretos reprimidos por LEC1 e envolvidos nos diferentes processos metabólicos (estatisticamente significativos, ver material e métodos) durante o início da maturação. As barras verdes representam a percentagem de genes descritos em soja para cada processo metabólico quando comparado ao número total de genes do genoma da espécie (*background*). A comparação entre barras azuis e verdes permite a visualização do enriquecimento de genes alvos de LEC1 em determinada categoria metabólica.

A)



	Motif	Logo	RC Logo	E-value	Unerased E-value
1.	ACGTGKM			1.8e-111	1.8e-111
2.	GARAGAGA			2.2e-055	2.2e-055
3.	CANGTGK			2.6e-034	5.5e-092
4.	AAWAATAW			4.4e-027	1.1e-027
5.	CACGNG			4.8e-024	7.7e-093
6.	GGHCCAM			4.8e-022	2.6e-028
7.	AAAANAAA			1.9e-016	3.2e-026
8.	RCATGCA			3.4e-016	3.6e-017
9.	CCAAYS			1.2e-015	4.9e-020
10.	ATAKATAK			2.5e-013	4.2e-014

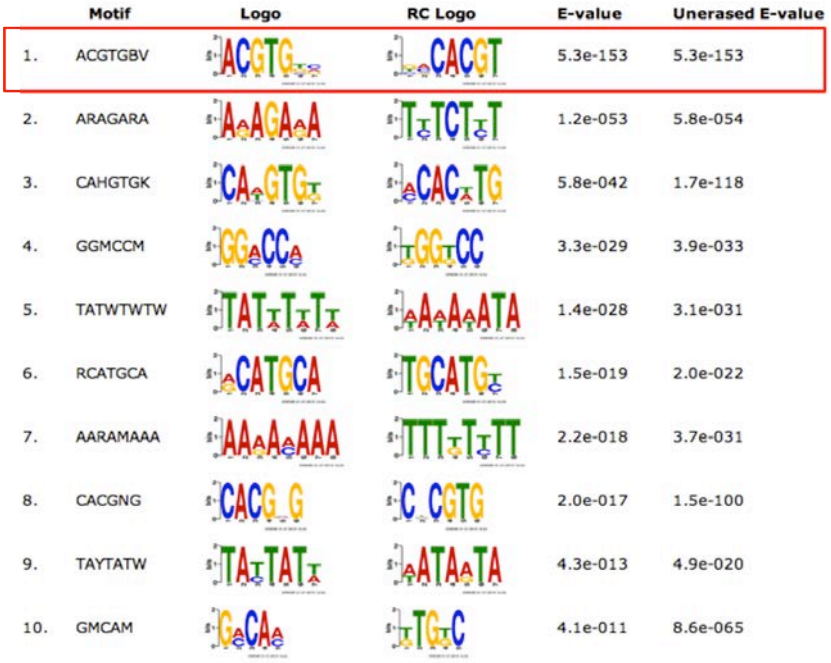
B)



	Motif	Logo	RC Logo	E-value	Unerased E-value
1.	ACGTGKM			8.6e-148	8.6e-148
2.	CCAATVR			1.3e-042	5.7e-043
3.	GADAGAGA			4.3e-031	2.5e-031
4.	DAWAATA			1.4e-029	2.9e-030
5.	CRCGTR			1.8e-025	3.2e-114
6.	GTGGS			2.8e-021	8.8e-050
7.	AGARRR			1.2e-017	9.8e-028
8.	AWTAAWTA			1.8e-011	1.0e-019
9.	CAAGTGK			2.8e-006	1.2e-013
10.	ATCCAAM			1.8e-005	7.1e-006

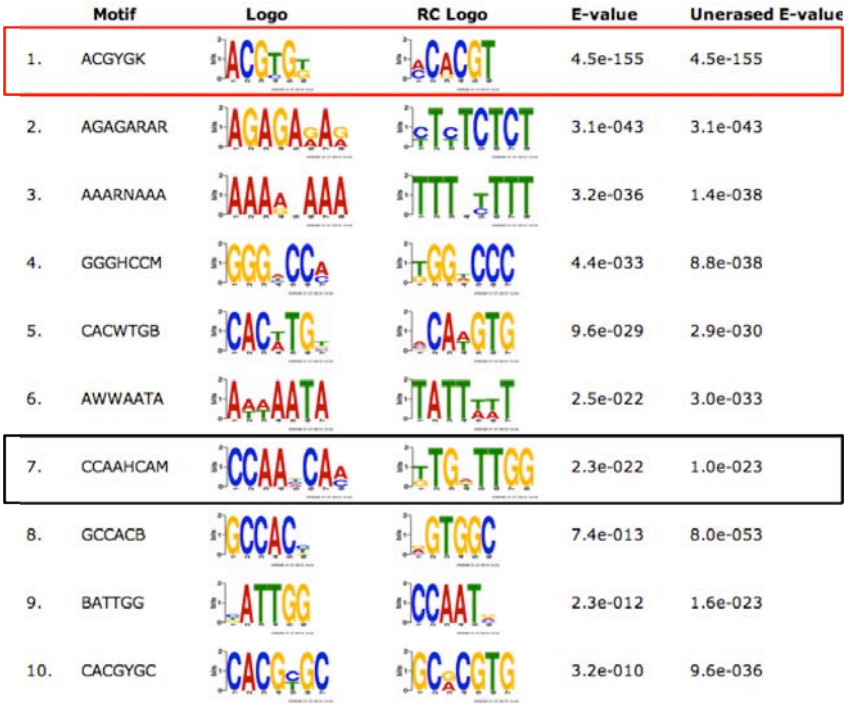
Figura 13. Identificação de motivos conservados nas sequencias de ligação de LEC1 nos genes alvos diretos induzidos **(A)** e reprimidos **(B)** durante o estágio cotiledonar. Retângulos vermelhos e pretos indicam motivo de ligação de fatores de transcrição do tipo bZIP e do tipo CCAAT (motivo canônico de ligação de LEC1), respectivamente.

A)



	Motif	Logo	RC Logo	E-value	Unerased E-value
1.	ACGTGBV			5.3e-153	5.3e-153
2.	ARAGARA			1.2e-053	5.8e-054
3.	CAHGTGK			5.8e-042	1.7e-118
4.	GGMCCM			3.3e-029	3.9e-033
5.	TATWTWTW			1.4e-028	3.1e-031
6.	RCATGCA			1.5e-019	2.0e-022
7.	AARAMAAA			2.2e-018	3.7e-031
8.	CACGNG			2.0e-017	1.5e-100
9.	TAYTATW			4.3e-013	4.9e-020
10.	GMCAM			4.1e-011	8.6e-065

B)



	Motif	Logo	RC Logo	E-value	Unerased E-value
1.	ACGYGK			4.5e-155	4.5e-155
2.	AGAGARAR			3.1e-043	3.1e-043
3.	AAARNAAA			3.2e-036	1.4e-038
4.	GGGHCCM			4.4e-033	8.8e-038
5.	CACWTGB			9.6e-029	2.9e-030
6.	AWWAATA			2.5e-022	3.0e-033
7.	CCAAHCAM			2.3e-022	1.0e-023
8.	GCCACB			7.4e-013	8.0e-053
9.	BATTGG			2.3e-012	1.6e-023
10.	CACGYGC			3.2e-010	9.6e-036

Figura 14. Identificação de motivos conservados nas sequencias de ligação de LEC1 nos genes alvos diretos induzidos **(A)** e reprimidos **(B)** durante o estágio de início da maturação. Retângulos vermelhos e pretos indicam motivo de ligação de fatores de transcrição do tipo bZIP e do tipo CCAAT (motivo canônico de ligação de LEC1), respectivamente.

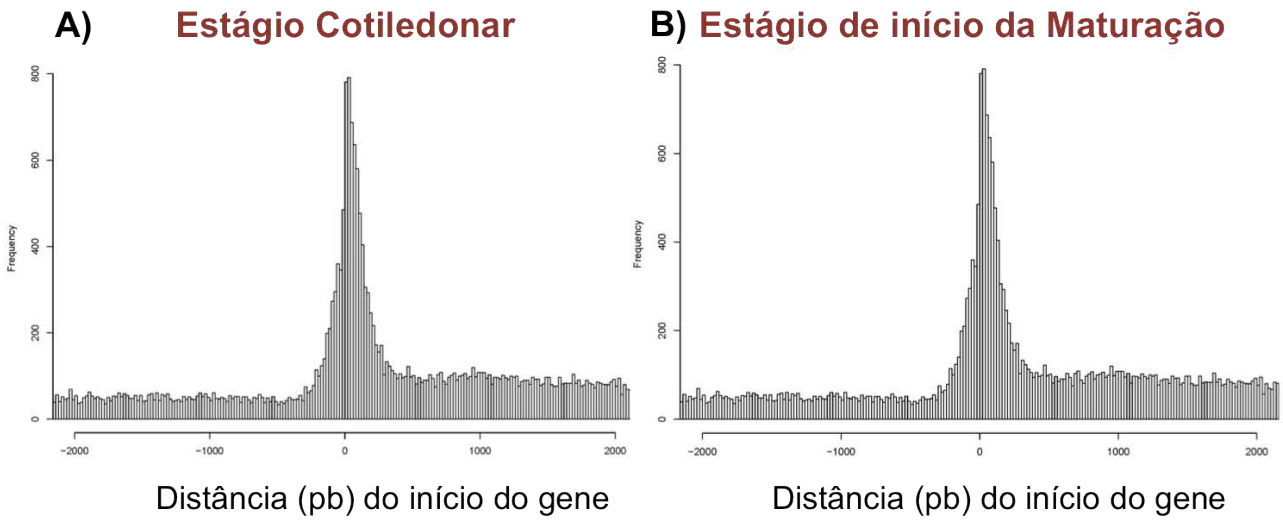


Figura S1. Distribuição dos sítios de ligação de LEC1 em seus genes alvos identificados por ChIP-seq. A) Estágio cotiledonar. **B)** Estágio de início da maturação.

Tabela S1. Oligonucleotídeos iniciadores utilizados em experimentos de ChIP-qPCR.

GENE	LOCUS	SEQUÊNCIA 5' 3'
ABI3	Glyma18g38490	For: AAATCTAACTTCCCAGAACAAACCAGAACCTACAC Rev: AGGAGCAACACACCACCACTACTACTACAAG
LEC1	Glyma07g39820	For: AGCCTCCACTTTTCTTGTCAAATCC Rev: GCCTTAGAATAGGGAAACGGATCAATC
RE1	Glyma17g04601	For: CAATGGGCCTGTGATCTTCT Rev: GAAGGCCACCTTGTTCTTCA
25SrRNA	Glyma13g11810	For: GATGTCGGCTCTTCCTATCATTGTG Rev: CTTAGAGGCGTTCAGTCATAATCCAG
β conglicinina	Glyma10g39150	For: AAAGCCATGCACAACAACACGTACTCACAAAGG Rev: GGCGGAAGAGAATACGAGTTTGAGGTTGGGTTAG

REFERÊNCIAS

Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada CM, Munoz MD, Le BH, Drews GN, Brady SM, Goldberg RB, Harada JJ. 2013. Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc Natl Acad Sci U S A* **110**, E435-44.

Braybrook SA, Harada JJ. 2008. LECs go crazy in embryo development. *Trends Plant Sci* **13**, 624-30.

Calvenzani V, Testoni B, Gusmaroli G, Lorenzo M, Gnesutta N, Petroni K, Mantovani R, Tonelli C. 2012. Interactions and CCAAT-binding of Arabidopsis thaliana NF-Y subunits. *Plos One* **7**, e42902.

Dorn A, Bollekens J, Staub A, Benoist C, Mathis D. 1987. A multiplicity of CCAAT box-binding proteins. *Cell* **50**, 863-72.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z. 2010. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Research* **38**, W64-70.

Duranti M. 2006. Grain legume proteins and nutraceutical properties. *Fitoterapia* **77**, 67-82.

Finkelstein RR, Gampala SS, Rock CD. 2002. Abscisic acid signaling in seeds and seedlings. *Plant Cell* **14 Suppl**, S15-45.

Gibson SI. 2004. Sugar and phytohormone response pathways: navigating a signalling network. *J Exp Bot* **55**, 253-64.

Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, Goodman HM. 1992. Isolation of the Arabidopsis ABI3 gene by positional cloning. *Plant Cell* **4**, 1251-61.

Gusmaroli G, Tonelli C, Mantovani R. 2002. Regulation of novel members of the Arabidopsis thaliana CCAAT-binding nuclear factor Y subunits. *Gene* **283**, 41-8.

Gutierrez L, Van Wuytswinkel O, Castelain M, Bellini C. 2007. Combined networks regulating seed maturation. *Trends Plant Sci* **12**, 294-300.

Izawa T, Foster R, Chua NH. 1993. Plant bZIP protein DNA binding specificity. *J Mol Biol* **230**, 1131-44.

Jeong CW, Roh H, Dang TV, Choi YD, Fischer RL, Lee JS, Choi Y. 2011. An E3 ligase complex regulates SET-domain polycomb group protein activity in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **108**, 8036-41.

Jones SI, Vodkin LO. 2013. Using RNA-Seq to profile soybean seed development from fertilization to maturity. *Plos One* **8**, e59270.

Kumimoto RW, Zhang Y, Siefers N, Holt BF, 3rd. 2010. NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in Arabidopsis thaliana. *Plant J*.

Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, Harada JJ. 2003. LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell* **15**, 5-18.

Le BH, Wagmaister JA, Kawashima T, Bui AQ, Harada JJ, Goldberg RB. 2007. Using genomics to study legume seed development. *Plant Physiol* **144**, 562-74.

Lee H, Fischer RL, Goldberg RB, Harada JJ. 2003. Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. *Proc Natl Acad Sci U S A* **100**, 2152-6.

Libault M, Thibivilliers S, Bilgin DD, O. R, Benitez M, Clough SJ, Stacey G. 2008. Identification of four soybean reference genes for gene expression normalization. *PLant Genome* **1**, 44-54.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-8.

Lotan T, Ohto M, Yee KM, West MA, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ. 1998. Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* **93**, 1195-205.

Meinke DW, Franzmann LH, Nickle TC, Yeung EC. 1994. Leafy Cotyledon Mutants of Arabidopsis. *Plant Cell* **6**, 1049-64.

Mu J, Tan H, Zheng Q, Fu F, Liang Y, Zhang J, Yang X, Wang T, Chong K, Wang XJ, Zuo J. 2008. LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in Arabidopsis. *Plant Physiol* **148**, 1042-54.
Parcy F, Valon C, Kohara A, Misera S, Giraudat J. 1997. The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in concert to control multiple aspects of Arabidopsis seed development. *Plant Cell* **9**, 1265-77.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-40.
Sanjaya, Durrett TP, Weise SE, Benning C. 2011. Increasing the energy density of vegetative tissues by diverting carbon from starch to oil biosynthesis in transgenic Arabidopsis. *Plant Biotechnol J* **9**, 874-83.
Schmittgen TD, Livak KJ. 2008. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* **3**, 1101-8.
Sha AH, Li C, Yan XH, Shan ZH, Zhou XA, Jiang ML, Mao H, Chen B, Wan X, Wei WH. 2012. Large-scale sequencing of normalized full-length cDNA library of soybean seed at different developmental stages and analysis of the gene expression profiles based on ESTs. *Mol Biol Rep* **39**, 2867-74.
Shen B, Allen WB, Zheng P, Li C, Glassman K, Ranch J, Nubel D, Tarczynski MC. 2010. Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize. *Plant Physiol* **153**, 980-7.
Sinha S, Maity SN, Lu J, de Crombrugghe B. 1995. Recombinant rat CBF-C, the third subunit of CBF/NFY, allows formation of a protein-DNA complex with CBF-A and CBF-B and with yeast HAP2 and HAP3. *Proc Natl Acad Sci U S A* **92**, 1624-8.
Stephenson TJ, McIntyre CL, Collet C, Xue GP. 2007. Genome-wide identification and expression analysis of the NF-Y family of transcription factors in Triticum aestivum. *Plant Mol Biol* **65**, 77-92.
Stone SL, Braybrook SA, Paula SL, Kwong LW, Meuser J, Pelletier J, Hsieh TF, Fischer RL, Goldberg RB, Harada JJ. 2008. Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc Natl Acad Sci U S A* **105**, 3151-6.
Tan H, Yang X, Zhang F, Zheng X, Qu C, Mu J, Fu F, Li J, Guan R, Zhang H, Wang G, Zuo J. 2011. Enhanced seed oil production in canola by conditional expression of Brassica napus LEAFY COTYLEDON1 and LEC1-LIKE in developing seeds. *Plant Physiol* **156**, 1577-88.
Truax AD, Greer SF. 2012. ChIP and Re-ChIP assays: investigating interactions between regulatory proteins, histone modifications, and the DNA sequences to which they bind. *Methods Mol Biol* **809**, 175-88.
Watanabe Y, Yamamoto S. 2009. Hōjinzei kihon tsūtatsu no gimonten. Tōkyō: Gyōsei.
West M, Harada JJ. 1993. Embryogenesis in Higher Plants: An Overview. *Plant Cell* **5**, 1361-9.
Wobus U, Weber H. 1999. Seed maturation: genetic programmes and control signals. *Curr Opin Plant Biol* **2**, 33-8.
Yamamoto A, Kagaya Y, Toyoshima R, Kagaya M, Takeda S, Hattori T. 2009. Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. *Plant J* **58**, 843-56.
Yazawa K, Takahata K, Kamada H. 2004. Isolation of the gene encoding Carrot leafy cotyledon1 and expression analysis during somatic and zygotic embryogenesis. *Plant Physiol Biochem* **42**, 215-23.
Yu Y, Li Y, Huang G, Meng Z, Zhang D, Wei J, Yan K, Zheng C, Zhang L. 2011. PwHAP5, a CCAAT-binding transcription factor, interacts with PwFKBP12 and plays a role in pollen tube growth orientation in Picea wilsonii. *J Exp Bot* **62**, 4805-17.

5 CONSIDERAÇÕES FINAIS

No presente trabalho buscamos, primeiramente contribuir para um melhor entendimento do processo de síntese de ácidos graxos e triacilgliceróis (TAGs) em plantas e algas, por meio de um estudo comparativo das diferentes rotas envolvidas na produção e acúmulo de lipídeos de reserva nestes grupos de organismos.

Como resultado, no primeiro capítulo concluímos que as principais rotas de biossíntese de ácidos graxos e TAGs em algas são análogas às apresentadas em plantas, baseado em homologia de sequência e características bioquímicas de muitos dos genes e enzimas envolvidas em diferentes etapas de síntese e acúmulo de TAGs (Hu *et al.*, 2008). Tanto em plantas quanto em algas é possível observar a existência de representantes capazes de sintetizar ácidos graxos não usuais, com diferentes comprimentos de cadeia carbônica e/ou apresentando modificações na composição de sua cadeia tais como duplas ligações em posições não comuns, ou a presença de grupos funcionais, todas estas, características importantes do ponto de vista econômico. Entretanto, diferentemente de plantas onde classes de lipídeos podem ser sintetizadas e localizadas em um local específico (semente, por exemplo), toda as etapas de síntese e acúmulo de TAGs ocorrem em uma única célula, que constitui a alga.

Em condições normais de crescimento as algas sintetizam principalmente lipídeos de membrana. Já sob condições desfavoráveis de crescimento ou de estresse, ocorre a mudança de uma síntese de lipídeos de membrana para síntese de TAGs (Tonon *et al.*, 2002). Ou seja, a biossíntese de TAGs em algas é uma via ativa na resposta a estresses, desempenhando importante papel como fonte de energia sob condições estressantes.

Ao contrário de plantas, as algas são capazes de sintetizar quantidades significativas de ácidos graxos poli-insaturados de cadeia longa (VLC-PUFAs), graças à ação de enzimas desaturases e elongases exclusivas. Inclusive, a excelente capacidade de adaptação a diferentes condições ambientais observada em algas é atribuída à sua capacidade de sintetizar um grande número de ácidos graxos não usuais (como VLC-PUFAs, por exemplo) e de modificar de maneira eficiente o seu metabolismo lipídico em resposta a condições de estresse, promovidos por mudanças ambientais.

Em termos de produção de óleo, as microalgas são capazes de produzir uma maior quantidade de óleo em uma mesma quantidade de área quando comparado à plantas oleaginosas (Huerlimann *et al.*, 2010) e, por este dentre outros motivos, vem atraindo muita atenção e incentivos buscando a identificação e caracterização de algas produtoras de óleos com características interessantes, sobretudo do ponto de vista industrial.

A diversidade natural observada em TAGs acumulados tanto em sementes quanto em algas, constitui uma importante fonte com potencial biotecnológico para a manipulação genética com o objetivo de produzir ácidos graxos não usuais em espécies vegetais agronomicamente importantes e em algas, sobretudo visando a substituição de nosso atual uso do petróleo por fontes alternativas e renováveis de óleo.

Nossas análises, utilizando sequências de todos os genomas eucarióticos disponíveis até o momento, mostraram que fungos e a grande maioria dos metazoários possuem um único gene codificando para cada subunidade (NF-YA, NF-YB e NF-YC) do fato de transcrição NF-Y, o qual apresenta alguns membros envolvidos com a regulação de processos como maturação, síntese e acúmulo de lipídeos em sementes. A exceção a essa regra ficou por conta do anfioxo *Branchiostoma floridae,* do nematóide *Caenorhabditis elegans* e do gastrópode *Lottia gigantean*, que apresentam dois genes codificando cada uma das três subunidades. Em plantas, cada subunidade é codificada por famílias gênicas, com exceção do musgo *Physcomitrella patens* e da licófita *Selaginella mollendorffii* que apresentam apenas um gene codificando para a subunidade NF-YA.

Análises filogenéticas mostraram que a diversificação gênica observada em todas as subunidades do fator de transcrição NF-Y é resultado de diversos eventos de duplicação ao longo da evolução e diversificação das plantas. Nós observamos também a formação de grupos independentes, sugerindo que cada grupo possa ter um gene ancestral independente a partir do qual os genes duplicados se originaram. Entretanto, em muitas espécies também foram observados eventos de duplicação gênica que ocorreram após a divergência entre mono e dicotiledôneas, especialmente em espécies que passaram por eventos recentes de duplicação de seus genomas (poliploidização).

Quando comparado com outras formas de mutação, uma importante característica da duplicação gênica é que ela cria redundância genética. Nesse contexto, a duplicação gênica contribui para a evolução fornecendo matéria-prima e criando oportunidades para que os genes duplicados evoluam para a aquisição de novas funções ou novos parceiros de interação que, por sua vez, podem contribuir para a plasticidade do genoma ou da espécie em se adaptar a novas funções metabólicas, mudanças ambientais ou condições de estresse. Um exemplo de aquisição de novas funções após duplicação para membros da subunidade NF-YB foi discutido no terceiro capítulo, que trata sobre a origem e evolução dos genes do tipo LEC1 em plantas vasculares a partir de um gene do tipo não LEC1 duplicado em plantas. O resultados apresentados relativos ao estudo de genômica comparativa, usando análise filogenética associada à dados de expressão, permitiu validar o modelo para a origem de genes do tipo LEC1 e fornecer novas informações sobre as relações evolutivas entre genes do tipo LEC1 e do tipo não LEC1 durante a evolução das plantas vasculares. Foram identificados genes do tipo LEC1 em todos os organismos vasculares, ao contrário de amoebozoa, algas, fungos, metazoários e plantas não vasculares, que apresentam exclusivamente genes do tipo não LEC1 como constituintes de sua subunidade NF-YB.

As taxas de substituição sinônima e não sinônimas (Ka/Ks) entre genes do tipo LEC1 e genes do tipo não LEC1 indicaram a presença de seleção positiva atuando nos membros do tipo LEC1 para a fixação de aminoácidos tipo LEC1 específicos.

As análises filogenéticas demonstraram que os genes do tipo LEC1 são evolutivamente divergentes dos genes do tipo não LEC1 dos outros organismos analisados. Nossos resultados apontam para um cenário no qual os genes do tipo LEC1 teriam se originado em plantas vasculares após a expansão gênica observada na subunidade NF-YB em plantas. Nós sugerimos um modelo no qual genes do tipo LEC1 teriam sido originados a partir de um ancestral eucariótico do tipo não LEC1 através de processos de neofuncionalização e/ou subfuncionalização. Tais processos seriam os responsáveis pela surgimento das novas funções apresentadas pelos genes do tipo LEC1 em plantas vasculares, especialmente em plantas com sementes.

Enquanto a expressão dos genes do tipo LEC1 parece ser restrita a sementes (ou estruturas que contém sementes, como inflorescências, por exemplo), a expressão dos genes do tipo não LEC1 foi detectada em diferentes órgãos e tecidos,

indicando que os genes do tipo LEC1, além de serem filogeneticamente divergentes, também apresentam perfil de expressão diferente (exclusivo em sementes) quando comparado a genes do tipo não LEC1.

Do ponto de vista evolutivo, o surgimento dos genes do tipo LEC1 poderia ter conferido vantagens adaptativas às plantas vasculares que podem ter sido importantes para a ampla distribuição desses organismos em diferentes habitats e condições ambientais. A origem dos genes do tipo LEC1 em plantas vasculares e sua posterior neofuncionalização e/ou subfuncionalização representam um importante passo no controle do processo de amadurecimento e maturação da semente e, consequentemente, podem ter contribuído com o sucesso reprodutivo apresentado pelas plantas com sementes.

A mamona (*Ricinus communis* L.) é uma espécie que vem surgindo como um importante modelo de estudo em plantas oleaginosas. Análises de expressão demonstraram uma sobreposição parcial no perfil de expressão entre LEC1, L1L com os membros da subunidade NF-YC ao longo do desenvolvimento da semente de mamona, especialmente nos estágios durante o qual o embrião e o endosperma encontram-se totalmente expandidos e onde previamente observamos a expressão máxima de genes que codificam enzimas envolvidas com a biossíntese de TAGs em mamona (Cagliari *et al.*, 2010). Tal sobreposição de expressão está de acordo com o fato de que algumas subunidades do fator de transcrição NF-Y estão diretamente envolvidas com o processo de síntese e acúmulo de óleo em sementes (Shen *et al.*, 2010; Tan *et al.*, 2011). Entretanto, estudos adicionais serão particularmente interessantes e contribuirão para a elucidação dos papéis desempenhados por LEC1, L1L e pelos membros da subunidade NF-YC no controle da síntese e acúmulo de lipídeos em sementes de mamona.

Análises de interação proteína-proteína demonstraram a existência de diferentes parceiros NF-YC para LEC1 e L1L, indicando que os genes do tipo LEC1 em mamona não são redundantes e podem ter evoluído para desempenhar diferentes papéis durante o desenvolvimento da semente de mamona. Nossos resultados sugerem que a divergência funcional, resultante de diferentes parceiros NF-YC, poderia permitir o controle de diferentes grupos de genes através da interação específica com distintos parceiros NF-YC. Estudos adicionais são importantes para

determinar o papel bem como os genes alvo dos complexos formados por LEC1, L1L e membros de NF-YC ao longo do desenvolvimento da semente de mamona.

Como perspectivas para o capítulo quatro, estamos atualmente iniciando ensaios de transativação em protoplastos de tabaco (ensaio GUS/Luciferase) com o objetivo de testar a habilidade dos dímeros NF-YB/NF-YC em ativar a expressão de promotores de proteínas de reserva em mamona. Para tanto, selecionamos a sequencia promotora de sete genes cuja expressão se sobrepõe à expressão dos genes que codificam os dímeros analisados. Com os ensaios de transativação, pretendemos mostrar dados quanto à funcionalidade dos dímeros NF-YB/NF-YC observados nos experimentos de interação proteína-proteína.

Por fim, identificamos os genes alvo diretos de LEC1 em diferentes estágios do desenvolvimento da semente de soja: estágio cotiledonar e início da maturação. Os resultados demonstraram que, apesar de ambos os estágios apresentarem muitos genes alvo diretos induzidos e reprimidos compartilhados, também apresentam características distintas quanto ao genes alvo diretos induzidos por LEC1: enquanto no estágio cotiledonar os genes alvo diretos são representados por genes envolvidos em ciclo celular, remodelação da cromatina e ativação transcricional, durante o início da maturação observamos um enriquecimento em processos como metabolismo e acúmulo de substâncias de reserva.

Além disso, nossas análises por CHiP-seq demonstraram que LEC1 se liga a promotores enriquecidos na sequencia canônica CCAAT, como observado em mamíferos, mas também em diversos outros motivos, dentre os quais merece destaque a sequência típica de ligação de fatores de transcrição do tipo bZIP. Em Arabidopsis, foi observado que, enquanto experimentos *in vitro* demonstraram que LEC1 se liga à sequencia CCAAT, experimentos *in planta* (ChIP-seq) apontaram para a preferência de ligação de LEC1 a promotores alvo com motivos para fatores de transcrição do tipo bZIP e não motivos CCAAT (dados não publicados do Dr Harada). Portanto, em plantas LEC1 possui a capacidade de controlar a expressão de genes alvo pela ligação à outros motivos além de CCAAT. Nesse contexto, fatores de transcrição do tipo bZIP podem estar diretamente relacionados com a regulação do desenvolvimento e maturação da semente exercida por LEC1, demonstrando que a

relação entre ABA, fatores de transcrição bZIP e LEC1 parece ser muito importante no controle das fases de desenvolvimento e da maturação de sementes.

Pela análise comparativa da organização dos genes envolvidos no processo de regulação da síntese e acúmulo de lipídeos em sementes, de suas histórias evolutivas, da expressão e da estrutura dos seus produtos gênicos e identificação de seus genes alvo pretende-se contribuir para a elucidação dos mecanismos que controlam o desenvolvimento e acúmulo de macromoléculas, em especial lipídeos, em sementes. Essas análises podem contribuir para a ampliação do conhecimento sobre o controle genético da síntese de ácidos graxos, acúmulo de TAGs e metabolismo e desenvolvimento da semente, de modo a fornecer ferramentas moleculares que possam ser úteis no delineamento de novas estratégias para o melhoramento das plantas de interesse econômico para o Brasil.

REFERÊNCIAS

Abbadi A, Domergue F, Bauer J, Napier JA, Welti R, Zahringer U, Cirpus P, Heinz E. 2004. Biosynthesis of very-long-chain polyunsaturated fatty acids in transgenic oilseeds: constraints on their accumulation. *Plant Cell* **16**, 2734-48.

Agrawal GK, Thelen JJ. 2006. Large scale identification and quantitative profiling of phosphoproteins expressed during seed filling in oilseed rape. *Mol Cell Proteomics* **5**, 2044-59.

Albani D, Hammond-Kosack MC, Smith C, Conlan S, Colot V, Holdsworth M, Bevan MW. 1997. The wheat transcriptional activator SPA: a seed-specific bZIP protein that recognizes the GCN4-like motif in the bifactorial endosperm box of prolamin genes. *Plant Cell* **9**, 171-84.

Angelovici R, Galili G, Fernie AR, Fait A. 2010. Seed desiccation: a bridge between maturation and germination. *Trends Plant Sci* **15**, 211-8.

Arents G, Moudrianakis EN. 1995. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc Natl Acad Sci U S A* **92**, 11170-4.

Bafor M, Smith MA, Jonsson L, Stobart K, Stymne S. 1991. Ricinoleic acid biosynthesis and triacylglycerol assembly in microsomal preparations from developing castor-bean (Ricinus communis) endosperm. *Biochem J* **280 (Pt 2)**, 507-14.

Bate NJ, Niu XP, Wang YW, Reimann KS, Helentjaris TG. 2004. An invertase inhibitor from maize localizes to the embryo surrounding region during early kernel development. *Plant Physiol* **134**, 246-54.

Baud S, Lepiniec L. 2010. Physiological and developmental regulation of seed oil production. *Prog Lipid Res* **49**, 235-49.

Baud S, Mendoza MS, To A, Harscoet E, Lepiniec L, Dubreucq B. 2007. WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. *Plant J* **50**, 825-38.

Baumlein H, Nagy I, Villarroel R, Inze D, Wobus U. 1992. Cis-analysis of a seed protein gene promoter: the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. *Plant J* **2**, 233-9.

Baxevanis AD, Arents G, Moudrianakis EN, Landsman D. 1995. A variety of DNA-binding and multimeric proteins contain the histone fold motif. *Nucleic Acids Res* **23**, 2685-91.

Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada CM, Munoz MD, Le BH, Drews GN, Brady SM, Goldberg RB, Harada JJ. 2013. Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc Natl Acad Sci U S A* **110**, E435-44.

Ben-Naim O, Eshed R, Parnis A, Teper-Bamnolker P, Shalit A, Coupland G, Samach A, Lifschitz E. 2006. The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA. *Plant J* **46**, 462-76.

Bensmihen S, Giraudat J, Parcy F. 2005. Characterization of three homologous basic leucine zipper transcription factors (bZIP) of the ABI5 family during Arabidopsis thaliana embryo maturation. *J Exp Bot* **56**, 597-603.

Bensmihen S, Rippa S, Lambert G, Jublot D, Pautot V, Granier F, Giraudat J, Parcy F. 2002. The homologous ABI5 and EEL transcription factors function antagonistically to fine-tune gene expression during late embryogenesis. *Plant Cell* **14**, 1391-403.

Bewley JD. 1997. Seed Germination and Dormancy. *Plant Cell* **9**, 1055-66.

Blackwell TK, Bowerman B, Priess JR, Weintraub H. 1994. Formation of a monomeric DNA binding domain by Skn-1 bZIP and homeodomain elements. *Science* **266**, 621-8.

Braybrook SA, Harada JJ. 2008. LECs go crazy in embryo development. *Trends Plant Sci* **13**, 624-30.

Braybrook SA, Stone SL, Park S, Bui AQ, Le BH, Fischer RL, Goldberg RB, Harada JJ. 2006. Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. *Proc Natl Acad Sci U S A* **103**, 3468-73.

Buitink J, Leger JJ, Guisle I, Vu BL, Wuilleme S, Lamirault G, Le Bars A, Le Meur N, Becker A, Kuster H, Leprince O. 2006. Transcriptome profiling uncovers metabolic and regulatory

processes occurring during the transition from desiccation-sensitive to desiccation-tolerant stages in Medicago truncatula seeds. *Plant J* **47**, 735-50.
Cadman CS, Toorop PE, Hilhorst HW, Finch-Savage WE. 2006. Gene expression profiles of Arabidopsis Cvi seeds during dormancy cycling indicate a common underlying dormancy control mechanism. *Plant J* **46**, 805-22.
Cagliari A, Margis-Pinheiro M, Loss G, Mastroberti AA, Mariath JED, Margis R. 2010. Identification and expression analysis of castor bean (Ricinus communis) genes encoding enzymes from the triacylglycerol biosynthesis pathway. *Plant Science* **179**, 499-509.
Cagliari A, Turchetto-Zolet AC, Maraschin FS, Loss G, Margis R, Margis-Pinheiro M. 2011. The Evolutionary History of CBF Transcription Factors: Gene Duplication of CCAAT – Binding Factors NF-Y in Plants*Gene Duplication*: Intech.
Cahoon EB, Shockey JM, Dietrich CR, Gidda SK, Mullen RT, Dyer JM. 2007. Engineering oilseeds for sustainable production of industrial and nutritional feedstocks: solving bottlenecks in fatty acid flux. *Curr Opin Plant Biol* **10**, 236-44.
Calvenzani V, Testoni B, Gusmaroli G, Lorenzo M, Gnesutta N, Petroni K, Mantovani R, Tonelli C. 2012. Interactions and CCAAT-binding of Arabidopsis thaliana NF-Y subunits. *Plos One* **7**, e42902.
Cao S, Kumimoto RW, Siriwardana CL, Risinger JR, Holt BF, 3rd. 2011. Identification and characterization of NF-Y transcription factor families in the monocot model plant Brachypodium distachyon. *PLoS One* **6**, e21805.
Casson SA, Lindsey K. 2006. The turnip mutant of Arabidopsis reveals that LEAFY COTYLEDON1 expression mediates the effects of auxin and sugars to promote embryonic cell identity. *Plant Physiol* **142**, 526-41.
Caupin H. 1997. Products from castor oil: past, present, and future. *Lipid Technol.Appl.*, 787-95.
Ceribelli M, Dolfini D, Merico D, Gatta R, Vigano AM, Pavesi G, Mantovani R. 2008. The histone-like NF-Y is a bifunctional transcription factor. *Mol Cell Biol* **28**, 2047-58.
Chakrabortee S, Boschetti C, Walton LJ, Sarkar S, Rubinsztein DC, Tunnacliffe A. 2007. Hydrophilic protein associated with desiccation tolerance exhibits broad protein stabilization function. *Proc Natl Acad Sci U S A* **104**, 18073-8.
Chen SL, Dai SX, Li JK, Wang SS, P.; A, Aloys H. 2002. Isolation of protoplast and ion channel recording in plasma membrane of suspension cells of Populus euphratica. *For. Stud. China* **4**, 1-4.
Chiaradia JJ. 2005. Avaliação agronômica e fluxo de gases do efeito estufa a partir de solo tratado com resíduos e cultivado com mamona (Ricinus communis L.) em área de reforma canavial*Escola Superior de Agricultura Luiz de Queiroz.*

. Piracicaba: ESALQ/USP.
Coelho I. 1979. Avaliação das exportações tradicionais baianas: caso de sisal e mamona.
CONAB CNdA. 2013.
De Grassi A, Lanave C, Saccone C. 2008. Genome duplication and gene-family evolution: the case of three OXPHOS gene families. *Gene* **421**, 1-6.
Dorn A, Bollekens J, Staub A, Benoist C, Mathis D. 1987. A multiplicity of CCAAT box-binding proteins. *Cell* **50**, 863-72.
Du Z, Zhou X, Ling Y, Zhang Z, Su Z. 2010. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Research* **38**, W64-70.
Duranti M. 2006. Grain legume proteins and nutraceutical properties. *Fitoterapia* **77**, 67-82.
Durrett TP, Benning C, Ohlrogge J. 2008. Plant triacylglycerols as feedstocks for the production of biofuels. *Plant J* **54**, 593-607.
Dyer JM, Mullen RT. 2008. Engineering plant oils as high-value industrial feedstocks for biorefining: the need for underpinning cell biology research. *Physiol Plant* **132**, 11-22.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-7.
Edwards D, Murray JA, Smith AG. 1998. Multiple genes encoding the conserved CCAAT-box transcription factor complex are expressed in Arabidopsis. *Plant Physiol* **117**, 1015-22.

EMBRAPA. 2004. Informações sobre os Sistemas de Produção Utilizados na Ricinocultura na Região Nordeste, em Especial o Semi-Árido e outros aspectos ligados a sua cadeia. Campina Grande, Paraíba.
Fait A, Angelovici R, Less H, Ohad I, Urbanczyk-Wochniak E, Fernie AR, Galili G. 2006. Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. *Plant Physiol* **142**, 839-54.
FAO FdNUpaAeaA. 2013.
Finch-Savage WE, Leubner-Metzger G. 2006. Seed dormancy and the control of germination. *New Phytol* **171**, 501-23.
Finkelstein RR, Gampala SS, Rock CD. 2002. Abscisic acid signaling in seeds and seedlings. *Plant Cell* **14 Suppl**, S15-45.
Flagel LE, Wendel JF. 2009. Gene duplication and evolutionary novelty in plants. *New Phytol* **183**, 557-64.
Focks N, Benning C. 1998. wrinkled1: A novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. *Plant Physiol* **118**, 91-101.
Frey A, Godin B, Bonnet M, Sotta B, Marion-Poll A. 2004. Maternal synthesis of abscisic acid controls seed development and yield in Nicotiana plumbaginifolia. *Planta* **218**, 958-64.
Frontini M, Imbriano C, Manni I, Mantovani R. 2004. Cell cycle regulation of NF-YC nuclear localization. *Cell Cycle* **3**, 217-22.
Gazzarrini S, Tsuchiya Y, Lumba S, Okamoto M, McCourt P. 2004. The transcription factor FUSCA3 controls developmental timing in Arabidopsis through the hormones gibberellin and abscisic acid. *Dev Cell* **7**, 373-85.
Gibson SI. 2004. Sugar and phytohormone response pathways: navigating a signalling network. *J Exp Bot* **55**, 253-64.
Gibson SI. 2005. Control of plant development and gene expression by sugar signaling. *Curr Opin Plant Biol* **8**, 93-102.
Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, Goodman HM. 1992. Isolation of the Arabidopsis ABI3 gene by positional cloning. *Plant Cell* **4**, 1251-61.
Gusmaroli G, Tonelli C, Mantovani R. 2002. Regulation of novel members of the Arabidopsis thaliana CCAAT-binding nuclear factor Y subunits. *Gene* **283**, 41-8.
Gutierrez L, Van Wuytswinkel O, Castelain M, Bellini C. 2007. Combined networks regulating seed maturation. *Trends Plant Sci* **12**, 294-300.
Hackenberg D, Wu Y, Voigt A, Adams R, Schramm P, Grimm B. 2011. Studies on Differential Nuclear Translocation Mechanism and Assembly of the Three Subunits of the Arabidopsis thaliana Transcription Factor NF-Y. *Mol Plant*.
Halim R, Danquah MK, Webley PA. 2012. Extraction of oil from microalgae for biodiesel production: A review. *Biotechnol Adv* **30**, 709-32.
Hammond-Kosack MC, Holdsworth MJ, Bevan MW. 1993. In vivo footprinting of a low molecular weight glutenin gene (LMWG-1D1) in wheat endosperm. *EMBO J* **12**, 545-54.
Harada JJ. 1997. Cellular and Molecular Biology of Plant Seed Development*Cellular and molecular* Dordrecht: Kluwer Academic, 545-92.
Hartig K, Beck E. 2006. Crosstalk between auxin, cytokinins, and sugars in the plant cell cycle. *Plant Biology* **8**, 389-96.
Hays DB, Wilen RW, Sheng C, Moloney MM, Pharis RP. 1999. Embryo-specific gene expression in microspore-derived embryos of brassica napus. An interaction between abscisic acid and jasmonic acid1,2. *Plant Physiol* **119**, 1065-72.
Hernandez-Sebastia C, Marsolais F, Saravitz C, Israel D, Dewey RE, Huber SC. 2005. Free amino acid profiles suggest a possible role for asparagine in the control of storage-product accumulation in developing seeds of low- and high-protein soybean lines. *J Exp Bot* **56**, 1951-63.
Holdsworth MJ, Bentsink L, Soppe WJ. 2008. Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination. *New Phytol* **179**, 33-54.
Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, Darzins A. 2008. Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *Plant J* **54**, 621-39.

Huang HS, Chen CJ, Chang WC. 1999. The CCAAT-box binding factor NF-Y is required for the expression of phospholipid hydroperoxide glutathione peroxidase in human epidermoid carcinoma A431 cells. *FEBS Lett* **455**, 111-6.
Huerlimann R, de Nys R, Heimann K. 2010. Growth, lipid content, productivity, and fatty acid composition of tropical microalgae for scale-up production. *Biotechnol Bioeng* **107**, 245-57.
Izawa T, Foster R, Chua NH. 1993. Plant bZIP protein DNA binding specificity. *J Mol Biol* **230**, 1131-44.
Jaworski J, Cahoon EB. 2003. Industrial oils from transgenic plants. *Curr Opin Plant Biol* **6**, 178-84.
Jeong CW, Roh H, Dang TV, Choi YD, Fischer RL, Lee JS, Choi Y. 2011. An E3 ligase complex regulates SET-domain polycomb group protein activity in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **108**, 8036-41.
Jones SI, Vodkin LO. 2013. Using RNA-Seq to profile soybean seed development from fertilization to maturity. *Plos One* **8**, e59270.
Kagaya Y, Okuda R, Ban A, Toyoshima R, Tsutsumida K, Usui H, Yamamoto A, Hattori T. 2005a. Indirect ABA-dependent regulation of seed storage protein genes by FUSCA3 transcription factor in Arabidopsis. *Plant Cell Physiol* **46**, 300-11.
Kagaya Y, Toyoshima R, Okuda R, Usui H, Yamamoto A, Hattori T. 2005b. LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. *Plant Cell Physiol* **46**, 399-406.
Keeling CI, Weisshaar S, Lin RP, Bohlmann J. 2008. Functional plasticity of paralogous diterpene synthases involved in conifer defense. *Proc Natl Acad Sci U S A* **105**, 1085-90.
Kim IS, Sinha S, de Crombrugghe B, Maity SN. 1996. Determination of functional domains in the C subunit of the CCAAT-binding factor (CBF) necessary for formation of a CBF-DNA complex: CBF-B interacts simultaneously with both the CBF-A and CBF-C subunits to form a heterotrimeric CBF molecule. *Mol Cell Biol* **16**, 4003-13.
Kouri JS, R F.; Santos, J W. 2004. Evolução da Cultura da Mamona no Brasil.*CONGRESSO BRASILEIRO DE MAMONA*. Campina Grande-PB: EMBRAPA ALGODÃO.
Kumimoto RW, Zhang Y, Siefers N, Holt BF, 3rd. 2010. NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in Arabidopsis thaliana. *Plant J*.
Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, Harada JJ. 2003. LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell* **15**, 5-18.
Le BH, Wagmaister JA, Kawashima T, Bui AQ, Harada JJ, Goldberg RB. 2007. Using genomics to study legume seed development. *Plant Physiol* **144**, 562-74.
Lee H, Fischer RL, Goldberg RB, Harada JJ. 2003. Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. *Proc Natl Acad Sci U S A* **100**, 2152-6.
Libault M, Thibivilliers S, Bilgin DD, O. R, Benitez M, Clough SJ, Stacey G. 2008. Identification of four soybean reference genes for gene expression normalization. *PLant Genome* **1**, 44-54.
Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-2.
Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-8.
Lotan T, Ohto M, Yee KM, West MA, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ. 1998. Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* **93**, 1195-205.
Luerssen H, Kirik V, Herrmann P, Misera S. 1998. FUSCA3 encodes a protein with a conserved VP1/AB13-like B3 domain which is of functional importance for the regulation of seed maturation in Arabidopsis thaliana. *Plant J* **15**, 755-64.
Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151-5.

Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459-73.
Maity SN, de Crombrugghe B. 1998. Role of the CCAAT-binding protein CBF/NF-Y in transcription. *Trends Biochem Sci* **23**, 174-8.
Mantovani R. 1998. A survey of 178 NF-Y binding CCAAT boxes. *Nucleic Acids Res* **26**, 1135-43.
Marzabal P, Busk PK, Ludevid MD, Torrent M. 1998. The bifactorial endosperm box of gamma-zein gene: characterisation and function of the Pb3 and GZM cis-acting elements. *Plant J* **16**, 41-52.
Masiero S, Imbriano C, Ravasio F, Favaro R, Pelucchi N, Gorla MS, Mantovani R, Colombo L, Kater MM. 2002. Ternary complex formation between MADS-box transcription factors and the histone fold protein NF-YB. *J Biol Chem* **277**, 26429-35.
Matuoka K, Chen KY. 2002. Transcriptional regulation of cellular ageing by the CCAAT box-binding factor CBF/NF-Y. *Ageing Res Rev* **1**, 639-51.
Matuoka K, Yu Chen K. 1999. Nuclear factor Y (NF-Y) and cellular senescence. *Exp Cell Res* **253**, 365-71.
Meinke DW, Franzmann LH, Nickle TC, Yeung EC. 1994. Leafy Cotyledon Mutants of Arabidopsis. *Plant Cell* **6**, 1049-64.
Michel D, Salamini F, Bartels D, Dale P, Baga M, Szalay A. 1993. Analysis of a desiccation and ABA-responsive promoter isolated from the resurrection plant Craterostigma plantagineum. *Plant J* **4**, 29-40.
Millar AA, Jacobsen JV, Ross JJ, Helliwell CA, Poole AT, Scofield G, Reid JB, Gubler F. 2006. Seed dormancy and ABA metabolism in Arabidopsis and barley: the role of ABA 8'-hydroxylase. *Plant J* **45**, 942-54.
Mu J, Tan H, Zheng Q, Fu F, Liang Y, Zhang J, Yang X, Wang T, Chong K, Wang XJ, Zuo J. 2008. LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in Arabidopsis. *Plant Physiol* **148**, 1042-54.
Mundy J, Yamaguchi-Shinozaki K, Chua NH. 1990. Nuclear proteins bind conserved elements in the abscisic acid-responsive promoter of a rice rab gene. *Proc Natl Acad Sci U S A* **87**, 1406-10.
Nakashima K, Fujita Y, Kanamori N, Katagiri T, Umezawa T, Kidokoro S, Maruyama K, Yoshida T, Ishiyama K, Kobayashi M, Shinozaki K, Yamaguchi-Shinozaki K. 2009. Three Arabidopsis SnRK2 protein kinases, SRK2D/SnRK2.2, SRK2E/SnRK2.6/OST1 and SRK2I/SnRK2.3, involved in ABA signaling are essential for the control of seed development and dormancy. *Plant Cell Physiol* **50**, 1345-63.
Nakashima K, Yamaguchi-Shinozaki K. 2013. ABA signaling in stress-response and seed development. *Plant Cell Reports* **32**, 959-70.
Nambara E, Marion-Poll A. 2003. ABA action and interactions in seeds. *Trends Plant Sci* **8**, 213-7.
Nambara E, Marion-Poll A. 2005. Abscisic acid biosynthesis and catabolism. *Annu Rev Plant Biol* **56**, 165-85.
Napier JA. 2007. The production of unusual fatty acids in transgenic plants. *Annu Rev Plant Biol* **58**, 295-319.
Napier JA, Graham IA. 2010. Tailoring plant lipid composition: designer oilseeds come of age. *Curr Opin Plant Biol* **13**, 330-7.
Nowack MK, Shirzadi R, Dissmeyer N, Dolf A, Endl E, Grini PE, Schnittger A. 2007. Bypassing genomic imprinting allows seed development. *Nature* **447**, 312-U3.
Ogas J, Kaufmann S, Henderson J, Somerville C. 1999. PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in Arabidopsis. *Proc Natl Acad Sci U S A* **96**, 13839-44.
Okamoto M, Kuwahara A, Seo M, Kushiro T, Asami T, Hirai N, Kamiya Y, Koshiba T, Nambara E. 2006. CYP707A1 and CYP707A2, which encode abscisic acid 8'-hydroxylases, are indispensable for proper control of seed dormancy and germination in Arabidopsis. *Plant Physiol* **141**, 97-107.
Parcy F, Valon C, Kohara A, Misera S, Giraudat J. 1997. The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in concert to control multiple aspects of Arabidopsis seed development. *Plant Cell* **9**, 1265-77.

Peng Y, Jahroudi N. 2002. The NFY transcription factor functions as a repressor and activator of the von Willebrand factor promoter. *Blood* **99**, 2408-17.
Peng Y, Jahroudi N. 2003. The NFY transcription factor inhibits von Willebrand factor promoter activation in non-endothelial cells through recruitment of histone deacetylases. *J Biol Chem* **278**, 8385-94.
Raz V, Bergervoet JH, Koornneef M. 2001. Sequential steps for developmental arrest in Arabidopsis seeds. *Development* **128**, 243-52.
Riechmann JL, Ratcliffe OJ. 2000. A genomic perspective on plant transcription factors. *Curr Opin Plant Biol* **3**, 423-34.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-40.
Rogalski M, Carrer H. 2011. Engineering plastid fatty acid biosynthesis to improve food quality and biofuel production in higher plants. *Plant Biotechnol J* **9**, 554-64.
Rogers JC, Rogers SW. 1992. Definition and functional implications of gibberellin and abscisic acid cis-acting hormone response complexes. *Plant Cell* **4**, 1443-51.
Rolland F, Baena-Gonzalez E, Sheen J. 2006. Sugar sensing and signaling in plants: conserved and novel mechanisms. *Annu Rev Plant Biol* **57**, 675-709.
Rolland F, Moore B, Sheen J. 2002. Sugar sensing and signaling in plants. *Plant Cell* **14 Suppl**, S185-205.
Rolletschek H, Hosein F, Miranda M, Heim U, Gotz KP, Schlereth A, Borisjuk L, Saalbach I, Wobus U, Weber H. 2005. Ectopic expression of an amino acid transporter (VfAAP1) in seeds of Vicia narbonensis and pea increases storage proteins. *Plant Physiol* **137**, 1236-49.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-4.
Ruiz-Lopez N, Sayanova O, Napier JA, Haslam RP. 2012. Metabolic engineering of the omega-3 long chain polyunsaturated fatty acid biosynthetic pathway into transgenic plants. *J Exp Bot* **63**, 2397-410.
Sanjaya, Durrett TP, Weise SE, Benning C. 2011. Increasing the energy density of vegetative tissues by diverting carbon from starch to oil biosynthesis in transgenic Arabidopsis. *Plant Biotechnol J* **9**, 874-83.
Santos Mendoza M, Dubreucq B, Miquel M, Caboche M, Lepiniec L. 2005. LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in Arabidopsis leaves. *FEBS Lett* **579**, 4666-70.
Santos-Filho SVLBO, M.; Santos, M.A. 2004. Absorção de nutrientes durante a fase vegetativa mamoneira em três solos da região de Mossoró, RN sob diferentes níveis de salinidade da água de irrigação*I Congresso Brasileiro da Mamona*. Campina Grande-PB.
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. *Plant J* **54**, 608-20.
Satoh S, Kamada H, Harada H, Fujii T. 1986. Auxin-controlled glycoprotein release into the medium of embryogenic carrot cells. *Plant Physiol* **81**, 931-3.
Schirawski J, Planchais S, Haenni AL. 2000. An improved protocol for the preparation of protoplasts from an established Arabidopsis thaliana cell suspension culture and infection with RNA of turnip yellow mosaic tymovirus: a simple and reliable method. *J Virol Methods* **86**, 85-94.
Schmittgen TD, Livak KJ. 2008. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* **3**, 1101-8.
Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA. 2010. Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178-83.

Sha AH, Li C, Yan XH, Shan ZH, Zhou XA, Jiang ML, Mao H, Chen B, Wan X, Wei WH. 2012. Large-scale sequencing of normalized full-length cDNA library of soybean seed at different developmental stages and analysis of the gene expression profiles based on ESTs. *Mol Biol Rep* **39**, 2867-74.
Shen B, Allen WB, Zheng P, Li C, Glassman K, Ranch J, Nubel D, Tarczynski MC. 2010. Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize. *Plant Physiol* **153**, 980-7.
Shiota H, Satoh R, Watabe K, Harada H, Kamada H. 1998. C-ABI3, the carrot homologue of the Arabidopsis ABI3, is expressed during both zygotic and somatic embryogenesis and functions in the regulation of embryo-specific ABA-inducible genes. *Plant Cell Physiol* **39**, 1184-93.
Siefers N, Dang KK, Kumimoto RW, Bynum WEt, Tayrose G, Holt BF, 3rd. 2009. Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. *Plant Physiol* **149**, 625-41.
Sinha S, Kim IS, Sohn KY, de Crombrugghe B, Maity SN. 1996. Three classes of mutations in the A subunit of the CCAAT-binding factor CBF delineate functional domains involved in the three-step assembly of the CBF-DNA complex. *Mol Cell Biol* **16**, 328-37.
Sinha S, Maity SN, Lu J, de Crombrugghe B. 1995. Recombinant rat CBF-C, the third subunit of CBF/NFY, allows formation of a protein-DNA complex with CBF-A and CBF-B and with yeast HAP2 and HAP3. *Proc Natl Acad Sci U S A* **92**, 1624-8.
Skriver K, Olsen FL, Rogers JC, Mundy J. 1991. cis-acting DNA elements responsive to gibberellin and its antagonist abscisic acid. *Proc Natl Acad Sci U S A* **88**, 7266-70.
Snyder CL, Yurchenko OP, Siloto RM, Chen X, Liu Q, Mietkiewska E, Weselake RJ. 2009. Acyltransferase action in the modification of seed oil biosynthesis. *N Biotechnol* **26**, 11-6.
Steimer A, Schob H, Grossniklaus U. 2004. Epigenetic control of plant development: new layers of complexity. *Curr Opin Plant Biol* **7**, 11-9.
Stephenson TJ, McIntyre CL, Collet C, Xue GP. 2007. Genome-wide identification and expression analysis of the NF-Y family of transcription factors in Triticum aestivum. *Plant Mol Biol* **65**, 77-92.
Stone SL, Braybrook SA, Paula SL, Kwong LW, Meuser J, Pelletier J, Hsieh TF, Fischer RL, Goldberg RB, Harada JJ. 2008. Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc Natl Acad Sci U S A* **105**, 3151-6.
Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, Goldberg RB, Harada JJ. 2001. LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proc Natl Acad Sci U S A* **98**, 11806-11.
Suzuki M, Wang HH, McCarty DR. 2007. Repression of the LEAFY COTYLEDON 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes. *Plant Physiol* **143**, 902-11.
Takaiwa F, Yamanouchi U, Yoshihara T, Washida H, Tanabe F, Kato A, Yamada K. 1996. Characterization of common cis-regulatory elements responsible for the endosperm-specific expression of members of the rice glutelin multigene family. *Plant Mol Biol* **30**, 1207-21.
Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-9.
Tan H, Yang X, Zhang F, Zheng X, Qu C, Mu J, Fu F, Li J, Guan R, Zhang H, Wang G, Zuo J. 2011. Enhanced seed oil production in canola by conditional expression of Brassica napus LEAFY COTYLEDON1 and LEC1-LIKE in developing seeds. *Plant Physiol* **156**, 1577-88.
Tanaka M, Kikuchi A, Kamada H. 2008. The Arabidopsis histone deacetylases HDA6 and HDA19 contribute to the repression of embryonic properties after germination. *Plant Physiol* **146**, 149-61.
Thirumurugan T, Ito Y, Kubo T, Serizawa A, Kurata N. 2008. Identification, characterization and interaction of HAP family genes in rice. *Mol Genet Genomics* **279**, 279-89.
To A, Valon C, Savino G, Guilleminot J, Devic M, Giraudat J, Parcy F. 2006. A network of local and redundant gene regulation governs Arabidopsis seed maturation. *Plant Cell* **18**, 1642-51.
Tollefson J. 2010. Food: The global farm. *Nature* **466**, 554-6.

Tonon T, Harvey D, Larson TR, Graham IA. 2002. Long chain polyunsaturated fatty acid production and partitioning to triacylglycerols in four microalgae. *Phytochemistry* **61**, 15-24.
Truax AD, Greer SF. 2012. ChIP and Re-ChIP assays: investigating interactions between regulatory proteins, histone modifications, and the DNA sequences to which they bind. *Methods Mol Biol* **809**, 175-88.
Tsukagoshi H, Morikami A, Nakamura K. 2007. Two B3 domain transcriptional repressors prevent sugar-inducible expression of seed maturation genes in Arabidopsis seedlings. *Proc Natl Acad Sci U S A* **104**, 2543-7.
USDA DAdA. 2013.
Vicente-Carbajosa J, Carbonero P. 2005. Seed maturation: developing an intrusive phase to accomplish a quiescent state. *Int J Dev Biol* **49**, 645-51.
Vicente-Carbajosa J, Moose SP, Parsons RL, Schmidt RJ. 1997. A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proc Natl Acad Sci U S A* **94**, 7685-90.
Vital MSM, M; Meneses, C.H.S.G.; Bezerra, CS. 2005. Comparação entre Protocolos para Extração de DNA Total de Ricinus communis L*EMBRAPA: Comunicado Técnico*.
Voelker T, Kinney AJ. 2001. Variations in the Biosynthesis of Seed-Storage Lipids. *Annu Rev Plant Physiol Plant Mol Biol* **52**, 335-61.
Voinnet O, Rivas S, Mestre P, Baulcombe D. 2003. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J* **33**, 949-56.
Wagner DB, Furnier GR, Saghai-Maroof MA, Williams SM, Dancik BP, Allard RW. 1987. Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. *Proc Natl Acad Sci U S A* **84**, 2097-100.
Walter M, Chaban C, Schutze K, Batistic O, Weckermann K, Nake C, Blazevic D, Grefen C, Schumacher K, Oecking C, Harter K, Kudla J. 2004. Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant J* **40**, 428-38.
Wang H, Guo J, Lambert KN, Lin Y. 2007. Developmental control of Arabidopsis seed oil biosynthesis. *Planta* **226**, 773-83.
Wang H, Qi Q, Schorr P, Cutler AJ, Crosby WL, Fowke LC. 1998. ICK1, a cyclin-dependent protein kinase inhibitor from Arabidopsis thaliana interacts with both Cdc2a and CycD3, and its expression is induced by abscisic acid. *Plant J* **15**, 501-10.
Wang L, Ruan YL. 2012. New Insights into Roles of Cell Wall Invertase in Early Seed Development Revealed by Comprehensive Spatial and Temporal Expression Patterns of GhCWIN1 in Cotton. *Plant Physiol* **160**, 777-87.
Wang L, Ruan YL. 2013. Regulation of cell division and expansion by sugar and auxin signaling. *Front Plant Sci* **4**, 163.
Watanabe Y, Yamamoto S. 2009. Hōjinzei kihon tsūtatsu no gimonten. Tōkyō: Gyōsei.
Weber H, Borisjuk L, Wobus U. 2005. Molecular physiology of legume seed development. *Annu Rev Plant Biol* **56**, 253-79.
West M, Harada JJ. 1993. Embryogenesis in Higher Plants: An Overview. *Plant Cell* **5**, 1361-9.
Wilhelm C, Jakob T. 2011. From photons to biomass and biofuels: evaluation of different strategies for the improvement of algal biotechnology based on comparative energy balances. *Appl Microbiol Biotechnol* **92**, 909-19.
Wobus U, Weber H. 1999. Seed maturation: genetic programmes and control signals. *Curr Opin Plant Biol* **2**, 33-8.
Wu CY, Suzuki A, Washida H, Takaiwa F. 1998. The GCN4 motif in a rice glutelin gene is essential for endosperm-specific gene expression and is activated by Opaque-2 in transgenic rice plants. *Plant J* **14**, 673-83.
Xi DM, Zheng CC. 2011. Transcriptional regulation of seed storage protein genes in Arabidopsis and cereals. *Seed Science Research* **21**, 247–54.
Xie T, Sun Y, Du K, Liang B, Cheng R, Zhang Y. 2012. Optimization of heterotrophic cultivation of Chlorella sp. for oil production. *Bioresour Technol* **118**, 235-42.

Xie Z, Li X, Glover BJ, Bai S, Rao GY, Luo J, Yang J. 2008. Duplication and functional diversification of HAP3 genes leading to the origin of the seed-developmental regulatory gene, LEAFY COTYLEDON1 (LEC1), in nonseed plant genomes. *Mol Biol Evol* **25**, 1581-92.
Xing Y, Fikes JD, Guarente L. 1993. Mutations in yeast HAP2/HAP3 define a hybrid CCAAT box binding domain. *EMBO J* **12**, 4647-55.
Yamaguchi-Shinozaki K, Mundy J, Chua NH. 1990. Four tightly linked rab genes are differentially expressed in rice. *Plant Mol Biol* **14**, 29-39.
Yamamoto A, Kagaya Y, Toyoshima R, Kagaya M, Takeda S, Hattori T. 2009. Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. *Plant J* **58**, 843-56.
Yang SH, Zeevaart JA. 2006. Expression of ABA 8'-hydroxylases in relation to leaf water relations and seed development in bean. *Plant J* **47**, 675-86.
Yazawa K, Kamada H. 2007. Identification and characterization of carrot HAP factors that form a complex with the embryo-specific transcription factor C-LEC1. *J Exp Bot* **58**, 3819-28.
Yazawa K, Takahata K, Kamada H. 2004. Isolation of the gene encoding Carrot leafy cotyledon1 and expression analysis during somatic and zygotic embryogenesis. *Plant Physiol Biochem* **42**, 215-23.
Yu Y, Li Y, Huang G, Meng Z, Zhang D, Wei J, Yan K, Zheng C, Zhang L. 2011. PwHAP5, a CCAAT-binding transcription factor, interacts with PwFKBP12 and plays a role in pollen tube growth orientation in Picea wilsonii. *J Exp Bot* **62**, 4805-17.