UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

Edição de RNA plastidial em *Glycine max*: caracterização de sítios de edição, componentes do editossomo e efeitos de estresse abiótico

Nureyev Ferreira Rodrigues

Porto Alegre
2018

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

Edição de RNA plastidial em *Glycine max*: caracterização de sítios de edição, componentes do editossomo e efeitos de estresse abiótico

Nureyev Ferreira Rodrigues

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Doutor em Ciências (Genética e Biologia Molecular).

Orientador: Prof. Dr. Rogerio Margis
Coorientadora: Profª Drª Franceli Rodrigues Kulcheski

Porto Alegre, janeiro de 2018

INSTITUIÇÕES E FONTES FINANCIADORAS

Este trabalho foi realizado no Laboratório de Genômica e Populações de Plantas, Centro de Biotecnologia e Departamento de Biofísica da Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil com apoio financeiro da FAPERGS e CNPq. O doutorando obteve bolsa de estudos do CNPq (48 meses).

À minha família

AGRADECIMENTOS

A Deus, por sua graça e misericórdia.

Aos meus pais, por acreditarem e contribuírem de todas as formas possíveis, sem impedimentos, sempre que precisei.

À minha esposa, Renata, pela parceria e confiança no projeto de vida que escolhemos.

Aos meus irmãos, que também, mesmo de longe e da forma que puderam, me ajudaram, motivaram e apoiaram.

Ao meu orientador, Rogerio Margis, pela oportunidade de trabalhar sob sua orientação; pela confiança e principalmente paciência.

À minha coorientadora, Franceli Rodrigues Kulcheski, pela orientação e contribuição nesse trabalho.

Aos colegas do LGPP, Frank, Guilherme Cordenonsi, Maria, Henrique, Isabel, Pabulo, Érika e Débora, pela motivação nas atividades, científicas ou não.

Às ex-colegas de LGPP, Ana e Priscila, pelo tempo compartilhado e trabalho produzido.

Aos amigos da Igreja Presbiteriana de Porto Alegre, que nos receberam e cuidaram de nós quando precisamos.

Aos componentes da banca, por aceitarem avaliar e contribuir na finalização desse trabalho.

Ao Elmo, pelo auxílio e presteza em todos os momentos.

À coordenação do PGGBM pelo apoio e empenho em fazer esse programa cada vez melhor.

Ao CNPq pela concessão da bolsa.

Ao povo brasileiro, o verdadeiro financiador desse trabalho.

SUMÁRIO

ABREVIATURAS

A-para-I – adenosina para inosina, do inglês Adenosine-to-Inosine
CDS – sequência codificante, do inglês coding sequence
coxII - citocromo c oxidase II, do inglês cytochrome c oxidase subunit II
coxIII - citocromo c oxidase III, do inglês cytochrome c oxidase subunit III
CP31 – do inglês chloroplast ribonucleoprotein 31
C-para-U – citidina para uridina, do inglês Citidine-to-Uridine
GluR-B – receptor de glutamato B, do inglês glutamate receptor B
ha – hectare
HMM – do inglês, hidden Markov model
miRNA – microRNA
mRNA – RNA mensageiro, do inglês messenger RNA
NADH – nicotinamida adenina dinucleotídio em sua forma reduzida, do inglês nicotinamide adenine dinucleotide
ncRNA - RNA não codificante, do inglês non-coding RNA
NGS – sequenciamento de nova geração, do inglês next generation sequencing
OCP3 – do inglês overexpressor of cationic peroxidase 3
ORRM – do inglês Organelle-localized RNA-Recognition Motif-containing
PPO1 – do inglês protoporphyrinogen IX oxidase 1
PPR – do inglês pentatricopeptide repeat
PS – fotossistema, do inglês photosystem
RIP – do inglês RNA editing factor interacting protein
RNA – ácido ribonucleico, do inglês ribonucleic acid
RNAi – RNA de interferência
RNA-seq – sequenciamento de RNA, do inglês RNA Sequencing
RT-qPCR – do inglês real time quantitative polymerase chain reaction
siRNAs - pequenos RNAs de interferência, do inglês small interfering RNA
SNP – do inglês single nucleotide polymorphism
sRNA – do inglês small RNA
tRNA – RNA transportador, do inglês transfer RNA
UTR – região não traduzida, do inglês untranslated region

RESUMO

Soja, uma cultura conhecida por sua importância econômica e nutricional, tem sido objeto de vários estudos que avaliam o impacto e as respostas efetivas das plantas aos estresses abióticos. O estresse salino é um dos principais estresses ambientais e afeta negativamente o crescimento e o rendimento das culturas, incluindo a soja. A edição de RNA é um processo pelo qual as sequências de nucleotídeos podem ser alteradas, revertendo mutações que podem mudar as sequências de proteínas para manter suas funções conservadas. As proteínas pentatricopeptide repeat (PPRs) são trans-elementos de edição caracterizados por reconhecer cis-elementos específicos de RNA e realizar a reação de edição.

Vários estudos descreveram estes trans-elementos e seus sítios de edição cognatos, mas nem todas as proteínas que compõem o complexo de edição foram identificadas. A perda de eventos de edição de plastídios, resultante de mutações em fatores de edição de RNA ou através de interferência por estresse, leva a alterações de desenvolvimento, de fisiologia e da fotossíntese. O objetivo do presente trabalho é caracterizar os sítios de edição e os fatores associados à edição de RNA em *Glycine max* e a influência de estresses abióticos no processo de edição de RNA em cloroplastos.

No capítulo 1, um método é apresentado para triar a edição de RNA de cloroplasto usando bibliotecas públicas de sRNAs de Arabidopsis, soja e arroz. Entre os sítios de edição previstos, 40,57, 34,78 e 25,31% foram confirmados utilizando sRNAs de Arabidopsis, soja e arroz, respectivamente. A análise de SNPs revelou alterações de C-to-U de 58,2, 43,9 e 37,5% nas respectivas espécies e identificou conhecidas e possíveis novas edições de RNA de adenosina para inosina (A-to-I) em tRNAs. O método e os dados revelam o potencial do uso de sRNA como uma fonte confiável para identificar novos e confirmar sítios de edição conhecidos.

No capítulo 2, o processo de edição de RNA foi avaliado em cloroplastos de plantas de soja sob estresse salino. A abordagem de bioinformática utilizando bibliotecas de sRNAs e mRNAs foi empregada para detectar sítios específicos que mostram diferenças na taxa de edição. RT-qPCR foi usado para medir a taxa de edição nos sítios selecionados. Observamos diferenças nas taxas de edição nos transcritos dos genes *ndhA*, *ndhB*, *rps14* e *rps16* ao comparar os dados das bibliotecas controle e das tratadas com NaCl. Os ensaios de RT-qPCR

demonstraram um aumento na edição dos genes selecionados. Esses aumentos podem ser uma resposta para manter a homeostase das funções das proteínas do cloroplasto em resposta ao estresse salino.

No capítulo 3, para identificar os fatores relacionados aos sítios de edição analisados, sondas biotiniladas de RNA foram projetadas com base nos sítios de edição de RNA de plastídio de soja para realizar um isolamento proteico específico do fator de edição. Proteínas que interagiram com as sondas foram isoladas através da ligação das sondas à biotina e foram identificadas utilizando espectrometria de massa. Entre os peptídeos detectados, cinco corresponderam a proteínas PPR. A comparação dos genes de Arabidopsis com as proteínas PPR da soja permitiu a identificação dos homólogos mais próximos.

O presente estudo representa a primeira identificação do conjunto de sítios de edição de RNA, de fatores associados aos sítios de edição de RNA e a caracterização dos efeitos do estresse abiótico na edição de RNA em *Glycine max*.

ABSTRACT

Soybean, a crop known by its economic and nutritional importance, has been the subject of several studies that assess the impact and the effective plant responses to abiotic stresses. Salt stress is one of the main environmental stresses and negatively impacts crop growth and yield. RNA editing is a process whereby nucleotide sequences can be altered, reverting mutations that could change protein sequences to maintain their conserved functions. Pentatricopeptide repeat proteins are editing trans-elements characterized by recognize specific RNA cis-elements and perform the editing reaction.

Several studies have described these trans-elements and their cognate editing sites, but not all proteins that compose the editing complex were identified. The loss of plastid editing events, resulting from mutations in RNA editing factors or through stress interference, leads to developmental, physiological and photosynthetic alterations. The aim of the present work is to characterize the editing sites and factors associated with RNA editing in *Glycine max* and the influence of abiotic stresses on the process of RNA editing in chloroplasts.

In chapter 1, a method is presented to screen chloroplast RNA editing using public sRNA libraries from Arabidopsis, soybean and rice. Among the predicted editing sites, 40.57, 34.78, and 25.31% were confirmed using sRNAs from Arabidopsis, soybean and rice, respectively. SNP analysis revealed 58.2, 43.9, and 37.5% new C-to-U changes in the respective species and identified known and new putative adenosine to inosine (A-to-I) RNA editing in tRNAs. The method and data reveal the potential of sRNA as a reliable source to identify new and confirm known editing sites.

In chapter 2, RNA editing process was evaluated in the chloroplast of soybean plants under salt stress. Bioinformatics approach using sRNA and mRNA libraries was employed to detect specific sites showing differences in editing efficiency. RT-qPCR was used to measure editing efficiency at selected sites. We observed differences in *ndhA, ndhB, rps14* and *rps16* editing rates between control and salt-treated libraries. RT-qPCR assays demonstrated an increase in editing efficiency of selected genes. These increases can be a response to keep the homeostasis of chloroplast protein functions in response to NaCl stress.

In chapter 3, to identify the trans-acting factors of editing sites analyzed, we have designed RNA biotinylated probes based in soybean plastid RNA editing sites to perform

specific isolation of proteins associated to editosomes. Proteins that interacted with the probes were isolated by binding the probes to biotin and were identified using mass spectrometry. Among the detected peptides, five corresponded to PPR proteins. Comparison of Arabidopsis genes to the soybean PPR proteins allow identification of the closest related homologs.

The present study represents the first identification of RNA editing sites set, associated factors to RNA editing sites and characterization of effects from abiotic stress in RNA editing in *Glycine max*.

1. INTRODUÇÃO

1.1. Soja

A soja (*Glycine max* (L.) Merrill) é uma leguminosa anual pertencente à família Fabaceae, subgênero *Soja* dentro do gênero *Glycine* (Doyle et al. 2004). Análises morfológicas, citogenéticas e moleculares indicam que a soja foi domesticada a partir da soja selvagem, *Glycine soja* (Broich and Palmer 1980; Kollipara et al. 1997; Doyle et al. 2004). *G. soja* e *G.max* têm ambas 20 cromossomos (2n = 40), hibridizam com facilidade, exibem emparelhamento normal de cromossomos meióticos e geram híbridos férteis viáveis (Kim et al. 2010). Estudos relacionados à domesticação da soja têm reforçado a hipótese de múltipla domesticação em diversos locais do leste da Ásia e datado esses eventos entre 9000 e 5000 anos atrás (Lee et al. 2011).

Hoje, a soja é uma das cultura mais valiosas do mundo, usada como alimento para bilhões de animais, como fonte de proteína e óleo por milhões de pessoas, bem como na fabricação industrial de milhares de produtos (Nwokolo 1996). A alta demanda de proteína em gêneros alimentícios para consumo humano e animal levou a uma maior expansão da produção de oleaginosas e favoreceu o aumento da produção de soja, especialmente no Brasil (Guevara et al. 2015). Atualmente, o Brasil é o segundo maior produtor, atrás dos Estados Unidos. Em 2015, a produção de soja alcançou seu recorde; no total, 97 464 936 toneladas foram colhidas, tendo um aumento de 10,7 milhões de toneladas (12,3%) em relação a produção de 2014 (IBGE 2015).

Em 2016, devido à seca que assolou alguns estados produtores, a produção foi de 96 296 714 toneladas, uma redução de 1,2% em relação à produção de 2015, e o valor dessa produção somou 104,9 bilhões de reais. Do total da produção, 67,3 milhões de toneladas (69,9%) foram exportados, tendo como principal destino o mercado chinês (IBGE 2016). Espera-se que a soja continue sendo o produto de exportação mais lucrativo com mais da metade da produção brasileira destinada aos mercados mundiais (Guevara et al. 2015). Mato Grosso, Paraná e Rio Grande do Sul são, nessa ordem, os maiores produtores de soja, correspondendo a cerca de 60% do total da produção. Em 2016 o Rio Grande do Sul produziu 16 209 892 toneladas de soja, alta de 3,2% em relação com 2015. O aumento da produção é devido à ampliação da área colhida, 5 436 653 hectares, 3,3% em relação a 2015, juntamente com uma estabilidade na produtividade (IBGE 2016).

Além de sua importância econômica, a soja também tem se destacado como planta modelo para diversos estudos genéticos. O genoma da soja foi totalmente sequenciado no final de 2008 e publicado em 2010 (Schmutz et al. 2010; Cannon and Shoemaker 2012). Associados ao avanço nas tecnologias de sequenciamento, a montagem e anotação do genoma de soja permitiu o desenvolvimento de diversos estudos de genômica comparativa, análises filogenéticas e evolutivas de famílias gênicas, bem como a associação entre variações genéticas e traços agrícolas importantes, incluindo tolerância à diversos tipos de estresses (Choi et al. 2004; Kim et al. 2012; Ma et al. 2012; Zhang et al. 2013; Zhou et al. 2015). Novas ferramentas para a análise de grandes conjuntos de dados têm permitido integrar dados de estudos genômicos, transcritômicos e proteômicos de uma coleção diversificada de tecidos sob diferentes condições, fornecendo dados valiosos que permitiram o avanço na agricultura de leguminosas (Komatsu and Ahsan 2009; Severin et al. 2010; Mathesius et al. 2011).

A evolução nas tecnologias de sequenciamento de nova geração (do inglês, next generation sequencing - NGS) e de ferramentas de bioinformática permitiu o avanço não só de estudos genômicos e transcritômicos, mas também de identificação de pequenos RNAs (sRNAs) (Kulcheski et al. 2011; Borges and Martienssen 2015). Esses estudos focam no papel dos sRNAs na manutenção do genoma, no desenvolvimento, nas respostas das plantas às mudanças ambientais e nas defesas contra patógenos (Ruiz-Ferrer and Voinnet 2009; Simon et al. 2009; Long et al. 2015; Xu et al. 2015). Tais conhecimentos permitiram a criação de tecnologias que utilizam sRNAs para o melhoramento genético. A tecnologia RNAi, baseada em pequenos RNAs de interferência (siRNAs), evoluiu como uma importante ferramenta de engenharia genética e genômica funcional destinada à melhoria das culturas (Kamthan et al. 2015). MicroRNAs (miRNAs), uma classe de pequenos RNAs que regulam a expressão gênica por meio da degradação ou do bloqueio de tradução dos mRNAs alvos (Bartel 2004), foram identificados no genoma de soja (Liu et al. 2010; Turner et al. 2012). Além disso, o papel de alguns desses miRNAs na resposta a estresses bióticos e abióticos têm sido elucidados (Zeng et al. 2010; Kulcheski et al. 2011; Li et al. 2011).

Nos últimos anos, os estudos de sRNAs aumentaram consideravelmente, particularmente associados à sequenciamento de miRNAs e outros pequenos RNAs não codificantes (ncRNAs) de origem nuclear, produzindo uma grande quantidade de novos

dados de sequenciamento. Dessa forma, quantidades consideráveis de dados de sRNA estão disponíveis em bancos de dados públicos e podem ser empregados em diversos estudos.

1.2. Estresses abióticos

Como organismos sésseis, as plantas estão mais expostas a uma série de condições de estresse como variações de temperatura e intensidade da luz, inundações, seca, salinidade e presença de metais pesados no solo. Plantas requerem, fundamentalmente, energia proveniente da luz, água, carbono e nutrientes minerais para seu crescimento. O estresse abiótico é definido como condições ambientais que reduzem o crescimento e a produção abaixo dos níveis ótimos (Cramer et al. 2011). Estresses abióticos afetam a planta inteira, comprometendo aspectos moleculares e fisiológicos básicos, da germinação às fases de reprodução (Mahajan and Tuteja 2005). Entre os estresses supracitados, o estresse salino é um dos principais estresses ambientais e afeta espécies de culturas economicamente importantes que são sensíveis à salinidade, como feijão (*Phaseolus vulgaris*), milho (*Zea mays*), arroz (*Oryza sativa*) e soja (Wang et al. 2003; Zheng et al. 2009). Os solos afetados pela salinidade ocorrem em mais de 100 países e sua extensão mundial é estimada em cerca de 1 bilhão de ha (FAO and ITPS 2015).

A salinidade afeta diversos componentes moleculares e funções fisiológicas como lipídios (Alvarez-Pizarro et al. 2009), níveis de íons (He et al. 2015), assimilação e metabolismo de nitrogênio (Silveira et al. 2001), enzimas antioxidantes (Gill and Tuteja 2010), componentes proteicos e estrutura dos cloroplastos (Feller et al. 2008; He et al. 2014) e especialmente a fotossíntese (Wang et al. 2001; Parida and Das 2005). Diversos estudos têm caracterizado o impacto da salinidade na atividade fotossintética (Lu et al. 2009), na assimilação de carbono (Chaves et al. 2009), na composição de pigmentos (Koyro 2006), no transporte de elétrons e na eficiência dos fotossistemas I e II (Lu et al. 2002; Munekage et al. 2004; Lu et al. 2008; Kalaji et al. 2011). Claramente, devido aos efeitos na fotossíntese, é necessário compreender quais processos moleculares podem estar sofrendo os efeitos do estresse salino no cloroplasto (Gomez 2003; Feller et al. 2008; Zhang et al. 2008; Zheng et al. 2009).

1.3. Edição de RNA

A edição de RNA é um processo pós-transcricional que altera a informação genética contida em moléculas de RNA pela inserção, remoção ou alteração de nucleotídeos (Takenaka et al. 2013). Esse processo ocorre em transcritos codificados pelo genoma

nuclear, mitocondrial ou de cloroplasto, em uma ampla gama de organismos. A edição de RNA foi descoberta em 1986 em *Trypanosoma brucei*, quando demostrou-se que uridinas foram inseridas em locais específicos em transcritos da citocromo c oxidase II (*coxII*) restaurando a sequência adequada para a codificação da proteína (Benne et al. 1986). Após isso, em 1988, foram descritas deleções de uridinas no transcrito de *coxIII* (Feagin et al. 1988). Em mamíferos, foram descritas conversões de citidina em uridina (C-para-U) em transcritos de apolipoproteína-B48 (Blanc and Davidson 2010) e de adenosina (A) para inosina (I) (A-para-I), em transcritos do gene *GluR-B* que codifica um receptor B de glutamato (Sommer et al. 1991), em elementos repetitivos presente em íntrons e 3'-UTRs de transcritos em cérebros humanos (Kim et al. 2004) e microRNAs (Chawla and Sokol 2014). No entanto, uma pequena fração de edições A-para-I estão localizadas em éxons, podendo mudar sítios de splicing ou levar a alteração não sinônimas de códons (Nishikura 2010). A edição A-para-I é comum nos metazoários (Albertin et al. 2015; Porath et al. 2017).

Em plantas, estudos de edição de RNA estão voltados especificamente à mitocôndrias e cloroplastos, embora recentemente, modificações no RNA de transcritos nucleares tenham sido identificadas em *A. thaliana*, todavia, somente através de análises *in silico* (Meng et al. 2010). A edição mais comum ocorre em citidinas específicas, onde através de uma reação de deaminação, elas são modificadas para uridinas (C-para-U). A edição reversa (U-para-C) também ocorre em transcritos de mitocôndrias e cloroplastos, mas parece ser restrita a algumas briófitas, licopódios e samambaias (Kugita 2003; Wolf et al. 2004; Grewe et al. 2011; Guo et al. 2015). Outro tipo de edição, pouco estudado em comparação a edição C-para-U é a edição de tRNAs. A edição de anticódons de tRNAs pela desaminação de adenina para inosina é usada por procariotos e eucariotos para expandir a capacidade de decodificação de tRNAs individuais (Schaub and Keller 2002). A edição A-para-I em tRNA em cloroplastos já foi descrita, bem como a enzima responsável pela reação. Em Arabidopsis, o mutante para a proteína codificada pelo locus At1g68720, uma tRNA arginina adenosina deaminase, demonstrou uma drástica redução de proteínas codificadas no cloroplasto e deficiência na função fotossintética (Delannoy et al. 2009).

A edição C-para-U de RNA em organelas tem sido descrita ocorrendo em íntrons, tRNAs e mRNAs. Sítios de edição dentro de íntrons do grupo II possuem importância funcional; a edição melhora o emparelhamento de bases, estabilizando o dobramento da estrutura necessária para o splicing (Carrillo 1997; Vogel et al. 1997; Castandet et al. 2010).

Em feijão, um evento de edição no $tRNA^{Phe}$ corrige a incompatibilidade do pareamento C:A para um pareamento U:A na extremidade aceptora (Mareéchal-Drouard et al. 1993). No $tRNA^{His}$ de coníferas do gênero *Larix*, três eventos de edição, na extremidade aceptora, na haste da alça D e na haste da alça do anticódon, são necessários para o correto processamento desse tRNA (Maréchal-Drouard et al. 1996). Na briófita *Takakia lepidozioides*, a edição gera um anticódon UAA canônico no $tRNA^{Leu}$(CAA) antes mesmo do splicing que gera o tRNA maduro (Miyata et al. 2008).

Em sequências codificantes, a maioria dos eventos de edição ocorrem na primeira ou segunda posição dos códons (Takenaka et al. 2013) e, portanto, geralmente resultam em alterações de códons, formando códons de iniciação e terminação (Oldenkott et al. 2014), bem como em alguns casos, à mudança de aminoácido, restaurando códons que são essenciais para a expressão de proteínas funcionais (Bock et al. 1994; Sasaki et al. 2001). Sem a edição de RNA várias proteínas da cadeia respiratória seriam produzidas com uma sequência que levariam a proteínas não funcionais, e assim nenhuma mitocôndria funcional poderia ser mantida em plantas (Takenaka et al. 2008). A caracterização da maior parte dos sítios de edição que tornam proteínas funcionais deu-se através da caracterização dos *trans*-elementos que são responsáveis pela edição nesses sítios.

A quantidade de sítios de edição em cloroplastos varia entre as espécies. Foram identificados 2 sítios de edição em *Physcomitrella patens* e 509 em *Anthoceros formorsae*, ambas briófitas (Kugita 2003; Miyata and Sugita 2004). As pteridófitas *Adiantum capillus-veneris* e *Ophioglossum californicum* apresentam 315 e 297 sítios de edição respectivamente (Wolf et al. 2004); foram identificados 21 sítios de edição em *Oryza sativa* (Corneille et al. 2000) e 43 sítios em *Arabidopsis thaliana* (Ruwe et al. 2013). Essa notável diferença no número de sítios de edição entre espécies tem sido alvo de pesquisas que visam compreender o papel da edição de RNA na evolução do genoma plastidial (Fiebig et al. 2004; Tillich et al. 2006a; Takenaka et al. 2013; Vu and Tsukahara 2017).

Estudos em diversas espécies têm demonstrado que os sítios de edição são reconhecidos pela maquinaria de edição através de *cis*-elementos (Hirose and Sugiura 1996; Neuwirt et al. 2005; Kobayashi et al. 2007). Trechos de 20 a 25 nucleotídeos a montante do local de edição proporcionam um sítio de reconhecimento sequência específico que é alvo da atividade de edição (Bock et al. 1996; Ruf and Bock 2011). Apesar do avanço do reconhecimento dos sítios de edição e seus *cis*-elementos, nem todos os componentes da

maquinaria de edição foram identificados, o que dificulta a compreensão da evolução do processo de edição de RNA.

1.4. Fatores proteicos associados à edição de RNA

1.4.1. Proteínas pentatricopeptide repeat (PPR)

Apesar da identificação dos sítios de edição, bem como seus *cis*-elementos, a primeira identificação de fatores de edição só foi possível através do estudo de mutantes fotossintéticos. O mutante *crr4* (chlororespiratory reduction 4), apresentou defeitos no acúmulo do complexo NADH desidrogenase (NDH) em cloroplastos. Esses defeitos foram relacionados a perda do evento de edição que gera o códon de início (AUG) nos transcritos do gene plastidial *ndhD* (Kotera et al. 2005). O gene *crr4* mutado corresponde à uma proteína pertencente à família PPR (pentatricopeptide repeat). CRR4 foi descrita posteriormente como responsável pelo reconhecimento do sítio de edição em *ndhD* (Okuda et al. 2006). Após a descoberta do CRR4, mapeamentos genéticos identificaram outros fatores adicionais para outros sítios de edição em cloroplasto, todos pertencentes à família de proteínas PPR (pentatricopeptide repeat).

A família de genes que codificam PPRs é encontrada em quase todas as linhagens eucarióticas, mas expandiu-se dramaticamente em plantas. A maioria das algas verdes aquáticas tem cerca de 20 genes PPR, enquanto que plantas terrestres possuem 100 ou mais genes PPR, e há cerca de 400 a 600 genes na maioria dos genomas de angiospermas (Cheng et al. 2016). *Selaginella moellendorffii* apresenta mais de 800 PPRs em seu genoma (Banks et al. 2011). Surpreendemente, em um trabalho recente, cerca de 4000 sequências de PPRs sem sobreposição e não redundantes foram identificadas no genoma de *S. moellendorffii* (Cheng et al. 2016). *A. thaliana*, possui cerca de 490 genes PPR (Lurin et al. 2004; Cheng et al. 2016). Soja possui 1024 genes PPRs identificados (Cheng et al. 2016). Outras angiospermas que passaram por recentes eventos de poliploidia também apresentam um número particularmente grande de genes de PPR; por exemplo, há 1181, 1646, 1392 e 1139 membros PPR em *Panicum virgatum*, *Triticum aestivum*, *Malus domestica* e *Brassica napus*, respectivamente (Cheng et al. 2016).

As proteínas PPRs pertencem à superfamília α-solenóide de proteínas de repetições helicoidais (Hammani et al. 2014). A família PPR é caracterizada por apresentar um motivo altamente degenerado constituído por 35 aminoácidos, que geralmente aparecem como repetições in tandem nas proteínas. Estruturalmente, cada motivo PPR compreende duas α-

hélices antiparalelas (Small and Peeters 2000). As proteínas da família PPR são reconhecidas por participarem no processamento de RNA em cloroplastos e mitocôndrias. PPRs canônicas foram incialmente descritas tendo como característica apresentar somente repetições de 35 aminoácidos, classificados como motivo P. Outros motivos foram posteriormente caracterizados; os de repetições um pouco mais longas que 35 e 36 aminoácidos (L), ou mais curtas, de 31 a 34 aminoácidos (S) (Lurin et al. 2004; O'Toole et al. 2008; Barkan and Small 2014). A família PPR foi então dividida em subfamílias P e PLS de acordo com os motivos que compõem essas proteínas.

Além dos motivos que as compõem, PPRs da subfamília PLS diferem-se da subfamília P na porção C-terminal. Após o último motivo PPR, as PLS comumente possuem os domínios extras E, E+ e DYW. O domínio E é específico de plantas e está presente em quase todas as PLS. Cerca de metade das PLS que possuem domínio E, também possuem o domínio DYW, caracterizado por ser uma extremidade C-terminal altamente conservada constituída por aspartato (D), tirosina (Y) e triptofano (W) (Liu et al. 2016; Ichinose and Sugita 2016). O domínio DYW contém uma assinatura conservada semelhante a de citidina deaminases e diversos estudos tem comprovado sua participação efetiva na reação de edição de RNA (Salone et al. 2007; Boussardon et al. 2014; Wagoner et al. 2015). A organização relativa dos três motivos na extremidade C-terminal segue regras bem caracterizadas: (1) os motivos são observados em cópia única na mesma proteína; (2) quando observados na mesma proteína, são ordenados de forma colinear E – E+ – DYW, DYW sendo o tripeptídeo C-terminal; (3) nas proteínas que possuem o motivo DYW estes quase sempre são precedidos de motivos E e E+; da mesma forma, as proteínas que possuem o motivo E+ sempre têm um motivo E anterior (Lurin et al. 2004). Estudos têm demonstrado que esses domínios são necessários para a atividade de edição de RNA, a qual parece ser a principal função de muitas PPRs tipo PLS (Okuda et al. 2009; Ohtani et al. 2010; Chateigner-Boutin et al. 2013; Pyo et al. 2013; Brehme et al. 2014; Wang et al. 2017).

Recentemente, uma nova classificação foi proposta baseada na análise de motivos PPR, usando Hidden Markov Model (HMM) em 41 genomas de plantas terrestres filogeneticamente diversas (Cheng et al. 2016). Assim, novos motivos foram propostos. Os motivos P1 e P2 foram criados derivados do motivo P, todavia, o motivo P se manteve. Os motivos L1 e L2 substituíram o motivo L. Os motivos S1, S2 e SS substituíram o motivo S. O motivo E foi substituído pelos motivos E1 e E2. Manteve-se o motivo E+ e DYW (Cheng

et al. 2016). Dessa forma, a subfamília PLS foi subdividida em subgrupos: PLS, E1, E2, E+ e DYW (Figura 1).

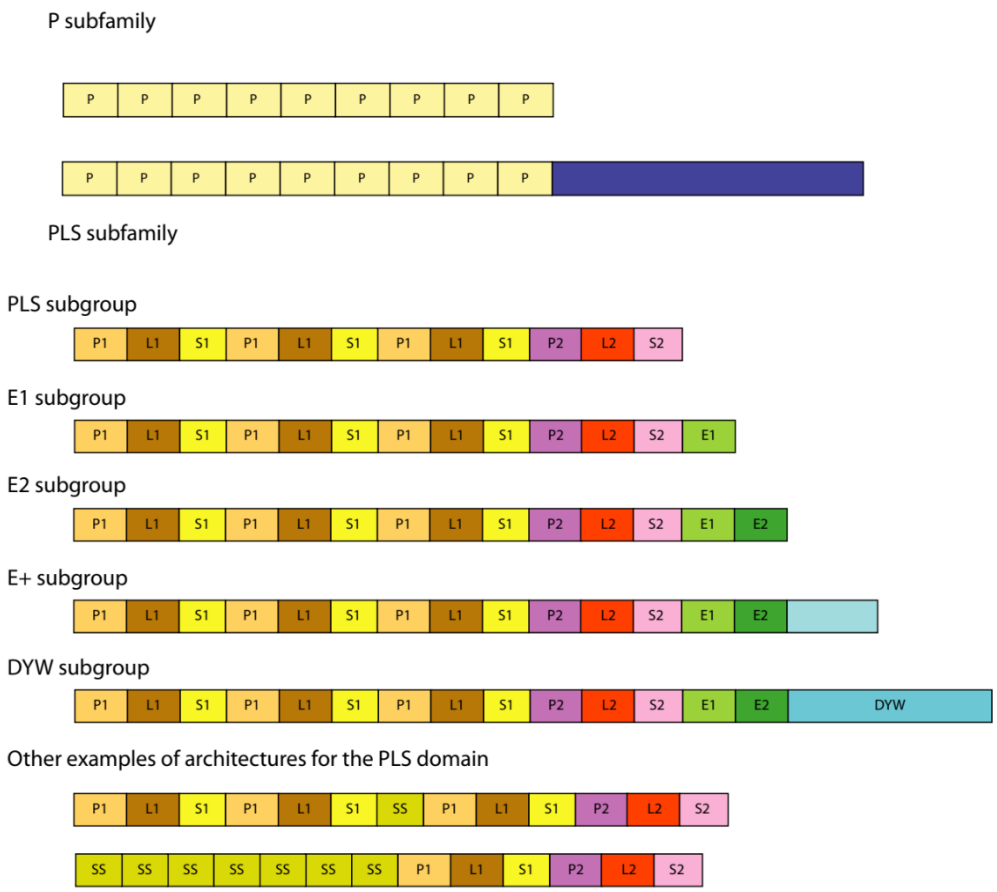

Figura 1. Representação dos motivos que compõe as subfamílias e subgrupos da família de proteínas PPR. O número de motivos em cada proteína pode variar de 2 a 35, e o primeiro motivo pode ser qualquer motivo P, P1, L1, S1 ou SS. O subgrupo E + consiste em proteínas com um domínio DYW degenerado ou truncado (Adaptado de Cheng et al, 2016).

O reconhecimento dos sítios de edição através dos *cis*-elementos é realizado pelos motivos PPRs (P, L e S). Os motivos têm a capacidade de reconhecer RNA de cadeia simples seguindo uma regra de um motivo para uma base (Figura 2) e a composição dos motivos de cada proteína determina a especificidade da ligação com o RNA (Barkan et al. 2012; Okuda

et al. 2014; Kindgren et al. 2015; Cheng et al. 2016). Todavia, apesar dessa especificidade, uma única PPR pode agir em diversos sítios de edição por reconhecer mais de um *cis*-elemento ou por esse *cis*-elemento ser compartilhado entre diferentes sítios de edição (Van Der Merwe et al. 2006; Kobayashi et al. 2007; Heller et al. 2008; Zehrmann et al. 2009; Okuda and Shikanai 2012). Além disso, devido à expansão das PPRs em plantas, alguns estudos têm demonstrado que mais de uma PPR pode agir em um único sítio de edição (Verbitskiy et al. 2012).

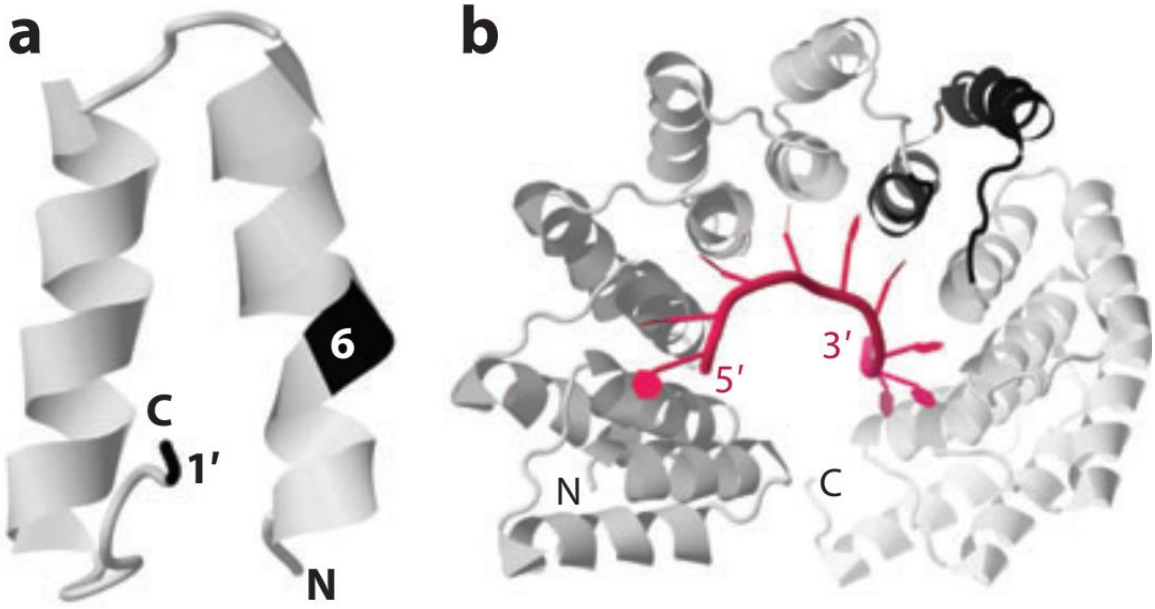

Figura 2. Estrutura de um motivo PPR e modelo da ligação da PPR ao RNA. Uma repetição consiste em um par de hélices antiparalelas, com a hélice N-terminal de cada motivo formando a face de ligação ao RNA e a hélice C-terminal formando a superfície externa da proteína. (a) Um único motivo PPR da RNase P organelar de Arabidopsis (PRORP1). Resíduos das posições 6 e 1' foram propostos para determinar a especificidade de ligação de nucleotídeos dos motivos PPR. (b) Um modelo de 10 repetições PPR (cinza) ligadas a um RNA composto por 9 nucleotídeos de uracila (U) (magenta), orientadas para mostrar que as bases estão previstas para interdigitar com as hélices PPR. O sexto motivo PPR é sombreado em preto para destacar uma única repetição. (Adaptado de Barkan and Small, 2014)

Duas PPRs do tipo DYW, RARE1 e AtECB2 (VAC1), foram identificadas como fatores de edição de um mesmo sítio em transcritos do gene *accD* (posição 794) em cloroplastos de Arabidopsis. A mutação de RARE1 resultou na abolição completa da edição de *accD*-794 (Robbins et al. 2009), enquanto que em AtECB2 levou a uma redução de edição de 60% em relação ao nível do tipo selvagem (Tseng et al. 2010). Análises *in silico* de atribuição de alvo sugeriram que RARE1, mas não AtECB2, seria de fato um fator de reconhecimento para edição de *accD*-794. AtECB2 estaria envolvido na edição de *accD*-794, mas não seria necessária para o reconhecimento do sítio de edição (Yagi et al. 2013b). Dessa forma, PPRs poderiam atuar de forma cooperativa na edição de RNA; a perda de função de uma PPR poderia reduzir, mas não abolir completamente a edição em um sítio de edição específico, já que a edição restante poderia ser realizada por outra PPR (Yagi et al. 2013b; Ichinose and Sugita 2016).

1.4.2. Outros fatores de edição

Proteínas não-PPR foram identificadas como componentes essenciais no editosomo: primeiro a família RIP/MORF (Multiple Organellar RNA-editing Factors), depois a família ORRM (Organelle-localized RNA-Recognition Motif-containing) e, mais tarde, um membro da família OZ (Tabela 1). A imunoprecipitação de RARE1 marcada com epítopo resultou na identificação de uma proteína não-PPR, a proteína RIP1 (RNA editing factor interacting protein 1) (Bentolila et al. 2012). O mutante *rip1* de Arabidopsis tem uma significativa redução da edição em mais de 400 sítios em mitocôndrias e 11 sítios em cloroplastos, tornando-se o fator de edição mais influente já identificado em plantas (Bentolila et al. 2013). A família RIP/MORF é composta por 10 membros em Arabidopsis; destes, cinco são os principais fatores de edição para mitocôndrias e/ou cloroplastos, enquanto membros restantes têm menor ou nenhum efeito na edição (Takenaka et al. 2012; Bentolila et al. 2013).

Tabela 1. Lista de fatores de edição não-PPR (Adaptado de Sun et al. 2016)

Família proteica	Proteína	Localização subcelular	% sítios de edição afetados em cloroplastos	% sítios de edição afetados em mitocôndrias
RIP/MORF	RIP1/MORF8	Dupla*	22%	77%
	RIP2/MORF2	Cloroplasto	100%	NA
	RIP3/MORF3	Mitocôndria	NA	26%
	RIP8/MORF1	Mitocôndria	NA	19%

	RIP9/MORF9	Cloroplasto	97%	NA
ORRM	ORRM1	Cloroplasto	62%	NA
	ORRM2	Mitocôndria	NA	6%
	ORRM3	Mitocôndria	NA	19%
	ORRM4	Mitocôndria	NA	44%
	ORRM5	Mitocôndria	NA	14%
	ORRM6	Cloroplasto	1%	NA
OZ	OZ1	Cloroplasto	81%	NA
	PPO1	Cloroplasto	50%	NA
	OCP3	Cloroplasto	12%	NA
	CP31	Cloroplasto	38%	NA

*Dupla: mitocôndria e cloroplasto

RIP2/MORF2 e RIP9/MORF9 são direcionadas para cloroplastos, enquanto que RIP1/MORF8 é direcionada para cloroplastos e mitocôndrias. Os membros restantes são direcionados para mitocôndrias, com a exceção de RIP10, que pode ser codificado por um pseudogene (Bentolila et al. 2013; Shikanai 2015). Excluindo-se RIP1/MORF8, RIP9/MORF9 e RIP8/MORF1 são responsáveis pela maior parte da edição em cloroplastos e mitocôndria, respectivamente. Devido à sua importância, estudos têm procurado descrever as características estruturais dessas proteínas (Haag et al. 2017). Duas proteínas RIP/MORF podem formar heterodímeros, cuja função pode ser substituída por homodímeros em alguns sítios (Takenaka et al. 2012; Zehrmann et al. 2015a; Glass et al. 2015). Interações seletivas entre PPR-PLS, bem como PPRs que possuem o domínio E e proteínas RIP/MORF sugerem que os domínios que compõem as PPRs e as proteínas MORF desempenham um papel fundamental para que complexos de proteínas específicos se agrupem em diferentes locais para a edição de RNA (Glass et al. 2015; Bayer-Császár et al. 2017; Yan et al. 2017).

O primeiro fator de edição da família ORRM descrito foi ORRM1. ORRM1 possui dois motivos RIP truncados que interagem com RIP1/MORF8 em sua extremidade N-terminal, além de um domínio RRM na extremidade C-terminal, sendo descoberto através de análises *in silico* em bancos de dados com a sequência da proteína RIP1/MORF8 (Sun et al. 2013). ORRM1 controla mais de 60% dos sítios de edição de cloroplastos em Arabidopsis (Sun et al. 2016). Embora o RRM seja um motivo muito comum em eucariotos, o motivo RRM em ORRM1 pertence a um clado distinto de aproximadamente 20 membros em Arabidopsis (Sun et al. 2013). Exceto para ORRM1, nenhum dos outros membros possui um domínio RIP, em vez disso, muitos contêm regiões ricas em glicina (Shi et al. 2017b).

A família ORRM possui 6 membros (ORRM1-6) que estão relacionados com edição de RNA em cloroplastos e mitocôndrias; ORRM1 e 6 em cloroplastos e ORRM2, 3, 4 e 5 em mitocôndrias (Sun et al. 2013; Shi et al. 2015; Shi et al. 2016a; Shi et al. 2016b; Hackett et al. 2017; Shi et al. 2017a). Experimentos de interação proteína-proteína demonstram que as proteínas ORRM interagem com outros componentes dos complexos de edição de RNA (Sun et al. 2016): com PPRs que são necessárias para a edição dos sítios regulados por cada ORRM (Sun et al. 2013); com RIP/MORFs (Hackett et al. 2017) e em alguns casos, com elas próprias, formando homo e heterodímeros (Shi et al. 2015; Shi et al. 2016b). Além disso, as proteínas ORRM se ligam a RNAs com uma ampla gama de afinidades e especificidades (Vermel et al. 2002; Sun et al. 2013; Hackett et al. 2017).

A descoberta da proteína membro da família OZ foi análoga à descoberta de RIP1; OZ1 foi encontrado em um complexo de proteínas coimunoprecipitada com ORRM1 (Sun et al. 2015). O mutante *oz1* em Arabidopsis tem uma perda na edição de 30 sítios em cloroplastos, sendo que em 14 destes, a diminuição é maior que 90% quando comparado ao nível do tipo selvagem. Em Arabidopsis, a família OZ contém 4 membros, OZ1-4, dos quais três estão previstos para serem localizados em cloroplastos enquanto um é mitocondrial (Ichinose and Sugita 2016). O único domínio anotado da família é o zinc-finger tipo RanBP2 que é repetido em diferentes quantidades nos membros OZ; no entanto, existe outro domínio conservado de função desconhecida nas quatro proteínas OZ (Sun et al. 2016). OZ1 interage seletivamente com PPRs e também se associa fortemente a ORRM1 e ORRM6, mas não parece se ligar diretamente a RIPs/MORFs, apesar de apresentar uma interação fraca com RIP1 em ensaios de duplo híbrido (Sun et al. 2013; Sun et al. 2015; Hackett et al. 2017). OZ1 também pode formar homodímeros (Sun et al. 2015). A função dos domínios zinc-finger e dos outros domínios não caracterizados das proteínas OZ na edição de RNA aguardam uma investigação mais aprofundada.

Além das proteínas já citadas e suas respectivas famílias, outras três proteínas adicionais - CP31, PPO1 e OCP3 - afetam a eficiência de edição de RNA em cloroplastos. CP31 é uma proteína que contém domínios RRM. No entanto, numa análise filogenética, CP31 não pertence ao mesmo clado das ORRMs e apesar da perda de CP31 levar a níveis de edição reduzidos em vários sítios, não afeta a edição em um padrão específico de sítios como é visto no mutante *orrm1* (Tillich et al. 2009; Sun et al. 2013; Shi et al. 2017b). Além disso, a transcrição é bastante reduzida em mutantes *cp31*, o que sugere que CP31 pode ser

principalmente um fator de estabilidade de RNA (Tillich et al. 2009; Kupsch et al. 2012; Sun et al. 2016).

A protoporfirinogênio IX oxidase 1 (PPO1) medeia o passo final da via comum compartilhada pela biossíntese de clorofila e heme (Koch et al. 2004). O mutante *ppo1* tem a edição reduzida em 18 sítios de edição, sendo a maioria em transcritos de genes do complexo Ndh (Zhang et al. 2014). Apesar da redução, com exceção de um sítio, nenhum dos sítios afetados perde completamente a edição quando PPO1 está ausente. PPO1 interage diretamente com RIP/MORFs de cloroplastos (RIP1/MORF8 e RIP9/MORF9), mas não com PPRs, sugerindo que PPO1 controle o nível de edição de cloroplasto através da estabilização de RIP/MORFs (Zhang et al. 2014; Sun et al. 2016).

A proteína que menos se conhece sobre sua relação com a edição de RNA até o momento é o fator de transcrição OVEREXPRESSOR OF CATIONIC PEROXIDASE3 (OCP3). OCP3 é endereçado ao cloroplasto e se combina a PPR. Além disso, uma análise demonstrou sua coexpressão com um conjunto de 9 genes que codificam PPRs. Mutantes *ocp3* exibe apenas uma edição suavemente reduzida, afetando a edição de múltiplos sítios em transcritos do gene plastidial *ndhB* (García-Andrade et al. 2013). Embora reduzida, foi suficiente para prejudicar a atividade de NDH, o que consequentemente aumentou a resistência da planta à infecção por fungos (Coego 2005; García-Andrade et al. 2013).

1.5. Efeito de estresses abióticos na edição de RNA

Poucos trabalhos caracterizaram os efeitos de estresses abióticos na edição de RNA em cloroplastos. O aumento da temperatura leva a uma diminuição das taxas de edição de transcritos dos genes *rps14* e *rpl20* em milho, apesar do aumento significante da taxa de transcrição destes genes (Nakajima et al. 2001). Além destes dois genes, efeitos do calor na edição dos transcritos do gene *ndhB* foi demonstrado em tabaco. Quando sob uma temperatura de 42°C, o bloqueio da edição ocorre especificamente em 3 sítios. Além da edição, o processamento através de splicing neste gene é comprometido (Karcher and Bock 2002). Uma perda da interação das proteínas que promovem a ligação aos sítios e o RNA alvo, devido alterações da conformação das proteínas pode levar a efeitos negativos no processo de edição de RNA.

Além da quantidade reduzida de informações dos efeitos dos estresses no processo de edição, os trabalhos que visam caracterizar sítios de edição em RNA plastidial limitam-se à caracterização dos sítios conservados. Dessa forma, o impacto de estresses na edição de

sítios espécie-específicos não pode ser avaliado. A caracterização dos sítios de edição, seus *cis*-elementos e respectivos fatores de edição é um importante passo para a compreensão da história do processo de edição de RNA dentro do processo evolutivo das plantas e suas organelas. A análise da influência de estresses abióticos no processo de edição pode contribuir para a compreensão do papel da edição de RNA nas respostas a esses estresses, bem como na seleção de genes responsivos frente aos estresses.

2. OBJETIVOS

O presente trabalho teve como objetivo a caracterização dos sítios de edição e de fatores associados à edição de RNA em *Glycine max*, e a influência do estresse abiótico no processo de edição de RNA em cloroplastos.

Objetivos específicos:

- Estabelecer um método *in silico* para identificação de sítios de edição em cloroplastos utilizando bibliotecas de sequenciamento de nova geração;
- Identificar os sítios de edição de RNA em cloroplastos de *Glycine max*;
- Avaliar o padrão de edição de RNA em cloroplastos de soja sob estresse salino;
- Identificar PPRs que se ligam a *cis*-elementos de sítios de edição de RNA em cloroplastos de *Glycine max*;

3. Capítulo 1 - Unveiling chloroplast RNA editing events using next generation small RNA sequencing data

Nureyev Ferreira Rodrigues[1], Ana Paula Christoff[1], Guilherme Cordenonsi da Fonseca[2], Franceli Rodrigues Kulcheski[3], Rogerio Margis[1,2,4*]

[1]PPGBM, Departamento de Genética, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brazil.
[2]PPGBCM, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brazil.
[3]PPGBCD, Departamento de Biologia Celular, Genética e Embriologia, Universidade Federal de Santa Catarina - UFSC, Florianópolis, SC, Brazil.
[4]Departamento de Biofísica, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brazil.

Artigo publicado no periódico Frontiers in Plant Science, na seção Bioinformatics and Computational Biology (2017)

frontiers in Plant Science

METHODS
published: 29 September 2017
doi: 10.3389/fpls.2017.01686

Unveiling Chloroplast RNA Editing Events Using Next Generation Small RNA Sequencing Data

*Nureyev F. Rodrigues[1], Ana P. Christoff[1], Guilherme C. da Fonseca[2], Franceli R. Kulcheski[3] and Rogerio Margis[1, 2, 4]**

[1] Programa de Posgraduação em Genética e Biologia Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, [2] Programa de Posgraduação em Biologia Celular e Molecular, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, [3] Programa de Pósgraduação em Biologia Celular e do Desenvolvimento, Departamento de Biologia Celular, Genética e Embriologia, Universidade Federal de Santa Catarina, Florianópolis, Brazil, [4] Departamento de Biofísica, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

OPEN ACCESS

Edited by:
Giovanni Nigita,
The Ohio State University Columbus,
United States

Reviewed by:
Lei Song,
National Cancer Institute (NIH),
United States
Xiyin Wang,
North China University of Science and Technology, China
Gaurav Sablok,
University of Helsinki, Finland
Fabio Iannelli,
IFOM-The FIRC Institute of Molecular Oncology, Italy

***Correspondence:**
Rogerio Margis
rogerio.margis@ufrgs.br

Specialty section:
This article was submitted to Bioinformatics and Computational Biology, a section of the journal Frontiers in Plant Science

Received: 06 June 2017
Accepted: 13 September 2017
Published: 29 September 2017

Citation:
Rodrigues NF, Christoff AP, da Fonseca GC, Kulcheski FR and Margis R (2017) Unveiling Chloroplast RNA Editing Events Using Next Generation Small RNA Sequencing Data. Front. Plant Sci. 8:1686. doi: 10.3389/fpls.2017.01686

Organellar RNA editing involves the modification of nucleotide sequences to maintain conserved protein functions, mainly by reverting non-neutral codon mutations. The loss of plastid editing events, resulting from mutations in RNA editing factors or through stress interference, leads to developmental, physiological and photosynthetic alterations. Recently, next generation sequencing technology has generated the massive discovery of sRNA sequences and expanded the number of sRNA data. Here, we present a method to screen chloroplast RNA editing using public sRNA libraries from Arabidopsis, soybean and rice. We mapped the sRNAs against the nuclear, mitochondrial and plastid genomes to confirm predicted cytosine to uracil (C-to-U) editing events and identify new editing sites in plastids. Among the predicted editing sites, 40.57, 34.78, and 25.31% were confirmed using sRNAs from Arabidopsis, soybean and rice, respectively. SNP analysis revealed 58.2, 43.9, and 37.5% new C-to-U changes in the respective species and identified known and new putative adenosine to inosine (A-to-I) RNA editing in tRNAs. The present method and data reveal the potential of sRNA as a reliable source to identify new and confirm known editing sites.

Keywords: small RNA, chloroplast, RNA editing, NGS, SNP genotyping

INTRODUCTION

Chloroplasts are notable examples of successful endosymbiosis in the early origin of modern life forms. These organelles possess their own gene expression machinery, with complex posttranscriptional processes and fine nucleus-cytosol crosstalk. In plants, these organelles undergo a posttranscriptional process called RNA editing, corresponding to nucleotide changes from cytosine to uracil (C-to-U) and less frequently from uracil to cytosine (U-to-C), in some sites of coding sequences (Tillich et al., 2006; Chateigner-Boutin and Small, 2010). These nucleotide changes correct the codons to encode appropriate amino acids, maintaining the functional amino acid sequence of the evolutionarily conserved protein (Takenaka et al., 2013). Another well-known mechanism of RNA editing is the adenine to inosine (A-to-I) editing, as observed in the chloroplast $tRNA^{Arg}$ (ACG). This type of editing enables hydrogen bond formation with more than one base in the corresponding codon position (Su and Randau, 2011). The A-to-I editing in position 34 of the $tRNA^{Arg}$ (ACG) produces the wobble nucleotide described as essential for efficient chloroplast

translation (Delannoy et al., 2009). In *Arabidopsis thaliana*, arginine tRNA adenosine deaminase (TAD or ADAT) performs this deamination (Elias and Huang, 2005; Delannoy et al., 2009).

RNA editing in coding sequences increases the conservation levels among proteins across several plants species. Evolutionarily, codons generated by RNA editing are more conserved than codons encoded by genomic DNA (Guo et al., 2015). Editing sites located within coding sequences have been well studied, despite the existence of editing sites in non-coding regions, such as introns and tRNAs. There are several cases of different editing efficiencies from plant to plant, and even among different plant tissues (Peeters and Hanson, 2002; Chateigner-Boutin and Hanson, 2003; Tseng et al., 2013), suggesting that several different RNA editing sites remain to be elucidated.

The identification of all components from the RNA editing machinery has not yet been achieved, although several proteins have been identified as important for the maintenance of editing processes. The pentatricopeptide repeat proteins (PPR) are a highly diverse protein family. In the plant evolutionary landscape of PPR proteins, 109 genomes/proteomes were analyzed, resulting in a total of 49,204 PPR genes and 616,206 motifs (Cheng et al., 2016). Some of these PPRs harbor a DYW motif, similar to the deaminase motifs observed in other proteins, which could explain the C-to-U nucleotide conversion (Salone et al., 2007; Schallenberg-Rüdinger et al., 2013; Hayes et al., 2015). In addition, several studies have reported PPRs associated with specific RNA editing events, demonstrating that these molecules bind to specific *cis*-elements located upstream of the RNA editing site (Okuda et al., 2006; Barkan and Small, 2014). Moreover, the PPR alone is not sufficient to promote RNA editing but requires other proteins, such as RNA editing-interacting (RIP/MORF), OMMR and OZ proteins, to achieve a successful editing event (Bentolila et al., 2013; Sun et al., 2016).

The most frequent plastid RNA editing type in flowering plants is the C-to-U change, with approximately 40 sites detected thus far in Arabidopsis (Takenaka et al., 2013). To facilitate RNA editing site prediction in organelles, software, such as PREP suite has been developed (Mower, 2009). These programs enable RNA editing site prediction in genes from organelles by considering homology and conservation among protein sequences compared to genomic databases. Currently, thousands of partial and complete plastid genomes are available in NCBI, which can be used to extensively search for RNA editing events.

Different experimental techniques have identified chloroplast RNA editing sites. A widely used method is the reverse transcription PCR (RT-PCR) of plastid messenger RNAs in which several chloroplast cDNA fragments are cloned into vectors and further sequenced (Rüdinger et al., 2009). Additionally, if a chloroplast candidate gene sequence is previously known, then specific primers can be designed to direct the gene amplification from cDNA samples, with subsequent sequencing (Wolf et al., 2004). RNA editing events can also be detected through the Poisoned Primer Extension method or High Resolution Melting (HRM) analysis (Chateigner-Boutin and Small, 2007), using chloroplast cDNA as a template for amplification. Another method to measure RNA editing is multiplex RT-PCR mass spectrometry, described as a robust and convenient method (Germain et al., 2015). Although robust, these methods are dependent on specific primers and are restricted to RNA editing studies only.

RNA sequencing has facilitated RNA editing analyses by comparing reads from RNA-seq data with organelle genome references. Currently, RNA-seq is primarily adapted to study polyadenylated transcripts. Thus, as their cyanobacterial ancestor, several plastid polyadenylated RNA transcripts are associated with the RNA decay pathway *via* degradation by 3′– 5′ exoribonucleases (Komine et al., 2002; Zimmer et al., 2009). Therefore, this approach generates RNA-seq libraries with smaller amounts of plastid reads than libraries generated from organelle-enriched RNA samples, with posterior reduction of ribosomal RNA (Guo et al., 2015). Furthermore, these approaches restrict the analysis to only transcripts located in chloroplasts, preventing a comparative analysis between nuclear and plastid transcripts.

In recent years, studies of small RNAs (sRNA) have considerably increased, particularly associated with the deep sequencing of microRNAs (miRNAs) and other small non-coding RNAs (ncRNAs) from nuclear origin, producing a large amount of new sequence data. These studies have focused on the roles of sRNAs in genome maintenance, development and plant responses to environmental stresses (Simon et al., 2009; Long et al., 2015; Xu et al., 2015). However, plastid-derived sRNA sequences have also been identified in these total sRNA libraries (Ruwe and Schmitz-Linneweber, 2012; Zhelyazkova et al., 2012; Ruwe et al., 2016). Therefore, considerable amounts of sRNA data are available in public databases and can be employed for RNA editing studies. In the present study, we propose that sRNA sequencing data could represent an additional resource to identify chloroplast RNA editing events, in addition to other approaches, such as strand-specific RNA sequencing and Single Nucleotide Polymorphism (SNP). Here, we describe a method for identifying a set of new editing sites in chloroplast transcripts using sRNA data. Analyses of sRNA libraries can provide a strong qualitative and reliable quantitative measure of plastid RNA editing events.

MATERIALS AND METHODS

sRNA Libraries and Chloroplast Genomes

Public RNA libraries deposited in NCBI GEO (www.ncbi.nlm.nih.gov/geo/) with accession numbers GSE85070 (Wu et al., 2016) (*Arabidopsis thaliana*, mRNA-seq and sRNA-seq), GSE69571 (da Fonseca et al., 2016) (*Glycine max*, soybean, mRNA-seq and sRNA-seq) and GSE77046 (Neto et al., 2015) (*Oryza sativa* japonica group, rice, sRNA-seq; mRNA-seq data unpublished) were used as input data to evaluate the proposed method. These libraries were produced from samples with no qualitative influence on RNA editing and did not use any method to enrich the isolation of plastid RNAs. The Arabidopsis mutant data present in the libraries were not used. For sRNA analyses, only reads with 18–24 nucleotides were selected from the libraries. Complete chloroplast genome, coding sequences and tRNAs from Arabidopsis (NC_000932), soybean (NC_007942), and rice (NC_001320) were obtained separately at the Index

of Genomes from The CpBase: Chloroplast Genome Database (http://chloroplast.ocean.washington.edu/).

Prediction of Conserved Editing Sites

The Predictive RNA Editor for Plants suite (PREP-Cp) (http://prep.unl.edu/) (Mower, 2009) was used to predict conserved plastid editing sites. These sites were used to evaluate read coverage and editing percentage using the sRNA data. Fasta files corresponding to plastid coding sequence data were manually formatted to be usedfor use as an input batch file in the PREP-Cp tool. To predict editing sites for each species, a less stringent cutoff value of 0.5 was used, despite the 0.8 default value. This lower cutoff value was used to evaluate the effective occurrence of the predicted editing sites and their efficacious detection from sRNA data.

RNA Mapping and Confirmation of Predicted Sites

The sRNA/mRNA libraries were primarily mapped using Bowtie (Langmead et al., 2009) with 0 mismatch and no reverse complement against the chloroplast genome, coding sequences and tRNAs. Mapped reads resulted in a new file (m0). Unmapped reads were submitted to a second round of mapping with no mismatches against nuclear and mitochondrial genomes. This step eliminates all reads with perfect matches against these genomes. Unmapped reads were further mapped with two mismatches and no reverse complement against chloroplast genome and coding sequences. This second group of mapped reads produced another file containing reads with editing events (m2). Both m0 and m2 fastq files were concatenated in an m0 + m2 file. The C-to-U editing sites predicted by PREP-Cp in the cpDNA coding sequence were subjected to m0 + m2 mapping and further manual inspection using Tablet software (Milne et al., 2013). The predicted editing sites were confirmed based on a C-to-T mapping change. The steps described above are summarized in **Figure 1**.

Single Nucleotide Polymorphism Analysis

The m0 + m2 fastq files from sRNA libraries were mapped against the whole chloroplast genome, coding sequences and tRNAs using Geneious-R8 (Kearse et al., 2012), with the Bowtie algorithm and the same parameters of the previous mapping (**Figure 1**). The Geneious find variation/SNPs tool was used to search for A-to-G and C-to-T changes in putative new editing sites that were not predicted by PREP. The following parameters were used: Minimum Coverage of 5, Maximum Variant *P*-value of 10^{-2}, option to find polymorphism Inside and Outside coding sequence and *P*-value calculation method as approximate. In the manual inspection of mapping, reads with putative editing events in the 5′ and 3′ end were discarded to improve prediction and selection for validation using RT-qPCR assay.

Validation and Analysis of the RNA Editing Sites Using RT-qPCR

To validate predicted and new C-to-U RNA editing sites from the sRNA data in soybean chloroplast transcripts [*Glycine max* (L.) Merrill], we collected the roots, leaves and petals from the soybean cultivar Conquista. These tissues were collected as biological triplicates. All samples were immediately frozen in liquid nitrogen, and total RNA was extracted using Trizol (Invitrogen, CA, USA). The RNA quality was evaluated through electrophoresis on a 1% agarose gel, and the RNA amount was verified using a Qubit fluorometer and Quant-iT RNA assay kit according to the manufacturer's instructions (Invitrogen, CA, USA).

Reverse transcription quantitative polymerase chain reaction (RT-qPCR) was performed to validate the C-to-U RNA editing rates for some predicted editing sites in soybean chloroplast genes across three different tissues (roots, leaves and petals). To validate and quantify new RNA editing sites, only leaf

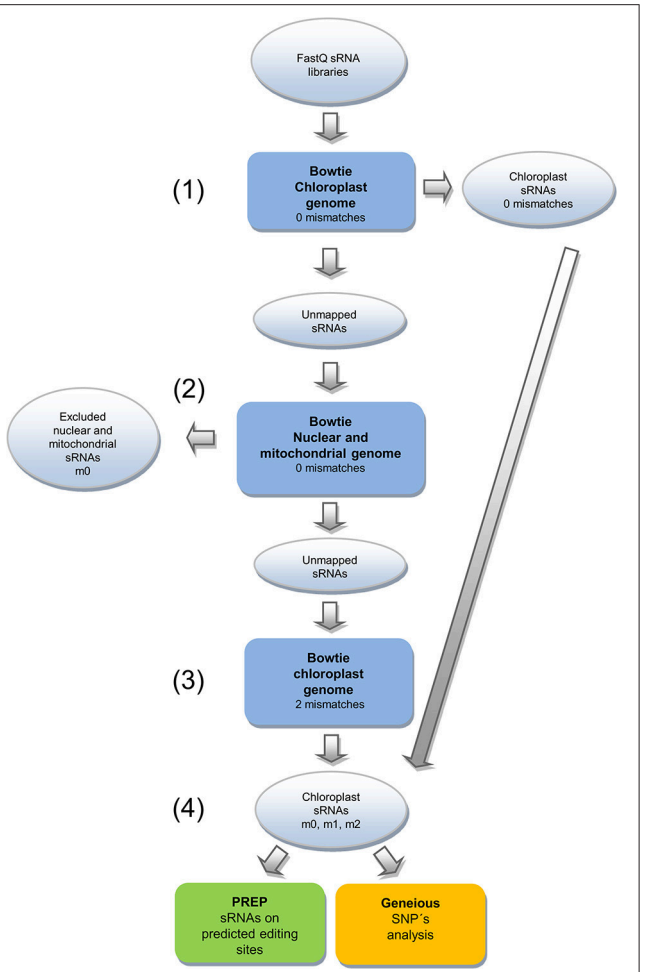

FIGURE 1 | Pipeline for identification of editing sites using chloroplast RNA transcripts. (1) sRNA-seq/mRNA-seq reads were filtered by mapping against the chloroplast reference genome. Mapped reads were saved as another file named as m0 (chloroplast RNAs m0). (2) Reads that did not map were subjected to a new round of mapping against nuclear and mitochondrial reference genomes, and those reads that did map were discarded. (3) The remaining unmapped reads were remapped against the chloroplast genome allowing up to 2 mismatches using Bowtie. (4) The resulting mapped reads (chloroplast m0 + m2), plus the m0 file, were used in the analysis to predict transcript editing sites through PREP and Geneious SNPs approaches.

samples were used. The cDNA synthesis was performed with approximately 1 μg of total RNA. Each reaction was primed with 1 μM dT25V oligonucleotide (Invitrogen, Carlsbad, CA, USA). Prior to transcription, RNA and the oligo(dT)25V primer oligo were mixed with RNase-free water to a total volume of 10 μL and incubated at 70°C for 5 min, followed by cooling on ice. The reactions were reverse transcribed with 1X M-MLV RT buffer, 0.5 mM dNTPs (Ludwig, Porto Alegre, RS, Brazil) and 200 U of M-MLV RT Enzyme (Promega, Madison, WI, USA) in a final volume of 30 μL. The synthesis was performed at 40°C for 60 min. All cDNA samples were diluted 100-fold with RNase-free water and subsequently used as templates in RT-qPCR analysis. The subsequent PCR amplification was performed using a set of primers designed according to Chen et al. (2008), with modifications. A set of primers, comprising two specific editing primers and one unique universal primer, were designed for each editing site. Specific editing primers were characterized by a unique difference in the last nucleotide at the 3′ end that recognizes and differentiates edited and unedited sites. All primers employed in the reaction are listed in Table S1.

All RT-qPCR reactions were performed on a Bio-Rad CFX384 real-time PCR detection system (Bio-Rad, Hercules, CA, USA) using SYBR Green I (Invitrogen, Carlsbad, CA, USA) to detect double-stranded cDNA synthesis. The reactions were conducted in a 10 μL volume containing 5 μL of diluted cDNA (1:100), 0.2X SYBR Green I, 0.1 mM dNTP, 1X PCR buffer, 3 mM $MgCl_2$, 0.25 U Platinum Taq DNA Polymerase (Invitrogen, Carlsbad, CA, USA) and 200 nM of each forward and reverse primer. The samples were analyzed as biological triplicates and technical quadruplicates in a 384-well plate. A non-template control was also included. The PCR reactions were run under the following conditions: an initial polymerase hot start at 94°C for 5 min, followed by 40 cycles at 94°C for 15 s, 60°C for 15 s and 72°C for 10 s. A melting curve analysis was programmed at the end of the PCR run over the range of 65 to 99°C, and the temperature increased stepwise by 0.5°C. The threshold and baseline were manually determined using Bio-Rad CFX manager software.

To calculate the RNA editing rates, we used the threshold cycle (Ct) generated during the qPCR amplifications. To calculate the percentage of editing, an equation that considered the difference between the Ct-values of each editing variant was used:

$$\%\ RNA\ editing\ = \frac{2^{(Ct\ mean\ of\ T\ variant\ -\ Ct\ mean\ of\ C\ variant)}}{2^{(Ct\ mean\ of\ T\ variant\ -\ Ct\ mean\ of\ C\ variant)}+1} \times 100$$

RESULTS

sRNA Reads Mapped to Chloroplast Genomes

The sRNA libraries sequenced without plastid RNA isolation were mapped to Arabidopsis, soybean and rice chloroplast genomes using an in-house pipeline (**Figure 1**). Approximately 3.2, 1.6, and 0.9 million reads did not map to nuclear and mitochondrial genomes but mapped to Arabidopsis, soybean and rice chloroplast genomes, respectively. These chloroplast (cp)-mapped reads represented approximately 22.9% (Arabidopsis), 4.79% (soybean), and 3.62% (rice) of the total reads in these libraries (**Table 1**). The editing informative m2 reads corresponded to 455,904 (Arabidopsis), 208,417 (soybean), and 144,609 (rice). The histograms representing the percentage length distribution of each individual class are shown in **Figure S1**. The mean coverage was 838.6 in Arabidopsis, 358.6 in soybean and 222 in rice. The maximum coverage values were 872,674 in Arabidopsis, 380,116 in soybean and 166,534 in rice. Some chloroplast regions were not covered by the sRNA library reads, with minimal coverage of zero. The number of plastid genome positions with no coverage was 47,057 in Arabidopsis, 24,505 in soybean and 3,039 in rice, representing approximately 30.46, 16.09, and 2.25% of each chloroplast genome, respectively. The genome fraction coverage for Arabidopsis, soybean and rice is represented in **Figure S2**.

sRNA Polymorphisms Confirm PREP Editing Site Prediction in Coding-Sequence Genes

The conserved chloroplast C-to-U RNA editing sites were predicted using the Predictive RNA Editor for Plants (PREP-Cp) (http://prep.unl.edu/) (Mower, 2009). The PREP suite predicted 69 potential editing sites in Arabidopsis, 92 sites in soybean and 79 sites in rice chloroplast genes. These predicted editing sites

TABLE 1 | Distribution of sRNA sequences among nuclear, mitochondrial and plastid genomes.

Organism	Total	Nuclear	mtDNA	cpDNA (m0)	cpDNA (m2)	cpDNA total	Not aligned
Arabidopsis	14,113,280	6,369,985	18,393	2,778,067	454,904	3,232,971	4,491,931
	100%	45.13%	0.13%	19.68%	3.22%	22.9%	31.82%
Soybean	34,313,559	28,219,467	46,399	1,438,193	208,417	1,646,610	4,401,083
	100%	82.23%	0.13%	4.19%	0.60%	4.79%	12.82%
Rice	25,247,958	21,479,400	12,003	768,437	144,609	913,046	2,843,509
	100%	85.07%	0.05%	3.04%	0.57%	3.62%	11.27%

m0, reads with no mismatches.
m2, reads with until 2 mismatches.

were distributed in 21 different coding sequences in Arabidopsis and rice and 23 coding sequences in soybean. The mapped chloroplast sRNA reads were analyzed using Tablet software to evaluate the presence/absence of C-to-U editing events in the predicted sites. Different numbers of confirmed editing sites were observed among the three species: 28 sites in Arabidopsis, 32 sites in soybean and 20 sites in rice, corresponding to 40.57, 34.78, and 25.31% of the total sites, respectively. The PREP score (values between 0 and 1) indicates editing site prediction confidence to control the relative proportion of false positive and false negative predictions. When a more stringent score value ($\geq$0.8) was considered, the predicted editing site numbers decreased to 45, 59, and 29 for Arabidopsis, soybean and rice, respectively. Analyses of chloroplast sRNA alignment confirmed the 23 predicted editing sites in Arabidopsis, 28 sites in soybean, and 14 sites in rice, corresponding to 51.1, 47.45, and 48.27% of the total predicted editing sites, respectively (**Figure 2A**). Even with a higher score value, some predicted sites were not confirmed, reflecting the absence of reads corresponding to editing or not enough coverage (Table S2). Four editing sites were conservatively predicted and confirmed among the three species. These sites corresponded to three sites inside the *ndhB* transcript and one site in the *rps14* transcript. Soybean and Arabidopsis shared 11 common editing sites in the *atpF*, *clpP*, *ndhB*, *ndhD*, *psbE*, *psbF*, *rpoB*, *rpoC1*, and *rps14* transcripts. Concerning the rice *atpF*, *clpP*, *ndhB*, *psbE*, and *psbF* genes, a thymine was already present in these editing sites. Rice shared a single editing site with Arabidopsis in the *ndhB* transcript at position 467, which in soybean corresponds to a thymine. The numbers of unique confirmed editing sites for each species were 12, 16, and 14

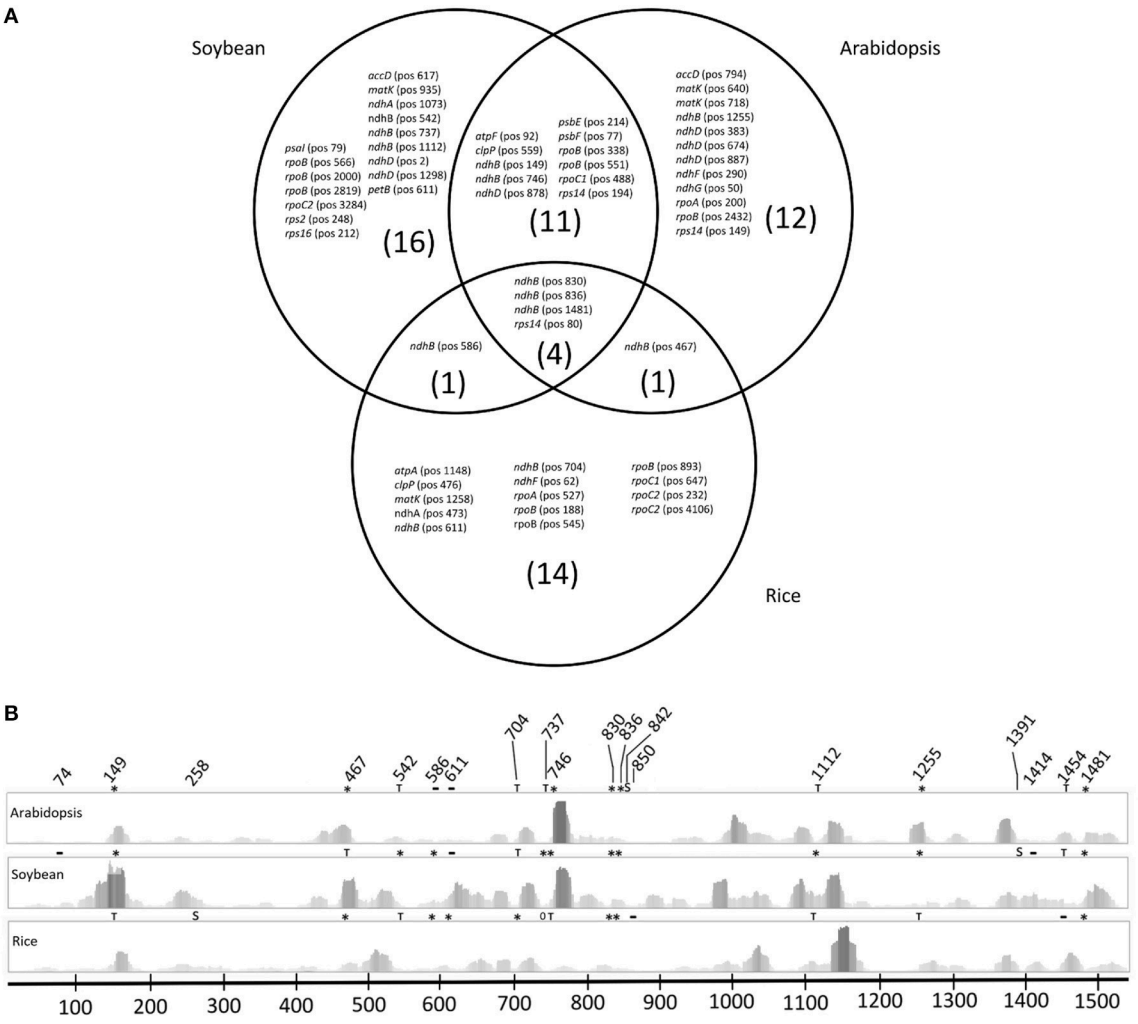

FIGURE 2 | PREP predicted editing sites and graphical read distribution and editing in the *ndhB* transcript. **(A)** Venn diagram with confirmed RNA editing sites predicted by PREP in Arabidopsis, soybean and rice. Gene names followed by the position numbering of the editing site in the coding sequence are indicated. **(B)** Graphical representation of sRNA coverage and predicted editing sites in the *ndhB* gene; (S) editing sites identified by SNP analysis, (T) predicted editing site in another species that already has a thymine in the species, (*) editing site predicted by PREP and confirmed by read mapping and coverage, (–) predicted sites with reads but not confirmed by editing and (0) predicted editing sites without read coverage.

for Arabidopsis, soybean and rice, respectively (**Figure 2A**). The complete distribution of PREP predicted editing sites according to species is described in Table S2.

mRNA-Seq and sRNA-Seq Differences in RNA Editing Analysis

To provide information concerning sRNA data reliability, the C-to-U RNA editing profiles were compared to the PREP predicted editing sites between the sRNA and mRNA (messenger RNA) libraries in Arabidopsis, soybean and rice. The mRNA-Seq data confirmed 27 predicted editing sites in Arabidopsis, 37 sites in soybean and 20 sites in rice, corresponding to 39.13, 40.21, and 25.31% of the predicted sites, respectively (Table S3). One predicted editing site was exclusively confirmed using mRNA-Seq libraries in Arabidopsis, and 11 predicted editing sites were confirmed in soybean and rice. However, analyses using sRNA-Seq libraries detected two exclusively confirmed editing sites in Arabidopsis, six sites in soybean and eight sites in rice. The confirmed predicted editing sites shared between mRNA and sRNA data corresponded to 37.68, 28.26, and 15.19% of the total predicted editing sites in Arabidopsis, soybean and rice, respectively (**Figure 3**).

Confirmation of PREP Predicted Editing Sites and New Editing Site Prediction through SNP Analysis in Coding-Sequences Using sRNA Data

In addition to the confirmation of the predicted editing sites, new candidates for editing sites were searched. A SNP analysis was used with a minimum *P*-value of $\leq 10^{-10}$ to identify sites with C-to-T changes. This parameter enabled the identification of 59 potential editing sites in Arabidopsis, 43 sites in soybean, and 19 sites in rice. Among these editing sites, 58, 37, and 15 sites encode amino acid changes in Arabidopsis, soybean and rice, respectively (Table S4). These editing sites were distributed in 27 genes in Arabidopsis, 24 genes in soybean and 11 genes in rice. Comparison of these editing sites against the editing sites predicted using PREP revealed that 20, 18, and 7 sites were previously predicted in Arabidopsis, soybean and rice, respectively (Table S5). Among these sites, 18, 18, and 6 sites were predicted with a higher score value in Arabidopsis, soybean and rice, respectively.

When the edited transcript distribution was evaluated in all species (**Figure 4A**), a higher editing frequency was associated with a core of genes (*clpP*, *ndhB*, *ndhF*, *rpoA*, *rpoB*, *rpoC1*, *rpoC2*, and *rps14)* and confirmed with at least one method used for all species evaluated. Considering exclusive edited genes, Arabidopsis showed 14 editing sites distributed among nine genes identified using SNP analysis. The editing in the rice *atpA* gene, detected through SNP analysis, was predicted by PREP. Soybean presented four exclusive editing sites confirmed by sRNA reads and predicted by PREP. They sites were distributed among the *petB*, *rps2*, and *rps14* genes. C-to-U changes promote a serine to leucine amino acid change in *petB* and *rps14* and

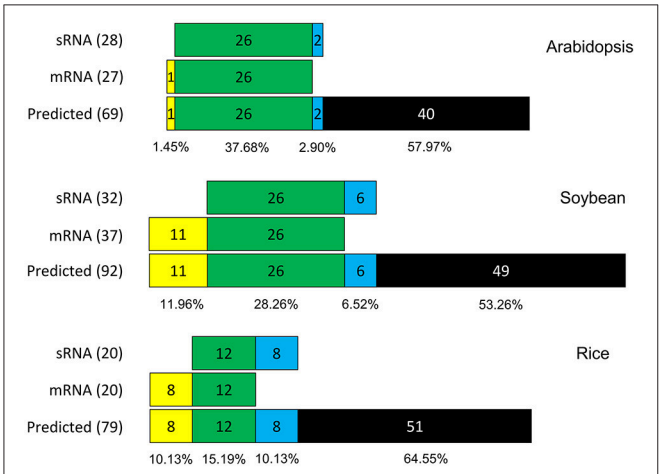

FIGURE 3 | Comparison of predicted editing site confirmation between sRNA and mRNA data. On the left, values of total confirmed predicted editing sites by data type (mRNA or sRNA). Green boxes represent editing sites confirmed in both data; yellow boxes represent editing sites confirmed only in mRNA data; blue boxes represent editing sites confirmed only in sRNA data; and black boxes represent unconfirmed predicted editing sites.

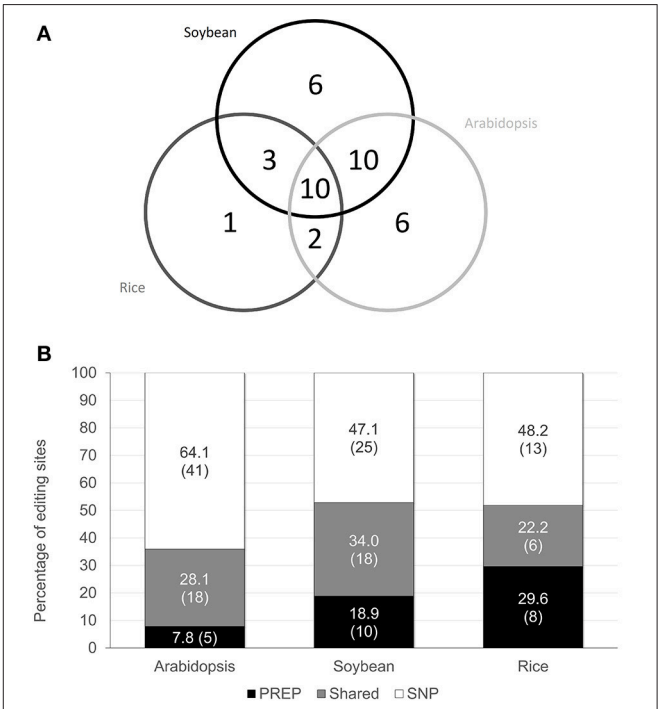

FIGURE 4 | Number of genes with C-to-U editing sites in the studied species. **(A)** Venn diagram with the total number of genes with editing sites in Arabidopsis, soybean and rice, when using both PREP (only confirmed) and SNP analysis. Not all genes share common editing sites among species. The gene identities are described in Table S6. **(B)** Percentages of total RNA editing sites identified by distinct approaches, as observed in Arabidopsis, soybean and rice. The absolute number of editing sites for each method is in parentheses. Black bars correspond to the percentage of total sites confirmed only by PREP prediction (>0.8 in prediction score); white bars indicate the percentage of total sites confirmed by the SNP approach; and gray bars show the percentage of total sites confirmed using both approaches.

a histidine to tyrosine amino acid change in *rps2*. Arabidopsis, soybean and rice SNP analysis revealed 19, 15, and 7 C-to-T changes distributed among 11, 10 and five exclusive genes, respectively. All genes and their respective editing sites are listed in Table S6. The comparative C-to-T analysis using different identification methods demonstrated that the SNP method could identify reliable C-to-U editing events, including events previously predicted using PREP at a lower PREP score (> 0.5) (**Figure S3**) or a more stringent cutoff (PREP score >0.8) (**Figure 4B**).

C-to-U RNA Editing in the ndhB Gene

The well-studied *ndhB* gene was the most frequently edited gene detected through PREP prediction in all plants. The number of editing sites predicted by PREP in this gene varied between species: 9 sites in Arabidopsis, 13 in soybean and 10 in rice. The number of editing sites confirmed by sRNA alignment was 7 sites in Arabidopsis, 9 sites in soybean and 7 sites in rice, representing 77.7, 69.23, and 70% of the predicted editing sites, respectively. Other editing sites could not be confirmed, reflecting insufficient read coverage (**Table 2**). In contrast, despite high predicted editing site numbers, 7 sites in Arabidopsis, 9 sites in soybean and 5 sites in rice, the *matK* gene had only two confirmed predicted editing sites in Arabidopsis and one confirmed predicted editing site in soybean and rice (Table S2).

In the *ndhB* gene, SNP analysis detected potential new editing sites in all three species (**Table 2**). However, this gene was not the most edited gene according to SNP analysis in rice. In this species, *ndhB* had three new potential editing sites, while *rpoC2* gene had four new sites. In Arabidopsis, *ndhD* had 8 new potential editing sites according to SNP analysis. In soybean, the *ndhB* gene remained as the most edited gene (Table S6). Comparative analyses showed a different read distribution of the predicted sites in *ndhB* among species (**Figure 2B**). Some regions showed higher coverage, not only in the editing site, but also in neighboring sites. For example, PREP predicted 467 editing sites (C-to-U), with varied coverage between species, but reads confirming the editing event were observed in both Arabidopsis and rice. Although soybean had a higher amount of reads in this site, a T was present in this genomic position. Notably, several sites showed more than 10 reads of coverage but did not confirm editing events. Some putative editing sites predicted using SNP analysis showed higher coverage than the predicted sites confirmed using PREP (**Table 2**).

A-to-I Editing Events Predicted Using SNP Analysis in Chloroplast tRNA Genes

Chloroplast sRNAs can also be useful in adenosine to inosine (A-to-I) RNA editing screening. tRNA genes were used to evaluate editing events, by searching for a guanosine (G) SNP in sRNA mapping since inosine is read as G by cellular machineries (Kim, 2004).

tRNA genes showed at least one position with an A-to-G change in at least two species (Table S7), totaling 11, 4, and 12 putative A-to-I editing events in Arabidopsis, soybean and rice, respectively. These A-to-G changes were distributed in 8, 4, and 10 tRNAs in Arabidopsis, soybean and rice, respectively. Among these sites, two sites were conserved between species: position 58 of tRNA-Trp (CCA) between soybean and rice and position 35 of tRNA-Arg (ACG) among all species evaluated. In tRNA-Arg (ACG), nucleotide 35 presented 40, 58.8, and 67.8% of the edited reads in Arabidopsis, soybean and rice, respectively (**Table 3**). The tRNAs most frequently edited were tRNA-Ser (UGA), with 3 A-to-G changes in Arabidopsis, and tRNA-Leu (UAG) and tRNA-Trp (CCA) with two A-to-G changes in Arabidopsis and rice, respectively.

Validation of C-to-U RNA Editing in Soybean Plastid Genes

To validate some predicted editing sites and demonstrate sRNA data reliability as a resourceful tool for the identification of RNA editing sites, four PREP predicted editing sites were selected for C-to-U RNA editing analysis using RT-qPCR. The *ndhA* (position 1073), *ndhB* (position 149), *rps14* (position 80), and *rps16* (position 212) editing sites were comparatively quantified in different soybean tissues (**Figures 5A–D**). Five new putative editing sites, identified by SNP analysis, were also confirmed and quantified in leaf samples: *accD* (position 617), *ndhE* (position 233), *petB* (position 611), *rps2* (position 248), and *rps3* (position 383) (**Figure 5E**). RT-qPCR showed that the percentage of *ndhA* editing was higher in leaves (76.75%) than in petals (20.11%) or roots (30.23%) (**Figure 5A**). The same editing pattern was observed for *ndhB* and *rps14*. In *ndhB,* the percentage editing was 72.41, 30.54, and 16.55% (**Figure 5B**), while values of 74, 17.86, and 8.15% were obtained in *rps14* editing in the leaves, petals and roots, respectively (**Figure 5C**). The *rps16* editing profile was different, with an editing percentage that was higher than 60% in all tissues (**Figure 5D**). With respect to putative new C-to-U editing sites identified using SNP analysis, RT-qPCR confirmed C-to-U editing events and demonstrated different editing rates among genes: *accD* (60.2%), *ndhE* (39.85%,) *petB* (54.3%), *rps2* (71.52%), and *rps3* (20.02%) (**Figure 5E**).

DISCUSSION

In the present study, we propose an additional resource and new method to identify conserved and new RNA editing sites in plastid RNA sequences. Currently, an increasing number of high-throughput sequencing data have become available. Among these datasets, there are substantial data corresponding to sRNA sequencing libraries. After analyzing some of these libraries, we observed that even without previous isolation of chloroplasts for further RNA extraction and sequencing, millions of chloroplast-derived sRNA reads could be recovered, reflecting mapping against the chloroplast genome. An important constraint of the presented method refers to the library quality and the read coverage of reference genomes.

In the present study, Arabidopsis libraries had the highest mean coverage using sRNA reads, which likely facilitated the recovery of the largest number of confirmed editing sites. The coverage percentage across genomes was different between species, with lower values detected in Arabidopsis. This result

TABLE 2 | *NdhB* C-to-U editing events by PREP and SNP approach using reads derived from sRNA-seq.

Organism	Codon change	Nucleotide position	AA change	AA position	Total coverage	Edited coverage	% Editing	SNP *P*-value	PREP score
Arabidopsis	TCA–TTA	149	S–L	50	40	32	80	4.8E-108	1
(1,539: 870)*	CCA–CTA	467	P–L	156	40	28	75	8.5E-109	1
	CAT–TAT	586	H–Y	196	1	0	no editing	-	1
	TCA–TTA	611	S–L	204	5	0	no editing	-	0.8
	TCT–TTT	746	S–F	249	12	5	41.7	5.3E-109	1
	TCA–TTA	830	S–L	277	20	9	45	8.4E-29	1
	TCA–TTA	836	S–L	279	21	10	47.6	1.1E-23	1
	GCC–GTC	842	T–I	281	19	2	10.5	1.3E-8	nd
	CAT–TAT	1,255	H–Y	419	47	47	100	nd	1
	CCA–CTA	1,481	P–L	494	34	14	41.2	5.5E-40	1
Soybean	CCT–CTT	74	P–L	25	4	0	no editing	nd	1
(1,533: 543)*	TCA–TTA	149	S–L	50	35	10	28	3.3E-11	1
	ACG–ATG	542	T–M	181	1	1	100	nd	1
	CAT–TAT	586	H–Y	196	11	2	18.2	0.0000038	1
	TCA–TTA	611	S–L	204	14	0	no editing	nd	0.8
	CCA–CTA	737	P–L	246	2	2	100	nd	1
	TCT–TTT	746	S–F	249	12	4	33.3	2.0E-14	1
	TCA–TTA	830	S–L	277	12	5	41.7	3E-17	1
	TCA–TTA	836	S–L	279	11	5	45.5	2.6E-15	1
	TCA–TTA	1,112	S–L	371	22	5	22.7	4.3E-17	1
	CAT–TAT	1,255	H–Y	419	1	0	no editing	nd	1
	CCT–CTT	1,391	P–L	464	9	2	22.7	0.0000036	
	CCC–TCC	1,414	P–S	472	10	0	no editing	nd	1
	CCA–CTA	1,481	P–L	494	13	8	64.3	1.3E-31	1
Rice	AGC–AGT	258	S–S	86	8	2	25	1.6E-8	nd
(1,533: 619)*	CCA–CTA	467	P–L	156	14	9	64.3	1.30E-31	1
	CAT–TAT	586	H–Y	196	5	3	60	4.00E-12	1
	TCA–TTA	611	S–L	204	2	1	50	nd	0.8
	TCC–TTC	704	S–F	235	16	3	18.8	7.10E-08	1
	CCA–CTA	737	P–L	246	0	0	nd	nd	1
	TCA–TTA	830	S–L	277	3	1	33	nd	1
	TCA–TTA	836	S–L	279	4	1	25	nd	1
	CTC–TTC	850	L–F	284	2	0	no editing	nd	0.6
	ACT–ATT	1,454	T–I	485	30	0	no editing	nd	0.6
	CCA–CTA	1,481	P–L	494	6	5	83	8.0E-17	1

*****Coding sequence length and coverage values.*

"Nucleotide position": position in base pair is from the A of the initiator codon.

"Total Coverage": total mapped reads in respective nucleotide position.

"Edited Coverage": number of reads shown T, instead C.

"% Editing": percentage of RNA editing using the edited reads divided by total mapped reads.

"PREP score": confidence value of prediction according PREP.

"nd": no defined.

demonstrated that the use of sRNA libraries for mapping editing events is not directly related to a significant coverage across the entire plastid genome. Although this method has the capacity to confirm and discover editing sites in chloroplasts, a smaller number of mitochondrial reads would likely affect RNA editing analysis in this organelle. In the present study, the approach for the identification of editing sites was compared to the PREP and SNP strategies. The editing sites and percentage editing may vary between species because some species may already possess a thymine in the genome. In these cases, C-to-U editing will not occur. The same situation can occur with some A-to-I editing sites, which could affect the general percentage of editing among species. The use of a different PREP score, resulting in distinct cut-off values, may also affect these percentages. In addition, editing factors and their editing sites may evolve differently among species.

The elementary step employed in the pipeline used in the present study was the initial sRNA library mapping against the chloroplast genome, considering 0 mismatches. Plastid DNA insertions in nuclear genomes have been demonstrated for partial, intact or even truncated coding sequences in several species (Chen et al., 2015). Thus, an initial filtration step against

TABLE 3 | A-to-I editing analysis of tRNA-Arg(ACG) sites by SNP approach with corresponding reads derived from sRNA-seq.

Organism	Nucleotide position	Nucleotide change	Total coverage	Edited coverage	% Editing	Variant *P*-value
Arabidopsis (74: 3,015)*	35	A–G	80	32	40	3.8E-655
Soybean (74: 65,787)*	35	A–G	80	47	58.5	2.3E-144
Rice (74: 1,673)*	35	A–G	214	145	67.8	1.5E-465

**tRNA sequence length and coverage values.*
"Total Coverage": total mapped reads in respective nucleotide position.
"Edited Coverage": number of reads shown G, instead A.
"% Editing": percentage of RNA editing using the edited reads divided by total mapped reads.

the chloroplast genome prevents the loss of unedited reads to those loci present in nuclear insertions. Unedited reads are necessary, particularly in quantitative editing analysis, where the editing percentage is measured and cannot be ruled out.

Some C-to-U editing studies have previously used mRNA-Seq to demonstrate and quantify editing events in plant mitochondria (Bentolila et al., 2013) and chloroplasts (Guo et al., 2015). Comparison of sRNAs and mRNA data sequences demonstrated that most of the confirmed editing sites can be recovered using both datasets. However, there are differences between these data, demonstrating that sRNAs can identify editing sites that were not detected using mRNA data and vice versa (**Figure 3**). The use of sRNA data to complement RNA editing analysis can improve the identification and measurement of RNA editing in various aspects.

In the present study, a new set of plastid editing sites was identified in soybean. The C-to-U editing events have previously been demonstrated in other species, and we recovered several edited transcripts, including *ndhB*, *ndhD*, *ndhG*, *rpoB*, and *rpoC1* (Corneille et al., 2000; Okuda et al., 2009; Zhou et al., 2009; Chateigner-Boutin et al., 2011; Boussardon et al., 2012; Tseng et al., 2013), in the present analysis. For most known C-to-U editing sites predicted through PREP and confirmed by sRNA reads in the present study, 21 sites have previously been demonstrated in Arabidopsis (Tsudzuki et al., 2001; Tillich et al., 2005) and 19 sites have previously been demonstrated in rice (Corneille et al., 2000; Tsudzuki et al., 2001), representing 30.43 and 24% of the total predicted editing sites, respectively (Table S2). Moreover, we showed editing events in soybean plastid genes, including *ndhA*, *psaI,* and *petB*, which had not previously been demonstrated for rice or Arabidopsis. In the SNP analysis, we identified new C-to-U editing sites. For example, in the Arabidopsis *ndhF* gene, a putative C-to-U editing site was identified at position 884, leading to a serine to phenylalanine change. In the soybean *ndhE* gene, a putative C-to-U editing site at position 233 was observed in 73.7% of the reads. This editing led to a proline to leucine change in the encoded protein. Despite this information, the impact of amino acid modifications on respective protein structures remains unclear. Both *ndh* genes encode thylakoid Ndh complex components involved in photosynthesis optimization under different stress conditions conditions (Casano, 2001; Martin et al., 2004; Rumeau et al., 2007). *NdhB* mutants under lower air humidity conditions or following exposure to ABA present a reduction in the photosynthetic level, likely mediated through stomatal closure triggered under these conditions (Horvath, 2000). Therefore, a protein structure modification, resulting from a loss or decrease in RNA editing events could affect adaptations to stress conditions or cause other unknown changes.

The coding sequence of protein D2, encoded by the *psbD* gene, a photosystem II (PSII) core protein, showed a putative new editing event in rice at positions 1006 and 1007. However, reflecting low coverage, these new editing sites still require further experimental confirmation. Maintenance of the D2 protein structure is important not only for proton transport (Pokhrel et al., 2013) but also for the phosphorylation dynamics of this protein (Tikkanen and Aro, 2012) and its interaction with the proteins responsible for PSII maintenance (Liu and Last, 2015). If this editing site is confirmed, then alterations in editing site patterns resulting from factors, such as abiotic stress could be associated with photo-oxidative damage susceptibility. Previous studies have demonstrated that abiotic stress influences the editing process and consequently plastid physiology (Nakajima and Mulligan, 2001; Karcher and Bock, 2002).

Five putative C-to-U editing sites predicted using SNP analysis were validated through RT-qPCR. This result demonstrates the reliability and accuracy of sRNA data resources and the method presented herein to confirm predicted sites *in silico* and identify new RNA editing sites. Position 1073 in the *ndhA* gene is an editing site identified only in the soybean chloroplast editome. RT-qPCR revealed that the editing percentage varies among different soybean tissues. The *ndhB* (position 149) gene was previously evaluated in the non-photosynthetic tissues of Arabidopsis. An RNA editing pattern previously demonstrated in Arabidopsis (Tseng et al., 2013), with a higher percentage in leaves (>75% edited), followed by flowers (25–75% edited) and roots (unedited), was similarly observed in the present study. An exception was observed for the root tissue, which showed a low editing percentage (16.5%) in soybean instead of an unedited rate, as observed in Arabidopsis. The editing site at position 80 in *rps14* also was evaluated across different tissues in Arabidopsis. A high editing percentage was demonstrated in Arabidopsis leaves (Tseng et al., 2013), a pattern also demonstrated in soybean using RT-qPCR. The RNA editing percentages observed in roots

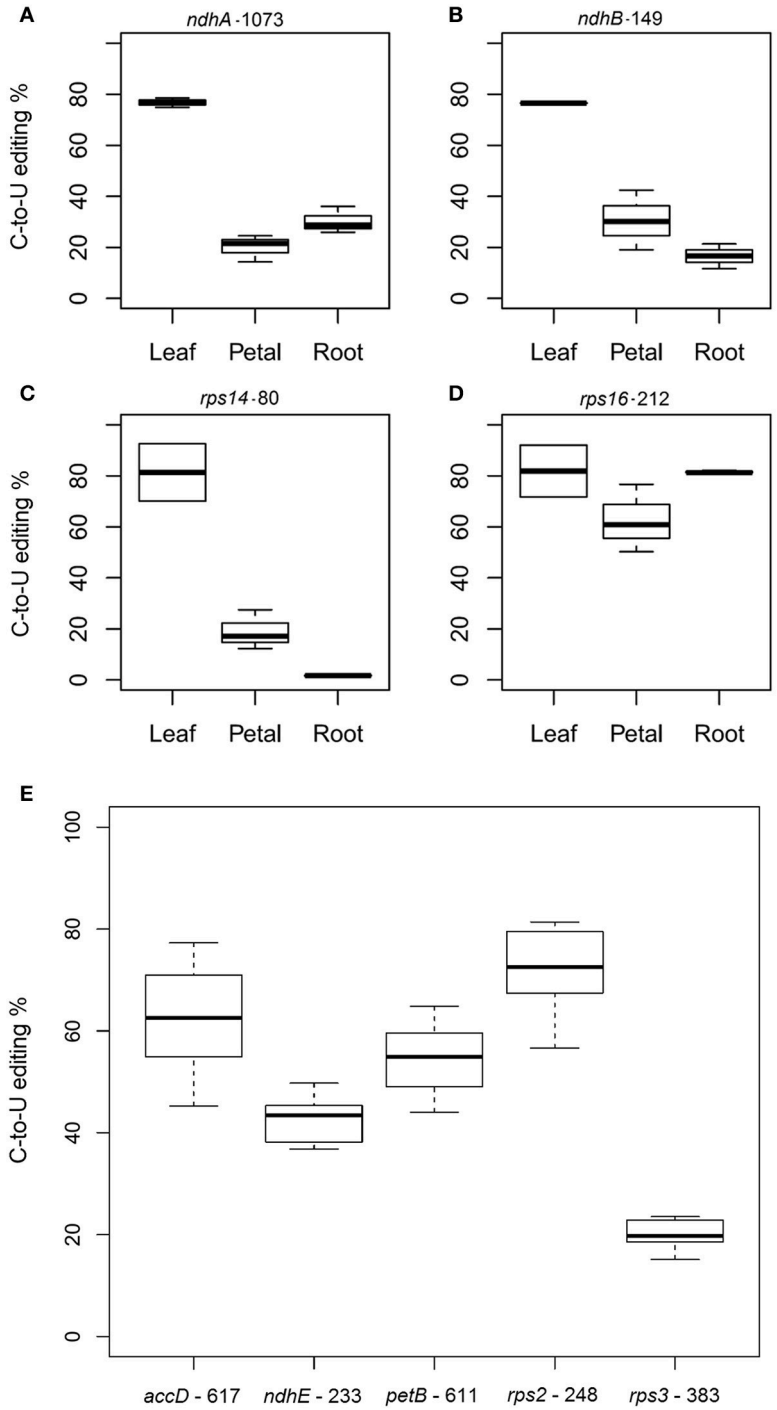

FIGURE 5 | Confirmation and quantitation of soybean editing sites predicted by PREP. **(A)** *ndhA-1053*, **(B)** *ndhB*-149, **(C)** *rps14*-80, and **(D)** *rps16*-212 were analyzed in leaves, petals and roots. Box area represents the lower and upper percentiles; **(E)** confirmation and quantitation of soybean editing sites identified by SNP analysis. Transcripts from soybean leaves were analyzed for C-to-U editing in specific nucleotide positions: *accD*-617, *ndhE*-233, *petB*-611, *rps2*-248, and *rps3*-383. Box area represents the lower and upper percentiles. The upper whisker of the boxplot indicates the highest editing value observed; the lower whisker, the lowest editing value; and the middle line, the median.

and petals showed different patterns between Arabidopsis and soybean, although a decrease in these values was observed in the root tissue of both species. The editing of *rps16* at position 212 was predicted and confirmed only in soybean and did not show differences in the editing percentage between leaf and root tissues. These results indicate that sRNA sequence mapping can not only

be used to confirm the predicted editing sites, but also to quantify the editing percentage.

The plastid acetyl-CoA carboxylase, necessary for *de novo* fatty acid synthesis, comprises two components, *accA* and *accD* proteins; *accD* encodes the β-carboxyl transferase subunit and is required in tobacco plants for a functional enzyme (Kode et al., 2005). The vanilla cream1 (vac1) albino mutant, reflecting a PPR-DYW protein required for editing in *accD* and *ndhF* in Arabidopsis, exhibits albino to pale yellow phenotype and an RNA editing reduction in those transcripts (Tseng et al., 2010). The requirement of plastid *accD* editing for functional protein has previously been demonstrated (Sasaki et al., 2001), and this new editing site, which promotes a serine to leucine change, could also be important for the maintenance of protein structure and functionality. The *ndhE* gene encodes a subunit of a membrane subcomplex of the NAD(P)H dehydrogenase complex (Peng et al., 2011). NdhE protein interacts with the membrane subcomplex proteins, NdhC and NdhG, and with subcomplex proteins, NhdH and NdhK (Efremov et al., 2010; Peng et al., 2011). The new editing site described here promotes a proline to leucine change, which could modify the interaction between these proteins and lead to changes in electron transfer to quinone. The *petB* gene encodes the cytochrome b_6 protein, a cytochrome b_6f complex component responsible for mediating electron transfer between photosystem I (PSI) and plastocyanin (Baniulis et al., 2008); mutants of *petB* in tobacco showed reduced levels of PSI, PSII and light-harvesting complex proteins (Monde et al., 2000), indicating a requirement of cytochrome b_6 to correct photosynthetic apparatus assembly. The new editing site involving a serine to leucine change in *petB* at position 611, identified in the present study, could be required for the maintenance of cytochrome b_6f complex structure and stability. Proteins S2 and S3 are located on the solvent side of ribosome small subunit (Manuell et al., 2004), and RNA editing events can modify their interactions among other ribosomal proteins and likely with mRNA, with potential effects on the regulatory aspects of plastid translation in response to stress or other homeostasis processes.

The SNP analysis facilitated the evaluation of not only C-to-U editing but also A-to-I editing events in chloroplast tRNAs. The tRNA-Arg (ACG) A-to-I editing event was also observed in all three species in the present study. This change corresponds to an inosine in the wobble position, which encodes three arginine codons CGU, CGC, and CGA that play a critical role in plastid protein synthesis (Rogalski et al., 2008). The enzyme involved in this mechanism in Arabidopsis, At1g68720, encodes a tRNA adenosine deaminase (TADA), which is targeted to plastids. RNAi lines of this gene show markedly reduced A-to-I editing efficiency, displaying phenotype consequences, such as growth and development delays (Elias and Huang, 2005; Delannoy et al., 2009; Karcher and Bock, 2009). Editing events in others tRNAs have been shown in some species and have been well studied in animals (Su and Randau, 2011) and previously demonstrated in moss *Takakia lepidozioides* (Miyata et al., 2008). The method described here can help to identify and measure other tRNA editing events not yet described in plants.

In addition to the high amount of data currently available in public databases that can readily be assessed, there are some plastid sRNAs biological features that can reveal important mechanisms of RNA editing. The precise plastid sRNA biogenesis remains unknown because there is no evidence of any RNAi machinery in organelles that could originate small RNAs thus far. Notably, there is evidence of a relaxed plastid genome transcription mechanism, resulting in full plastid genome transcription (Hotto et al., 2012). It has been suggested that plastid sRNAs originated from RNA sequence regions protected against degradation by forming secondary structures or from associations with RNA-binding proteins regions (Pfalz et al., 2009). The results of the present study demonstrated that sRNAs are not necessarily over-represented in regions of editing sites but are also evident in coding sequences with smaller lengths, where these sRNAs can still be observed. These biological features enable the use of sRNA datasets to confirm the results of different RNA editing prediction tools and enable the analysis of editing events not only in a qualitative but also a quantitative manner, depending on the library quality and read coverage.

The identification of editing sites and measurement of editing levels have demonstrated differences among tissues (Tseng et al., 2013) and developmental stages (Miyata and Sugita, 2004). These findings can be used to evaluate the impact of different stresses on these mechanisms (Nakajima and Mulligan, 2001; Van Den Bekerom et al., 2013). Thus, the use of sRNA data to confirm predicted editing sites in association with SNP searches can provide a powerful and reliable plastid editome characterization and measurement, and the results can be applied to compare editing levels in different tissues, developmental stages and physiological conditions.

CONCLUSION

Analysis of sRNA libraries can be used to identify and quantify RNA editing events. Using this source of sequence data and pipeline of analyses, we obtained, for the first time, a consistent set of non-conserved and new editing sites in soybean. We propose the use of plastid sRNA libraries as a novel source and approach to study RNA editing events. Until recently, no other studies have taken advantage of such data to screen for RNA editing sites. Thus, the results from the present study should encourage researchers to use small RNA libraries to compare RNA editing in different plants under different conditions to improve knowledge on the editing role of plastid RNA in plant biology.

AUTHOR CONTRIBUTIONS

RM, NR, and AC conceived and designed the study. NR conducted *in silico* analysis. NR and FK conducted the RT-qPCR experiments. NR and GdF analyzed the data. NR and AC drafted the manuscript. All authors have read and approved the manuscript.

FUNDING

RM is the recipient of a research fellowship 309030/2015-3, NR is the recipient of a Ph.D. fellowship, and AC and GdF are

the recipients of Post-Doctoral fellowships from CNPq. FK was sponsored by a FAPERGS/CAPES-DOCFIX (1634-2551/13-9) grant. The present study was also partially supported through a grant from INCT-MCTIC.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2017.01686/full#supplementary-material

Figure S1 | sRNA length distribution. The histograms represent the percentage of length distribution of each individual class. In black, gray and white bars, Arabidopsis, soybean, and rice read data, respectively.

Figure S2 | Number of plastid genomic sites (Y-axis) and their respective sRNA reads coverage (X-axis). In black, gray and white bars, Arabidopsis, soybean and rice read data, respectively.

Figure S3 | RNA editing site numbers identified by the PREP and SNP approaches in Arabidopsis, soybean and rice. Black bars correspond to sites confirmed only by PREP prediction (>0.5 in prediction score); white bars indicate sites confirmed using the SNP approach; and gray bars show sites confirmed using both approaches.

REFERENCES

Baniulis, D., Yamashita, E., Zhang, H., Hasan, S. S., and Cramer, W. A. (2008). Structure-function of the cytochrome b 6 f complex. *Photochem. Photobiol.* 84, 1349–1358. doi: 10.1111/j.1751-1097.2008.00444.x

Barkan, A., and Small, I. (2014). Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* 65, 415–442. doi: 10.1146/annurev-arplant-050213-040159

Manuell, A., Beligni, M. V., Yamaguchi, K., and Mayfield, S. P. (2004). Regulation of chloroplast translation: interactions of RNA elements, RNA-binding proteins and the plastid ribosome. *Biochem. Soc. Trans.* 32, 601–605. doi: 10.1042/BST0320601

Bentolila, S., Oh, J., Hanson, M. R., and Bukowski, R. (2013). Comprehensive high-resolution analysis of the role of an Arabidopsis gene family in RNA editing. *PLoS Genet.* 9:e1003584. doi: 10.1371/journal.pgen.1003584

Boussardon, C., Salone, V., Avon, A., Berthome, R., Hammani, K., Okuda, K., et al. (2012). Two interacting proteins are necessary for the editing of the NdhD-1 Site in Arabidopsis plastids. *Plant Cell* 24, 3684–3694. doi: 10.1105/tpc.112.099507

Casano, L. M. (2001). Hydrogen peroxide mediates the induction of chloroplastic Ndh complex under photooxidative stress in barley. *Plant Physiol.* 125, 1450–1458. doi: 10.1104/pp.125.3.1450

Chateigner-Boutin, A.-L., des Francs-Small, C. C., Delannoy, E., Kahlau, S., Tanz, S. K., de Longevialle, A. F., et al. (2011). OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript rpoC1. *Plant J.* 65, 532–542. doi: 10.1111/j.1365-313X.2010.04441.x

Chateigner-Boutin, A. L., and Hanson, M. R. (2003). Developmental co-variation of RNA editing extent of plastid editing sites exhibiting similar cis-elements. *Nucleic Acids Res.* 31, 2586–2594. doi: 10.1093/nar/gkg354

Chateigner-Boutin, A.-L., and Small, I. (2010). Plant RNA editing. *RNA Biol.* 7, 213–219. doi: 10.4161/rna.7.2.11343

Chen, H., Yu, Y., Chen, X., Zhang, Z., Gong, C., Li, J., et al. (2015). Plastid DNA insertions in plant nuclear genomes: the sites, abundance and ages, and a predicted promoter analysis. *Funct. Integr. Genomics* 15, 131–139. doi: 10.1007/s10142-014-0422-z

Chen, Y., Kao, S., Chou, H., Lin, W., Wong, F.-H., and Chow, W.-Y. (2008). A real-time PCR method for the quantitative analysis of RNA editing at specific sites. *Anal. Biochem.* 375, 46–52. doi: 10.1016/j.ab.2007.12.037

Cheng, S., Gutmann, B., Zhong, X., Ye, Y., Fisher, M. F., Bai, F., et al. (2016). Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J.* 85, 532–547. doi: 10.1111/tpj.13121

Corneille, S., Lutz, K., and Maliga, P. (2000). Conservation of RNA editing between rice and maize plastids: are most editing events dispensable? *Mol. Gen. Genet.* 264, 419–424. doi: 10.1007/s004380000295

da Fonseca, G. C., de Oliveira, L. F. V., de Morais, G. L., Abdelnor, R. V., Nepomuceno, A. L., Waterhouse, P. M., et al. (2016). Unusual RNA plant virus integration in the soybean genome leads to the production of small RNAs. *Plant Sci.* 246, 62–69. doi: 10.1016/j.plantsci.2016.01.011

Delannoy, E., Le Ret, M., Faivre-Nitschke, E., Estavillo, G. M., Bergdoll, M., Taylor, N. L., et al. (2009). Arabidopsis tRNA adenosine deaminase arginine edits the wobble nucleotide of chloroplast tRNAArg(ACG) and Is essential for efficient chloroplast translation. *Plant Cell* 21, 2058–2071. doi: 10.1105/tpc.109.066654

Efremov, R. G., Baradaran, R., and Sazanov, L. A (2010). The architecture of respiratory complex I. *Nature* 465, 441–445. doi: 10.1038/nature09066

Elias, Y., and Huang, R. H. (2005). Biochemical and Structural Studies of A-to-I Editing by tRNA:A34 Deaminases at the Wobble Position of Transfer RNA. *Biochemistry* 44, 12057–12065. doi: 10.1021/bi050499f

Germain, A., Hanson, M. R., and Bentolila, S. (2015). High-throughput quantification of chloroplast RNA editing extent using multiplex RT-PCR mass spectrometry. *Plant J.* 83, 546–554. doi: 10.1111/tpj.12892

Guo, W., Grewe, F., and Mower, J. P. (2015). Variable frequency of plastid RNA editing among ferns and repeated loss of uridine-to-cytidine editing from vascular plants. *PLoS ONE* 10:e0117075. doi: 10.1371/journal.pone.0117075

Hayes, M. L., Dang, K. N., Diaz, M. F., and Mulligan, R. M. (2015). A conserved glutamate residue in the C-terminal deaminase domain of pentatricopeptide repeat proteins is required for RNA editing activity. *J. Biol. Chem.* 290, 10136–10142. doi: 10.1074/jbc.M114.631630

Horvath, E. M. (2000). Targeted inactivation of the plastid ndhB gene in tobacco results in an enhanced sensitivity of photosynthesis to moderate stomatal closure. *Plant Physiol.* 123, 1337–1350. doi: 10.1104/pp.123.4.1337

Hotto, A. M., Germain, A., and Stern, D. B. (2012). Plastid non-coding RNAs: emerging candidates for gene regulation. *Trends Plant Sci.* 17, 737–744. doi: 10.1016/j.tplants.2012.08.002

Karcher, D., and Bock, R. (2002). Temperature sensitivity of RNA editing and intron splicing reactions in the plastid ndhB transcript. *Curr. Genet.* 41, 48–52. doi: 10.1007/s00294-002-0278-y

Karcher, D., and Bock, R. (2009). Identification of the chloroplast adenosine-to-inosine tRNA editing enzyme. *RNA* 15, 1251–1257. doi: 10.1261/rna.1600609

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kim, D. D. Y. (2004). Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res.* 14, 1719–1725. doi: 10.1101/gr.2855504

Kode, V., Mudd, E. A., Iamtham, S., and Day, A. (2005). The tobacco plastid accD gene is essential and is required for leaf development. *Plant J.* 44, 237–244. doi: 10.1111/j.1365-313X.2005.02533.x

Komine, Y., Kikis, E., Schuster, G., and Stern, D. (2002). Evidence for *in vivo* modulation of chloroplast RNA stability by 3'-UTR homopolymeric tails in *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 4085–4090. doi: 10.1073/pnas.052327599

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25

Liu, J., and Last, R. L. (2015). A land plant-specific thylakoid membrane protein contributes to photosystem II maintenance in *Arabidopsis thaliana*. *Plant J.* 82, 731–743. doi: 10.1111/tpj.12845

Long, R. C., Li, M. N., Kang, J. M., Zhang, T. J., Sun, Y., and Yang, Q. C. (2015). Small RNA deep sequencing identifies novel and salt-stress-regulated microRNAs from roots of *Medicago sativa* and *Medicago truncatula*. *Physiol. Plant.* 154, 13–27. doi: 10.1111/ppl.12266

Chateigner-Boutin, A. L., and Small, I. (2007). A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. *Nucleic Acids Res.* 35:e114. doi: 10.1093/nar/gkm640

Martin, M., Casano, L. M., Zapata, J. M., Guera, A., del Campo, E. M., Schmitz-Linneweber, C., et al. (2004). Role of thylakoid Ndh complex and peroxidase in the protection against photo-oxidative stress: fluorescence and enzyme

activities in wild-type and ndhF-deficient tobacco. *Physiol. Plant.* 122, 443–452. doi: 10.1111/j.1399-3054.2004.00417.x

Milne, I., Stephen, G., Bayer, M., Cock, P. J. A., Pritchard, L., Cardle, L., et al. (2013). Using tablet for visual exploration of second-generation sequencing data. *Brief. Bioinformatics* 14, 193–202. doi: 10.1093/bib/bbs012

Miyata, Y., Sugita, C., Maruyama, K., and Sugita, M. (2008). RNA editing in the anticodon of tRNA Leu (CAA) occurs before group I intron splicing in plastids of a moss *Takakia lepidozioides* S. Hatt. & Inoue. *Plant Biol.* 10, 250–255. doi: 10.1111/j.1438-8677.2007.00027.x

Miyata, Y., and Sugita, M. (2004). Tissue- and stage-specific RNA editing of rps14 transcripts in moss (*Physcomitrella patens*) chloroplasts. *J. Plant Physiol.* 161, 113–115. doi: 10.1078/0176-1617-01220

Monde, R.-A., Zito, F., Olive, J., Wollman, F.-A., and Stern, D. B. (2000). Post-transcriptional defects in tobacco chloroplast mutants lacking the cytochrome b6/f complex. *Plant J.* 21, 61–72. doi: 10.1046/j.1365-313x.2000.00653.x

Mower, J. P. (2009). The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res.* 37, W253–W259. doi: 10.1093/nar/gkp337

Nakajima, Y., and Mulligan, M. (2001). Heat stress results in incomplete C-to-U editing of maize chloroplast mRNAs and correlates with changes in chloroplast transcription rate. *Curr. Genet.* 40, 209–213. doi: 10.1007/s002940100249

Neto, L. B., Arenhart, R. A., de Oliveira, L. F. V., de Lima, J. C., Bodanese-Zanettini, M. H., Margis, R., et al. (2015). ASR5 is involved in the regulation of miRNA expression in rice. *Plant Cell Rep.* 34, 1899–1907. doi: 10.1007/s00299-015-1836-3

Okuda, K., Hammani, K., Tanz, S. K., Peng, L., Fukao, Y., Myouga, F., et al. (2009). The pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid ndhB and ndhG transcripts. *Plant J.* 61, 339–349. doi: 10.1111/j.1365-313X.2009.04059.x

Okuda, K., Nakamura, T., Sugita, M., Shimizu, T., and Shikanai, T. (2006). A Pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J. Biol. Chem.* 281, 37661–37667. doi: 10.1074/jbc.M608184200

Peeters, N. M., and Hanson, M. R. (2002). Transcript abundance supercedes editing efficiency as a factor in developmental variation of chloroplast gene expression. *RNA* 8, 497–511. doi: 10.1017/S1355838202029424

Peng, L., Yamamoto, H., and Shikanai, T. (2011). Structure and biogenesis of the chloroplast NAD(P)H dehydrogenase complex. *Biochim. Biophys. Acta Bioenerg.* 1807, 945–953. doi: 10.1016/j.bbabio.2010.10.015

Pfalz, J., Bayraktar, O. A., Prikryl, J., and Barkan, A. (2009). Site-specific binding of a PPR protein defines and stabilizes 5′ and 3′ mRNA termini in chloroplasts. *EMBO J.* 28, 2042–2052. doi: 10.1038/emboj.2009.121

Pokhrel, R., Service, R. J., Debus, R. J., and Brudvig, G. W. (2013). Mutation of lysine 317 in the D2 subunit of photosystem II alters chloride binding and proton transport. *Biochemistry* 52, 4758–4773. doi: 10.1021/bi301700u

Rogalski, M., Karcher, D., and Bock, R. (2008). Superwobbling facilitates translation with reduced tRNA sets. *Nat. Struct. Mol. Biol.* 15, 192–198. doi: 10.1038/nsmb.1370

Rüdinger, M., Funk, H. T., Rensing, S. A., Maier, U. G., and Knoop, V. (2009). RNA editing: only eleven sites are present in the *Physcomitrella patens* mitochondrial transcriptome and a universal nomenclature proposal. *Mol. Genet. Genomics* 281, 473–481. doi: 10.1007/s00438-009-0424-z

Rumeau, D., Peltier, G., and Cournac, L. (2007). Chlororespiration and cyclic electron flow around PSI during photosynthesis and plant stress response. *Plant Cell Environ.* 30, 1041–1051. doi: 10.1111/j.1365-3040.2007.01675.x

Ruwe, H., and Schmitz-Linneweber, C. (2012). Short non-coding RNA fragments accumulating in chloroplasts: footprints of RNA binding proteins? *Nucleic Acids Res.* 40, 3106–3116. doi: 10.1093/nar/gkr1138

Ruwe, H., Wang, G., Gusewski, S., and Schmitz-Linneweber, C. (2016). Systematic analysis of plant mitochondrial and chloroplast small RNAs suggests organelle-specific mRNA stabilization mechanisms. *Nucleic Acids Res.* 44, 7406–7417. doi: 10.1093/nar/gkw466

Salone, V., Rüdinger, M., Polsakiewicz, M., Hoffmann, B., Groth-Malonek, M., Szurek, B., et al. (2007). A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett.* 581, 4132–4138. doi: 10.1016/j.febslet.2007.07.075

Sasaki, Y., Kozaki, A., Ohmori, A., Iguchi, H., and Nagano, Y. (2001). Chloroplast RNA editing required for functional acetyl-CoA Carboxylase in Plants. *J. Biol. Chem.* 276, 3937–3940. doi: 10.1074/jbc.M008166200

Schallenberg-Rüdinger, M., Lenz, H., Polsakiewicz, M., Gott, J. M., and Knoop, V. (2013). A survey of PPR proteins identifies DYW domains like those of land plant RNA editing factors in diverse eukaryotes. *RNA Biol.* 10, 1549–1556. doi: 10.4161/rna.25755

Simon, S. A., Zhai, J., Nandety, R. S., McCormick, K. P., Zeng, J., Mejia, D., et al. (2009). Short-read sequencing technologies for transcriptional analyses. *Annu. Rev. Plant Biol.* 60, 305–333. doi: 10.1146/annurev.arplant.043008.092032

Su, A. A. H., and Randau, L. (2011). A-to-I and C-to-U editing within transfer RNAs. *Biochemistry* 76, 932–937. doi: 10.1134/S0006297911080098

Sun, T., Bentolila, S., and Hanson, M. R. (2016). The unexpected diversity of plant organelle RNA editosomes. *Trends Plant Sci.* 21, 962–973. doi: 10.1016/j.tplants.2016.07.005

Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B., Brennicke, A., Barbara, H., et al. (2013). RNA editing in plants and its evolution. *Annu. Rev. Genet.* 47, 335–352. doi: 10.1146/annurev-genet-111212-133519

Tikkanen, M., and Aro, E. M. (2012). Thylakoid protein phosphorylation in dynamic regulation of photosystem II in higher plants. *Biochim. Biophys. Acta Bioenerg.* 1817, 232–238. doi: 10.1016/j.bbabio.2011.05.005

Tillich, M., Funk, H. T., Schmitz-Linneweber, C., Poltnigg, P., Sabater, B., Martin, M., et al. (2005). Editing of plastid RNA in *Arabidopsis thaliana* ecotypes. *Plant J.* 43, 708–715. doi: 10.1111/j.1365-313X.2005.02484.x

Tillich, M., Lehwark, P., Morton, B. R., and Maier, U. G. (2006). The evolution of chloroplast RNA editing. *Mol. Biol. Evol.* 23:1912. doi: 10.1093/molbev/msl054

Tseng, C. C., Lee, C. J., Chung, Y. T., Sung, T. Y., and Hsieh, M. H. (2013). Differential regulation of Arabidopsis plastid gene expression and RNA editing in non-photosynthetic tissues. *Plant Mol. Biol.* 82, 375–392. doi: 10.1007/s11103-013-0069-5

Tseng, C. C., Sung, T. Y., Li, Y. C., Hsu, S. J., Lin, C. L., and Hsieh, M. H. (2010). Editing of accD and ndhF chloroplast transcripts is partially affected in the *Arabidopsis vanilla* cream1 mutant. *Plant Mol. Biol.* 73, 309–323. doi: 10.1007/s11103-010-9616-5

Tsudzuki, T., Wakasugi, T., and Sugiura, M. (2001). Comparative analysis of RNA editing sites in higher plant chloroplasts. *J. Mol. Evol.* 53, 327–332. doi: 10.1007/s002390010222

Van Den Bekerom, R. J. M., Dix, P. J., Diekmann, K., and Barth, S. (2013). Variations in efficiency of plastidial RNA editing within ndh transcripts of perennial ryegrass (*Lolium perenne*) are not linked to differences in drought tolerance. *AoB Plants* 5:plt035. doi: 10.1093/aobpla/plt035

Wolf, P. G., Rowe, C. A., and Hasebe, M. (2004). High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern *Adiantum capillus-veneris*. *Gene* 339, 89–97. doi: 10.1016/j.gene.2004.06.018

Wu, C., Li, X., Guo, S., and Wong, S.-M. (2016). Analyses of RNA-Seq and sRNA-Seq data reveal a complex network of anti-viral defense in TCV-infected *Arabidopsis thaliana*. *Sci. Rep.* 6:36007. doi: 10.1038/srep36007

Xu, T., Wang, Y., Liu, X., Lv, S., Feng, C., Qi, M., et al. (2015). Small, RNAs and degradome sequencing reveals microRNAs and their targets involved in tomato pedicel abscission. *Planta* 242, 963–984. doi: 10.1007/s00425-015-2318-0

Zhelyazkova, P., Hammani, K., Rojas, M., Voelker, R., Vargas-Suarez, M., Borner, T., et al. (2012). Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. *Nucleic Acids Res.* 40, 3092–3105. doi: 10.1093/nar/gkr1137

Zhou, W., Cheng, Y., Yap, A., Chateigner-Boutin, A.-L., Delannoy, E., Hammani, K., et al. (2009). The Arabidopsis gene YS1 encoding a DYW protein is required for editing of rpoB transcripts and the rapid development of chloroplasts during early growth. *Plant J.* 58, 82–96. doi: 10.1111/j.1365-313X.2008.03766.x

Zimmer, S. L., Schein, A., Zipor, G., Stern, D. B., and Schuster, G. (2009). Polyadenylation in Arabidopsis and *Chlamydomonas organelles*: the input of nucleotidyltransferases, poly(A) polymerases and polynucleotide phosphorylase. *Plant J.* 59, 88–99. doi: 10.1111/j.1365-313X.2009.03853.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

4. Capítulo 2 - Salt stress affects the mRNA editing in soybean chloroplasts

N.F. Rodrigues[a]; G.C. da Fonseca[b]; F.R. Kulcheski[b] and R. Margis[a,b,c*]

[a]PPGBM, Departamento de Genética, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brasil.
[b]PPGBCM, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brasil.
[c]Departamento de Biofísica, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brasil.

Artigo publicado no periódico Genetics and Molecular Biology (2017)

Genetics and Molecular Biology, 40, 1(suppl), 200-208 (2017)

DOI: http://dx.doi.org/ 10.1590/1678-4685-GMB-2016-0055

Research Article

Salt stress affects mRNA editing in soybean chloroplasts

Nureyev F. Rodrigues [1], Guilherme C. da Fonseca [2], Franceli R. Kulcheski [2] and Rogério Margis [1,2,3]

[1]*Departamento de Genética, PPGBM, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil.*
[2]*Centro de Biotecnologia, PPGBCM, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil.*
[3]*Departamento de Biofísica, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil.*

Abstract

Soybean, a crop known by its economic and nutritional importance, has been the subject of several studies that assess the impact and the effective plant responses to abiotic stresses. Salt stress is one of the main environmental stresses and negatively impacts crop growth and yield. In this work, the RNA editing process in the chloroplast of soybean plants was evaluated in response to a salt stress. Bioinformatics approach using sRNA and mRNA libraries were employed to detect specific sites showing differences in editing efficiency. RT-qPCR was used to measure editing efficiency at selected sites. We observed that transcripts of *NDHA*, *NDHB*, *RPS14* and *RPS16* genes presented differences in coverage and editing rates between control and salt-treated libraries. RT-qPCR assays demonstrated an increase in editing efficiency of selected genes. The salt stress enhanced the RNA editing process in transcripts, indicating responses to components of the electron transfer chain, photosystem and translation complexes. These increases can be a response to keep the homeostasis of chloroplast protein functions in response to salt stress.

Keywords: small RNA, chloroplast, RNA editing, PPR, salt stress.

Received: March 23, 2016; Accepted: June 20, 2016.

Introduction

Soybean (*Glycine max* L.) is one of the major legume crops in the world, providing an abundant source of oil and protein-rich food for human and animal consumption (Le *et al.*, 2012). The high demand for protein in meals drove to further expansion of oilseed production and has favored an increase of soybean production, especially in Brazil (Guevara *et al.*, 2015). In Brazilian agriculture, soybean is the most important crop. Currently, Brazil is the second largest producer behind the United States. Soybeans are expected to continue being the most lucrative export product with more than half of Brazilian production destined for world markets (Guevara *et al.*, 2015). However, like many crops, soybean is subject to several abiotic stresses that reduce its yield.

Plants are exposed to a range of stress conditions such as oxidative stress, variant temperature, light intensity, waterlogging, drought and salinity. These abiotic stresses affect the whole plant, compromising basic molecular and physiological aspects from germination to the reproduction phases (Mahajan and Tuteja, 2005). Salt stress is one of the main environmental stresses, and it affects economically important crop species that are very sensitive to salinity, such as bean (*Phaseolus vulgaris*), maize (*Zea mays*), rice (*Oryza sativa*) and soybean (Wang *et al.*, 2003; Zheng *et al.*, 2009). Salt-affected soils occur in more than 100 countries and their worldwide extent is estimated at about 1 billion ha (FAO and ITPS, 2015). Salinity stress affects mainly lipids, ions levels, malate and nitrogen metabolism, anti-oxidative enzymes and antioxidants, chloroplast structure and photosynthesis (Parida and Das, 2005). Many studies have been dedicated to the impact of salinity on photosynthetic activity, carbon assimilation, pigment composition, electron transport, and photosystem I and II efficiency (Sudhir *et al.*, 2005; Parida and Das, 2005; Koyro, 2006). Clearly, there is a link between effects on photosynthesis and chloroplast, however, certain works have looked specifically at plastid salt stress effects (Gomez, 2003; Zhang *et al.*, 2008; Zheng *et al.*, 2009).

Chloroplasts are complex organelles that have their own gene expression machinery, intricate post-transcriptional processes and a fine coordination with nuclear gene expression. Chloroplasts have received particular interest because they are responsible for photosynthesis. Al-

Send correspondence to Rogerio Margis. Centro de Biotecnologia, sala 213, prédio 43431, Universidade Federal do Rio Grande do Sul (UFRGS), PO Box 15005, 91501-970, Porto Alegre, RS, Brazil. E-mail: rogerio.margis@ufrgs.br

terations in metabolic pathways, in specific signals like redox state, or in protein structures can lead to disruption in plastid activity and, consecutively, in plant yield. RNA editing, a post transcriptional process, consists in nucleotide conversions from cytosine (C) to uracil (U), or, less frequently, from U to C. This process, also present in mitochondria, is performed by deamination and amination reactions (Chateigner-Boutin and Small, 2011; Hayes *et al.*, 2015). Usually, editing events preserve amino acids that are phylogenetically conserved by restoring the codon sequence. The most frequent change is serine to leucine, but other alterations, including silent or non-conservative changes, have also been described (Inada *et al.*, 2004; Chateigner-Boutin and Small, 2010). In both organelles, editing can create an initiation codon, and create or remove stop codons. Editing can also be found in introns (prerequisite for splicing in some cases) and in untranslated regions (UTR) (Takenaka *et al.*, 2008; Castandet and Araya, 2011). This powerful and intriguing process has been studied due its essential function and also because of the impact in the evolutionary process (Takenaka *et al.*, 2013).

Plastid RNA editing depends on the editosome machinery to precisely process the emerging transcripts. The editosome composition has not yet been completely identified. However, some components of the editing machinery, like the pentatricopeptide repeat (PPR) proteins, were already recognized. The PPR motif is a 35-amino-acid repeat that folds into a pair of antiparallel alpha helices. Arrays of tandem PPR motifs form a superhelical ribbon-like sheet (Small and Peeters, 2000; Barkan and Small 2014). In land plants, the PPR gene family contains from 400 to more than 1000 members (Barkan and Small, 2014). The PPR proteins are classified into two major subfamilies, P-type and PLS-type PPRs. The PLS-type PPR proteins can be further divided into three subgroups: E, E+, and DYW, that differ in the presence of an optional C-terminal region (Lurin *et al.*, 2004). Most PLS-type PPR proteins involved in editing act as site-recognition factors, recognizing the 5' region upstream of the editable C residue (Yagi *et al.*, 2013). PLS-type PPR proteins presenting cytidine deaminase motifs within the DYW domain have been described as being directly responsible for RNA editing activity (Boussardon *et al.*, 2014; Wagoner *et al.*, 2015). Other PPR proteins, as HCF152 and PPR10, are involved in intercistronic processing of polycistronic precursor transcripts or in stabilizing specific RNAs (Barkan and Small, 2014; Yap *et al.*, 2015).

Diverse studies have been done to analyze editing regulation of plastids under various situations, such as tissue-specific differences, responses to molecular signals, effects in immunity, and responses to abiotic stress (Kakizaki *et al.*, 2012; García-Andrade *et al.*, 2013; Tseng *et al.*, 2013). The potential of the RNA editing efficiency as a marker for stress tolerance or as a target for genetic modification was evaluated in some studies. For example, incomplete editing caused by increased temperature is correlated with change in plastid translation in maize (Nakajima and Mulligan, 2001). Specifically, heat stress leads to loss of editing sites and intron splicing reactions in *NDHB* transcripts (Karcher and Bock 2002). Variations in the efficiency of plastid editing in *NDH* transcripts was evaluated and not linked to differences in drought tolerance in perennial ryegrass (*Lolium perenne*) (Van Den Bekerom *et al.*, 2013).

Most of the studies on RNA editing have used the reverse transcription PCR (RT-PCR) method of total chloroplast mRNAs and cloning of several chloroplast cDNA fragments into vectors to be sequenced (Rüdinger *et al.*, 2009). Another method is to design primers to amplify target genes from cDNA samples and sequence them (Wolf *et al.*, 2004). RNA editing events could also be detected by using chloroplast cDNA datasets as templates for amplification in Poisoned Primer Extension methodology, or also by High Resolution Melting (HRM) analysis (Chateigner-Boutin and Small, 2007). Many plastid small RNAs (sRNAs) showed sequence similarities to PPR-binding sites, which provides support to the idea that large amounts of sRNAs remnants resulted from PPR protein targets (Ruwe and Schmitz-Linneweber, 2012). In this way, several chloroplast sRNAs are recovered as RNA-binding protein footprints, including PPR-editosome components, which remain in the sequencing results due to protein protection against ribonucleases.

Despite several different methodologies already described in the literature for RNA-editing recognition, in this work we evaluated the impact of salt stress on soybean C to T editing efficiency by a new method comprised by *in silico* screening of editing sequences of sRNA libraries obtained by high-throughput sequencing, followed by RT-qPCR assays.

Materials and Methods

Plant material, stress treatment and RNA isolation

Soybean plants were grown over 8 days using Hoagland solution. After this period, six plants were transferred into a new Hoagland solution (establishing the control group), and six plants were submitted to a salt-stress treatment using a Hoagland solution supplemented with 200 mM NaCl. Leaves were collected after intervals of 4 and 24 hours and stored in liquid nitrogen until RNA extraction. Total RNA from leaves was isolated using TRIzol reagent (Invitrogen, CA, USA), and the RNA quality was evaluated by Nanodrop quantification and gel inspection.

sRNA/mRNA libraries, chloroplast genome, and prediction of conserved editing sites

Public sRNAs and mRNAs libraries of *G. max* leaves, deposited in NCBI GEO (http://www.ncbi.nlm.nih.gov/geo/), accession number GSE69571, were used in this study to evaluate the differen-

tial RNA editing rate when exposed to saline stress. Complete chloroplast genome and coding sequences, as well as tRNAs from soybean (NC_007942) were obtained separately from the Index of Genomes from the Chloroplast Genome Database (http://chloroplast.ocean.washington.edu/). To predict editing sites and evaluate their editing rates, the PREP-Cp tool (http://prep.unl.edu/) (Mower 2009) was used with a cutoff value of 0.5, in spite of the 0.8 default value, using the coding sequences of the chloroplast genome mentioned above.

Analyses of edited sRNAs

The sRNAs libraries were primarily aligned against the chloroplast genome, coding sequences and tRNAs, using Bowtie software (Langmead *et al.*, 2009) with 0 mismatch and not allowing reverse complement matches. The aligned reads resulted in a new file called cp_m0. The unaligned reads were submitted to a second round of alignment with 0 mismatch, against nuclear and mitochondrial genomes. The unaligned reads were further aligned with two mismatches, and no reverse complement matches were allowed against the chloroplast genome and coding sequences. This second group of aligned reads produced another file called cp_m2. Both cp_DNA fastq files were concatenated in a cp_m0_m2 file. The cp_m0_m2 files were aligned against chloroplast coding sequences using Geneious (Kearse *et al.*, 2012) R8 with the Bowtie algorithm, using the same parameters of the previous alignments. The Geneious Find Variation/SNPs tool was used with parameters set as follows: Minimum Coverage of 5, Maximum Variant P-Value of 10^{-2}, to find polymorphism Inside and Outside coding sequence and P-value calculation method as approximate. The coverage values of edited and non-edited reads were transposed to the implementation of statistical analysis. The same pipeline was used to analyze editing rates with mRNA data.

Differential expression analysis

SAM files created in the bowtie alignment were utilized to generate a count table containing data from all libraries. This table was the input file to differential expression analysis performed using DeSeq2 package (Anders and Huber, 2010) implemented in R package (R Core Team, 2015). Heatmaps were generated with normalized counts of all plastid genes for data visualization.

Editing analysis by RT-qPCR

The cDNA synthesis was carried out using approximately 1 µg of total RNA. The d26T primer was used in each reaction. Before transcription, RNA and primers were mixed with RNase-free water to a total volume of 10 µL and incubated at 70 °C for 5 min followed by ice-cooling. Then, 3 µL of 5 RT-Buffer (Promega, Madison, WI, USA), 1 µL of 5 mM dNTP (Ludwig, Porto Alegre, RS, Brazil) and 1 µL of MMLV-RT Enzyme 200 U (Promega, Madison, WI, USA) were added for a final volume of 20 µL. The synthesis was performed at 42 °C for 30 min in a Veriti Thermal Cycler (Applied Biosystems, Foster City, CA, USA), and inactivation of the enzyme was completed at 85 °C for 5 min. All cDNA samples were 100-fold diluted with RNase-free water before being used as a template in RT-qPCR analysis.

A set of primers was designed according to (Chen *et al.*, 2008) with modifications. For each editing site, we designed a set of primers composed by two specific editing primers and one unique universal primer. When the specific editing primers were designed as forward, the universal primer was designed as reverse and vice-versa. The specific editing primers containing a unique difference in the first nucleotide recognized the edited or unedited site (Figure 1). The RT-qPCR reactions were performed in a Bio-Rad CFX384 real time PCR detection system (Bio-Rad, Hercules, CA, USA) using SYBR Green I (Invitrogen, Carlsbad, CA, USA) to detect double-stranded cDNA synthesis. Reactions were completed in a volume of 10 µL containing 5 µL of diluted cDNA (1:100), 1 SYBR Green I (Invitrogen, CA, USA), 0.025 mM dNTP, 1 PCR Buffer, 3 mM MgCl2, 0.25 U Platinum Taq DNA Polymerase (Invitrogen, CA, USA) and 200 nM of each universal and C or T-specific primer set. Samples were analyzed in technical quadruplicate in a 384-well plate, and a no-template control was included. The conditions were set as follows: an initial polymerase activation step for 5 min at 95 °C, 40 cycles for 15 s at 95 °C for denaturation, 10 s at 60 °C for annealing and 10 s at 72 °C for elongation. A melting curve analysis was programmed at the end of the PCR run over the range of 65 to 99 °C, and the temperature increased stepwise by 0.5 °C.

Threshold and baselines were manually determined using the Bio-Rad CFX manager software. To calculate the

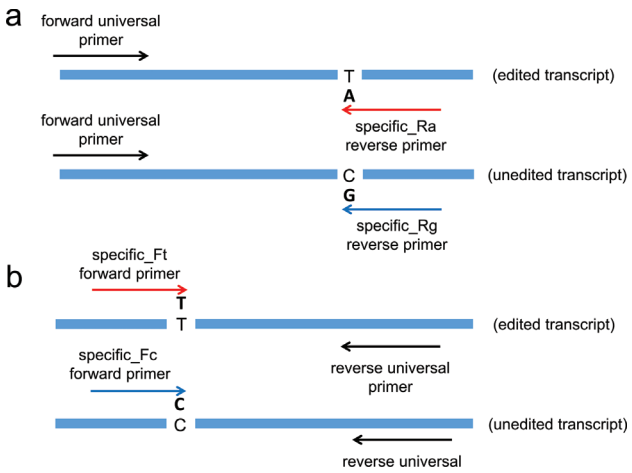

Figure 1 - Schematic illustration of qPCR analysis of RNA editing frequency showing relative locations of (A) specific-reverse and (B) specific-forward qPCR primers. Arrows depict the annealing sites of qPCR primers.

relative expression of transcripts we used the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). Primer efficiencies were calculated by LinRegPCR software (Ruijter *et al.*, 2009) to evaluate a possible amplification by primer efficiency bias. By doing so we obtained independent estimates of amplification efficiency for each primer in each treatment. Differences in plastid transcript editing among treatments were detected using two-tailed Student's *t*-tests between means. Significance was set at $p < 0.05$. Tests were performed with R package software (R Core Team, 2015).

Results

Rates of editing in sRNAs libraries

The PREP analysis carried out on soybean chloroplasts identified 20 different genes that contained RNA editing sites (Table S1). All predicted editing sites were confronted with the aligned sRNA reads in order to evaluate the presence/absence of editing events. Edited reads were identified in a set of 16 genes from at least one of the sRNAs library (Table 1). Among 87 predicted edited sites, 34 were confirmed by sRNAs reads. Other predicted sites,

Table 1 - Quantitative distribution of sRNAs reads in plastid editing sites, editing percentages and p-values (t-test).

Gene	Position (nt)	PREP score	Cnt-1	% edition	Cnt-2	% edition	Salt-1	% edition	Salt-2	% edition	p-value
NDHA	**1073**	**1**	**4**	**0.75**	**1**	**1**	**1**	**0.20**	**9**	**0.60**	**0.033**
NDHB	**149**	**1**	**11**	**0.55**	**4**	**0.80**	**6**	**0.33**	**5**	**0.36**	**0.046**
PSBF	**77**	**1**	**8**	**1**	**10**	**1**	**14**	**1**	**7**	**1**	**-**
RPS14	**80**	**1**	**24**	**0.75**	**17**	**0.85**	**14**	**0.88**	**19**	**0.90**	**0.079**
RPS16	**212**	**0.83**	**10**	**0.90**	**6**	**0.75**	**4**	**0.57**	**9**	**0.75**	**0.073**
ACCD	617	0.8	7	0.86	5	1	8	1	0	nd	0.275
ATPF	92	0.86	0	nd	3	1	3	1	3	1	0.225
CLPP	559	1	16	0.81	13	0.81	8	1	10	0.71	0.643
MATK	935	0.57	6	ne	0	nd	0	nd	1	0.08	0.225
NDHB	542	1	0	nd	1	1	1	1	0	nd	1.00
	586	1	0	ne	1	1	2	1	0	nd	1.00
	737	1	1	1	2	1	0	nd	0	nd	-
	746	1	1	1	4	1	0	nd	1	0.50	0.035
	830	1	0	ne	1	0.50	1	1	4	0.67	0.035
	836	1	0	ne	2	1	0	nd	6	0.86	0.860
	1112	1	6	0.67	4	1	5	0.83	3	0.60	0.383
	1255	1	1	1	0	nd	0	nd	0	nd	0.225
	1481	1	3	1	3	1	2	0.67	4	1	0.225
NDHD	2	1	1	ne	1	1	0	nd	0	nd	0.225
	674	1	0	nd	0	nd	1	1	0	nd	0.225
	878	1	1	ne	2	0.67	2	1	2	0.67	0.104
	1298	0.8	0	nd	2	1	0	nd	0	nd	0.225
NDHF	586	0.8	1	ne	1	0.33	0	nd	0	nd	0.225
PSAI	79	1	0	nd	1	1	3	1	0	nd	1.000
PSBE	214	1	23	0.91	20	0.91	20	0.91	24	1	0.239
RPOB	338	1	2	0.50	1	1	0	nd	1	1	0.496
	551	1	0	nd	1	1	0	nd	0	nd	0.225
	566	1	0	nd	1	0.33	0	nd	1	0.50	0.660
	2000	1	1	1	0	nd	1	1.00	1	0.20	0.801
	2819	1	2	0.50	0	nd	0	nd	0	nd	0.225
RPOC1	41	1	0	nd	1	1	0	nd	0	nd	0.225
	488	0.71	0	nd	0	nd	0	nd	2	0.67	0.225
RPOC2	3284	0.57	2	0.50	0	nd	0	nd	0	nd	0.225
RPS14	194	0.71	20	0.05	26	0.04	9	0.11	11	0.09	0.003

Ne: no edition; nd: not defined (without coverage)

even with a higher PREP score value, that should indicate a higher confidence, could not be confirmed because they did not present enough coverage (Table S1). A group of four genes was selected considering their total coverage and for being sites with statistical differential values of edited reads between control and salt treatment: *NDHA*-1073 (p = 0.033), *NDHB*-149 (p = 0.046), *RPS14*-80 (p = 0.079) and *RPS16*-212 (p = 0.073) (Table 1). Other editing sites showed relevant p-value in leaves libraries, however, they were not selected when their total coverage was lower than four reads (Table 1).

Specific primers were designed to detect edition in the four genes and also in *PSBF*-77 (Table S2) that presented 100% of edited reads in all anchored sRNAs. Except for *RPS14*-80, sRNA analysis demonstrated that in the selected genes, the editing percentage was higher in control libraries than in salt-treated ones (Table 1). A parallel analysis of editing sites using mRNA data showed relevant values in coverage and edited reads that shared similar patterns to those observed with sRNA, except for *NDHA*-1073 and *NDHB*-149 (Table S3).

Rate of editing of chloroplast transcripts by RT-qPCR

RT-qPCR was used to measure the relative amount of edited and unedited plastid transcripts at 4 and 24 hours, comparing control and salt treatment. Using LinRegPCR software, the efficiency of each amplification was calculated; for each editing primer, only reactions with efficiency higher than 1.75 were maintained in the analysis. The mean efficiency of all primers was higher than 1.80, and was not significantly different when compared with the pairs of C/G and T/A specific primers (Table S4).

The rate of edition was affected in all four genes when leaf samples were collected 4 hours after the salt treatment. The percentage of C to T editing varied in all genes. A statistically significant increase in RNA edition was observed for salt-treated samples: *NDHB*-149 presented an increase in editing from 88.7% to 93.7% (p = 0.004) (Figure 2a), *RPS14*-80 from 94.76% to 96.20% (p = 0.05) (Figure 2c) and *RPS16*-212 from 74.5% to 78.99% (p = 0.003) (Figure 2d). *NDHA*-1073 presented an absolute reduction in the average of editing percentage, but due to variance, without statistical significance (from 77.79% to 70.53%, p = 0.285) (Figure S3); the *PSBF*-77 editing percentage was not significantly different (from 83.36% to 84%, p = 0.629) (Figure 2b). When salt treatment was extended to 24 hours, an increase in editing percentage was verified in *PSBF*-77 from 88.75% to 94.70% (p = 0.0001) (Figure 2b), *RPS14*-80 from 96.31% to 97.76% (p = 0.025) (Figure 2c) and *RPS16*-212 from 73.10% to 91.65% (p = 0.0002) (Figure 2d). *NDHA*-1073 and *NDHB*-149 presented no statistical differences in their editing percentages, with values from 61.51% to 60.97% (p = 0.861) (Figure S3), and from 82.18% to 84.39% (p = 0.395) (Figure 2a) respectively.

In order to evaluate if differences in editing efficiency could be correlated with transcriptional rate, a differential gene expression of chloroplast editing genes was performed using RNA sequence libraries. In sRNAs libraries, no differences were found between control and salt treatment for the analyzed chloroplast editing genes (Figure S1). The same analysis of chloroplast gene expression was performed with mRNA libraries, and no differences were found (Figure S2a). Contrarily, when all nuclear genes were compared, a differential expression was detected.

Discussion

Plant responses to salt stress have been examined due to their agronomic implications. Our results demonstrated variability in plastid transcript editing in soybeans, in response to salt treatment. The selected editing sites showed different coverage of sRNAs when control samples were compared to salt treated ones. Plastid sRNAs present as peaks of sequence reads indicated that they are found at coverage levels similar to, or even higher than matching mRNAs (Zhelyazkova *et al.*, 2011). The parameters that determine the rate of the initiating endonucleolytic cleavage for chloroplast RNA decay are not known. These parameters are likely to include sequence and structure of mRNAs, their extent of ribosome association, and the presence of other RNA-binding proteins that mask or expose potential RNase cleavage sites (Barkan, 2011). Therefore, an increase in translation and consequent protection by the ribosome and PPR-like proteins association can lead to a reduction in the degradation of edited transcripts. This could explain the reverse correlation between total sRNA coverage decrease in editing sites and the increase in editing percentage demonstrated by RT-qPCR assays, as observed for *NDHB*-149.

The *NDHB* gene encodes part of the hydrophobic thylakoid-inserted arm in the NAD(P)H dehydrogenase (NDH) complex; this complex plays a role in alleviating over-reduction in the stroma under stress conditions (Martín and Sabater, 2010; Peng *et al.*, 2011); therefore, the increase in *NDHB*-149 editing found after 4 hours of salt treatment could contribute to the maintenance of the NDH complex, avoiding an initial impact in the redox state of plastids in treated plants. Moreover, *NDHB* editing maintenance is also essential to cyclic electron flow around photosystem 1 (CEF1), that has been demonstrated as a correlated process in salt tolerance (Lu *et al.*, 2008). In *G. max* varieties, chlorophyll fluorescence, NDH-dependent CEF activity, *NDHB* mRNA abundance, and constitutive levels of NDH-B protein were much higher in a salt-tolerant variety than in the salt-sensitive one (He *et al.*, 2015);. The elevated editing percentage, observed 4 hours after salt treatment, can be linked to this increase in translation of the *NDHB* gene and NDH-dependent CEF activity enhancement in the salt-tolerance response. Our chloroplast gene expression data presented no differences, but other experi-

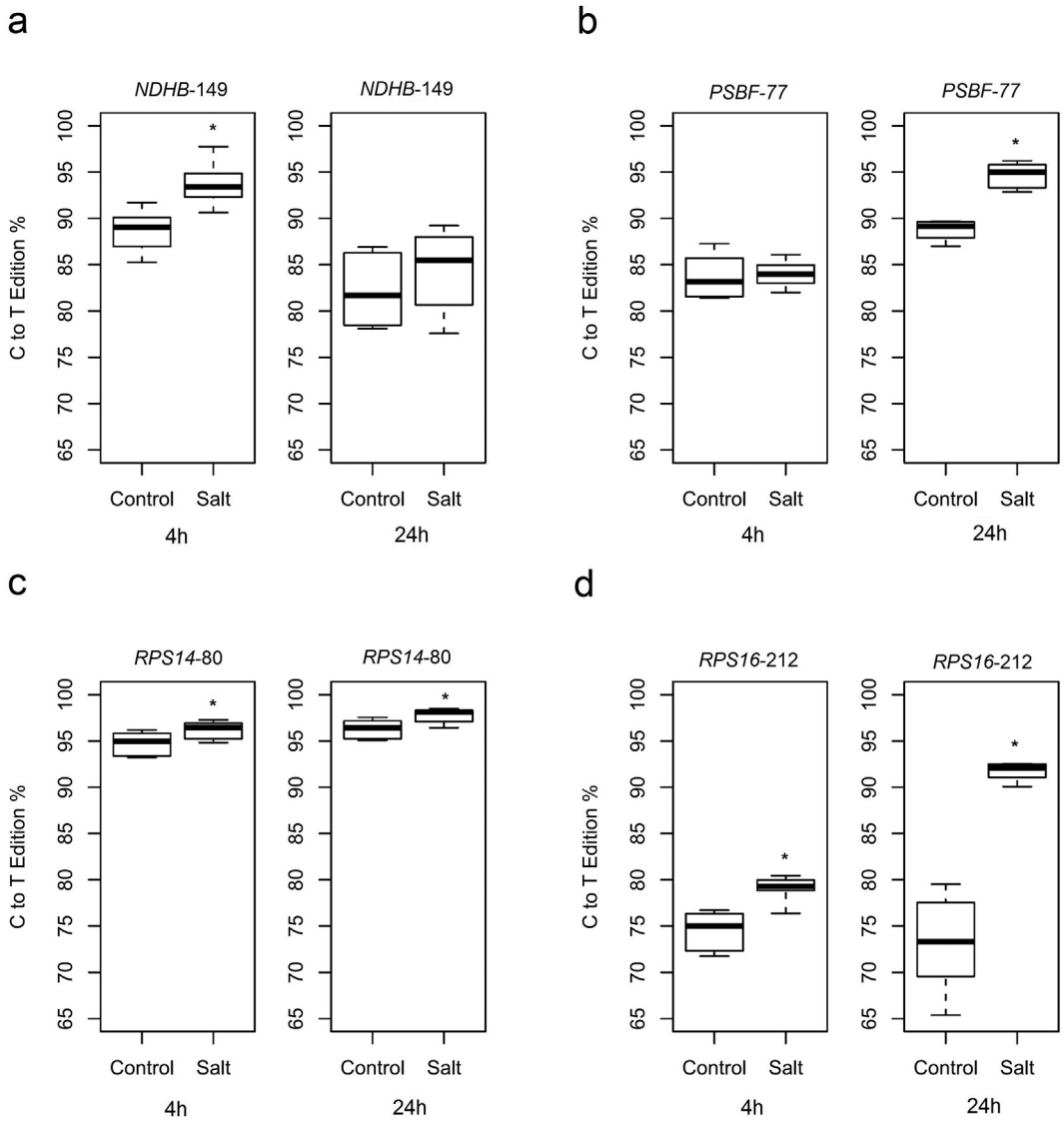

Figure 2 - Boxplot indicating the editing of (a) *NDHB*-149, (b) *PSBF*-77, (c) *RPS14*-80, and (d) *RPS16*-212 sites of control and salt stress plants, in 4h and 24 hours treatment. Box area represents the lower and the upper percentiles. The upper whisker of the boxplot indicates the highest editing value observed; the lower whisker, the lowest editing value; and the middle line, the median editing value. Asterisk indicate significantly different values at $P < 0.05$.

mental approaches are necessary to confirm a possible role of transcriptional changes in the increase of editing. After 24 hours of treatment, the *NDHB* editing level returned to normal baseline, possibly causing a mechanism by which the photosynthesis system can be impaired, when ROS begin to cause effects, such as inhibition of PSII repair and of protein synthesis.

The impact of non-editing of the *PSBF* plastid gene has been described in an *LPA66* mutant for which a PPR responsible for editing *PSBF*-77 should be encoded. Its morphological aspects were reduced growth, and pale green leaves under optimal growth, due to perturbed PSII functions (Cai *et al.*, 2009). In our results, the editing percentage of *PSBF*-77 showed an increase during the salt stressed condition, probably aiming at translation and repair enhancement of PSII. Although after 24 hours of treatment an increase in editing percentage of *PSBF* transcripts (component of PSII) occurred, salt stress has been reported to enhance photodamage to PSII by excess ROS suppressing transcription and translation of the *PSBA* gene and inhibiting the repair of PSII in *Synechocystis* (Kreslavski *et al.*, 2007; Murata *et al.*, 2007).

The *RPS14* and *RPS16* genes encode small ribosomal subunits, and among the plastid ribosomal genes, *RPS16* is an essential plastid gene that cannot be inactivated, having thus, an important role in the translation process (Tiller *et al.*, 2012). In both treatment intervals, the editing percentage showed an increase, being higher at 24 hours than at 4 hours of treatment. This increase can be related to a need for further translation of plastid proteins under salt stress.

Decreased or incomplete editing of *RPS14* and *RPS16* transcripts can affect the plastid-encoded protein synthesis. Effects of incomplete editing in *RPS12* were reported, resulting in the synthesis of polymorphic polypeptides in plant mitochondria (Phreaner, 1996). In heat stress, the editing status of *RPS14* decreased rapidly in response to change in temperature, and it remained low after an extended period of acclimatization (Nakajima and Mulligan, 2001). *RPS14* and *RPS16* gene expression is regulated by cytokinins (CK) and abscisic acid (ABA) (Cherepneva *et al.*, 2003; Yamburenko *et al.*, 2013). Chloroplast transcription can be stimulated by CK in response to ABA, drought, and salt-induced senescence. Specific ABA and stress-responsive CK receptors have been described, and maybe a crosstalk among CK, ABA and stress signaling pathways exists (Tran *et al.*, 2007). The increase in editing of *RPS14* and *RPS16* transcripts can be linked to a CK response against salt-induced senescence.

Based on our results, salt stress enhances the editing process in transcript components of the NDH, PSII, and translation complexes. All analyzed editing sites had a percentage of increase that can be a response to keep homeostasis of chloroplast functions. The maintenance of edited codons seems to be essential for protein function, and the editing process responds to this demand. Other studies that measure transcription, editing and translation of edited genes in different time intervals and salt concentrations can help to reveal the floating diversity in all edited transcripts and correlate these to other salt stress-induced responses of the editing process.

Acknowledgments

This work was sponsored by CNPq, Brazil. FRK was sponsored by a FAPERGS/CAPES-DOCFIX (1634-2551/13-9) grant.

References

Anders S and Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106.

Barkan A (2011) Expression of plastid genes: Organelle-specific elaborations on a prokaryotic scaffold. Plant Physiol 155:1520-1532.

Barkan A and Small I (2014) Pentatricopeptide repeat proteins in plants. Annu Rev Plant Biol 65:415-442.

Boussardon C, Avon A, Kindgren P, Bond CS, Challenor M, Lurin C and Small I (2014) The cytidine deaminase signature HxE(x)(n)CxxC of DYW1 binds zinc and is necessary for RNA editing of ndhD-1. New Phytol 203:1090-1095.

Cai W, Ji D, Peng L, Guo J, Ma J, Zou M, Lu C and Zhang L (2009) LPA66 Is required for editing psbF chloroplast transcripts in Arabidopsis. Plant Physiol 150:1260-1271.

Castandet B and Araya A (2011) RNA editing in plant organelles. Why make it easy? Biochemistry 76:924-931.

Chateigner-Boutin A.-L and Small I (2007) A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. Nucleic Acids Res 35:e114

Chateigner-Boutin A-L and Small I (2010) Plant RNA editing. RNA Biol 7:213-219.

Chateigner-Boutin AL and Small I (2011) Organellar RNA editing. Wiley Interdiscip Rev RNA 2:493-506.

Chen Y, Kao S, Chou H and Lin W (2008) A real-time PCR method for the quantitative analysis of RNA editing at specific sites. Anal Biochem 375:46-52.

Cherepneva GN, Schmidt K-H, Kulaeva ON, Oelmüller R and Kusnetsov VV (2003) Expression of the ribosomal proteins S14, S16, L13a and L30 is regulated by cytokinin and abscisic acid. Plant Sci 165:925-932.

FAO and ITPS (2015) Global Soil Status, Processes and Trends. Status of the World's Soil Resources (SWSR) - Main Report. Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils, Rome, pp 98-168.

García-Andrade J, Ramírez V, López A and Vera P (2013) Mediated plastid RNA editing in plant immunity. PLoS Pathog 9:e1003713.

Gomez JM (2003) Location and effects of long-term NaCl stress on superoxide dismutase and ascorbate peroxidase isoenzymes of pea (*Pisum sativum* cv. Puget) chloroplasts. J Exp Bot 55:119-130.

Guevara J, Sukerman M and Velasco M (2015) OECD-FAO Agricultural Outlook 2015. http://www.oecd-ilibrary.org/docserver/download/5115021e.pdf.

Hayes ML, Dang KN, Diaz MF and Mulligan RM (2015) A conserved glutamate residue in the C-terminal deaminase domain of Pentatricopeptide Repeat Proteins is required for RNA editing activity. J Biol Chem 290:10136-10142.

He Y, Fu J, Yu C, Wang X, Jiang Q, Hong J, Lu K, Xue G, Yan C, James A, *et al.* (2015) Increasing cyclic electron flow is related to Na+ sequestration into vacuoles for salt tolerance in soybean. J Exp Bot 66:6877-6889.

Inada M, Sasaki T, Yukawa M, Tsudzuki T and Sugiura M (2004) A systematic search for RNA editing sites in pea chloroplasts: An editing event causes diversification from the evolutionarily conserved amino acid sequence. Plant Cell Physiol 45:1615-1622.

Kakizaki T, Yazu F, Nakayama K, Ito-Inaba Y and Inaba T (2012) Plastid signalling under multiple conditions is accompanied by a common defect in RNA editing in plastids. J Exp Bot 63:251-260.

Karcher D and Bock R (2002) Temperature sensitivity of RNA editing and intron splicing reactions in the plastid ndhB transcript. Curr Genet 41:48-52.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, *et al.* (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647-1649.

Koyro H-W (2006) Effect of salinity on growth, photosynthesis, water relations and solute composition of the potential cash crop halophyte *Plantago coronopus* (L.). Environ Exp Bot 56:136-146.

Kreslavski VD, Carpentier R, Klimov VV, Murata N and Allakhverdiev SI (2007) Molecular mechanisms of stress resistance of the photosynthetic apparatus. Biochem Suppl Ser A Membr Cell Biol 1:185-205.

Langmead B, Trapnell C, Pop M and Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25.

Le DT, Nishiyama R, Watanabe Y, Tanaka M, Seki M, Ham LH, Yamaguchi-Shinozaki K, Shinozaki K and Tran LSP (2012) Differential gene expression in soybean leaf tissues at late developmental stages under drought stress revealed by genome-wide transcriptome analysis. PLoS One 7:e49522.

Livak KJ and Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2^(-ΔΔCT) method. Methods 25:402-408.

Lu KX, Yang Y, He Y and Jiang DA (2008) Induction of cyclic electron flow around photosystem 1 and state transition are correlated with salt tolerance in soybean. Photosynthetica 46:10-16.

Lurin C, Andrés C, Aubourg S, Bellaoui M, Bitton F, Bruyère C, Caboche M, Debast C, Gualberto J, Hoffmann B, *et al.* (2004) Genome-wide analysis of Arabidopsis Pentatricopeptide Repeat Proteins reveals their essential role in organelle biogenesis. Plant Cell 16:2089-2103.

Mahajan S and Tuteja N (2005) Cold, salinity and drought stresses: An overview. Arch Biochem Biophys 444:139-158.

Martín M and Sabater B (2010) Plastid ndh genes in plant evolution. Plant Physiol Biochem 48:636-645.

Mower JP (2009) The PREP suite: Predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucleic Acids Res 37:W253-W259.

Murata N, Takahashi S, Nishiyama Y and Allakhverdiev SI (2007) Photoinhibition of photosystem II under environmental stress. Biochim Biophys Acta - Bioenerg 1767:414-421.

Nakajima Y and Mulligan M (2001) Heat stress results in incomplete C-to-U editing of maize chloroplast mRNAs and correlates with changes in chloroplast transcription rate. Curr Genet 40:209-213.

Parida AK and Das AB (2005) Salt tolerance and salinity effects on plants: A review. Ecotoxicol Environ Saf 60:324-349.

Peng L, Yamamoto H and Shikanai T (2011) Structure and biogenesis of the chloroplast NAD(P)H dehydrogenase complex. Biochim Biophys Acta - Bioenerg 1807:945-953.

Phreaner CGG (1996) Incomplete editing of rps12 transcripts results in the synthesis of polymorphic polypeptides in plant mitochondria. Planty Cell Online 8:107-117.

R Development Core Team (2011), R: A Language and Environment for Statistical Computing. The R Foundation for Statistical Computing,Vienna. Available online at http://www.R-project.org/.

Rüdinger M, Funk HT, Rensing SA, Maier UG and Knoop V (2009) RNA editing: Only eleven sites are present in the Physcomitrella patens mitochondrial transcriptome and a universal nomenclature proposal. Mol Genet Genomics 281:473-481.

Ruijter JM, Ramakers C, Hoogaars WMH, Karlen Y, Bakker O, van den Hoff MJB and Moorman AFM (2009) Amplification efficiency: Linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res 37:e45-e45.

Ruwe H and Schmitz-Linneweber C (2012) Short non-coding RNA fragments accumulating in chloroplasts: Footprints of RNA binding proteins? Nucleic Acids Res 40:3106-3116.

Small ID and Peeters N (2000) The PPR motif - A TPR-related motif prevalent in plant organellar proteins. Trends Biochem Sci 25:45-47.

Sudhir P, Pogoryelov D, Kovács L, Garab G and Murthy SDS (2005) The effects of salt stress on photosynthetic electron transport and thylakoid membrane proteins in the Cyanobacterium. J Biochem Mol Biol 38:481-485.

Takenaka M, Verbitskiy D, van der Merwe JA, Zehrmann A and Brennicke A (2008) The process of RNA editing in plant mitochondria. Mitochondrion 8:35-46.

Takenaka M, Zehrmann A, Verbitskiy D, Härtel B and Brennicke A (2013) RNA editing in plants and its evolution. Annu Rev Genet 47:335-352.

Tiller N, Weingartner M, Thiele W, Maximova E, Schöttler MA and Bock R (2012) The plastid-specific ribosomal proteins of *Arabidopsis thaliana* can be divided into non-essential proteins and genuine ribosomal proteins. Plant J 69:302-316.

Tran L-SP, Urao T, Qin F, Maruyama K, Kakimoto T, Shinozaki K and Yamaguchi-Shinozaki K (2007) Functional analysis of AHK1/ATHK1 and cytokinin receptor histidine kinases in response to abscisic acid, drought, and salt stress in Arabidopsis. Proc Natl Acad Sci U S A 104:20623-20628.

Tseng CC, Lee CJ, Chung YT, Sung TY and Hsieh MH (2013) Differential regulation of Arabidopsis plastid gene expression and RNA editing in non-photosynthetic tissues. Plant Mol Biol 82:375-392.

Van Den Bekerom RJM, Dix PJ, Diekmann K and Barth S (2013) Variations in efficiency of plastidial RNA editing within ndh transcripts of perennial ryegrass (*Lolium perenne*) are not linked to differences in drought tolerance. AoB Plants 5:plt035-plt035.

Wagoner JA, Sun T, Lin L and Hanson MR (2015) Cytidine deaminase motifs within the DYW domain of two Pentatricopeptide Repeat-containing proteins are required for site-specific chloroplast RNA editing. J Biol Chem 290:2957-2968.

Wang W, Vinocur B and Altman A (2003) Plant responses to drought, salinity and extreme temperatures: Towards genetic engineering for stress tolerance. Planta 218:1-14.

Wolf PG, Rowe CA and Hasebe M (2004) High levels of RNA editing in a vascular plant chloroplast genome: Analysis of transcripts from the fern *Adiantum capillus-veneris*. Gene 339:89-97.

Yagi Y, Tachikawa M, Noguchi H, Satoh S, Obokata J and Nakamura T (2013) Pentatricopeptide repeat proteins involved in plant organellar RNA editing. RNA Biol 10:1419-1425.

Yamburenko MV, Zubo YO, Vanková R, Kusnetsov VV, Kulaeva ON and Börner T (2013) Abscisic acid represses the transcription of chloroplast genes. J Exp Bot 64:4491-4502.

Yap A, Kindgren P, Colas des Francs-Small C, Kazama T, Tanz SK, Toriyama K and Small I (2015) AEF1/MPR25 is implicated in RNA editing of plastid atpF and mitochondrial nad5 , and also promotes atpF splicing in Arabidopsis and rice. Plant J 81:661-669.

Zhang J, Tan W, Yang XH and Zhang HX (2008) Plastid-expressed choline monooxygenase gene improves salt and drought tolerance through accumulation of glycine betaine in tobacco. Plant Cell Rep 27:1113-1124.

Zhelyazkova P, Hammani K, Rojas M, Voelker R, Vargas-Suárez M, Börner T and Barkan A (2011) Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. Nucleic Acids Res 40:3092-3105.

Zheng C, Jiang D, Liu F, Dai T, Jing Q and Cao W (2009) Effects of salt and waterlogging stresses and their combination on leaf photosynthesis, chloroplast ATP synthesis, and antioxidant capacity in wheat. Plant Sci 176:575-582.

Supplementary Material

The following online material is available for this article:

Table S1 - Editing analyses of plastid CDS from small RNA seq.

Table S2 - Sequences and descriptions of real time primers.

Table S3 - Editing analyses of plastid CDS from mRNA seq.

Table S4 - Means of RT-qPCR primer efficiency and correlation.

Table S5 - Identification of chloroplast genes in heatmap.

Figure S1 - Heatmap of relative expression of plastid genes.

Figure S2 - Heatmap of relative expression of plastid genes and differentially expressed nuclear genes.

Figure S3 - Boxplot of percentage editing of the NDHA-1073 editing site.

Associate Editor: Nelson Saibo

5. Capítulo 3 - Identification of *Glycine max* trans-factors associated to plastid *atpF*, *ndhB* and *rps14* RNA editing

Nureyev F. Rodrigues[a], Fábio C. S. Nogueira[b,c], Gilberto B. Domont[b], Rogerio Margis[a,d,e*]

Affiliations:

[a] PPGBM, Departamento de Genética, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brasil.

[b] Unidade Proteômica, PPGBq, Departamento de Bioquímica, Instituto de Química, Universidade Federal do Rio de Janeiro - UFRJ, Rio de Janeiro, RJ, Brasil.

[c] Laboratório de Proteômica, LADETEC, Instituto de Química, Universidade Federal do Rio de Janeiro - UFRJ, Rio de Janeiro, RJ, Brasil.

[d] Departamento de Biofísica, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brasil.

[e] PPGBCM, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brasil.

Artigo a ser submetido à revista Plant Cell Reports

Title:

Identification of *Glycine max trans*-factors associated to plastid *atpF*, *ndhB* and *rps14* RNA editing.

Authors:

Nureyev F. Rodrigues[a], Fábio C. S. Nogueira[b,c], Gilberto B. Domont[b], Rogerio Margis[a,d,e*]

Affiliations:

[a] PPGBM, Departamento de Genética, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brasil.

[b] Unidade Proteômica, PPGBq, Departamento de Bioquímica, Instituto de Química, Universidade Federal do Rio de Janeiro - UFRJ, Rio de Janeiro, RJ, Brasil.

[c] Laboratório de Proteômica, LADETEC, Instituto de Química, Universidade Federal do Rio de Janeiro - UFRJ, Rio de Janeiro, RJ, Brasil.

[d] Departamento de Biofísica, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brasil.

[e] PPGBCM, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brasil.

***Correspondence**

Rogerio Margis, Centro de Biotecnologia, sala 206, prédio 43422, Universidade Federal do Rio Grande do Sul – UFRGS, PO Box 15005, CEP 91501-970, Porto Alegre, RS, Brazil. Tel: 55 (51) 3308-7766 Fax: 55 (51) 3308-6072

E-mail: rogerio.margis@ufrgs.br

Keywords: chloroplast; soybean; RNA editing; PPR; proteomics.

Author contribution statement

RM, and NFR conceived and designed research. NFR conducted *in silico* analysis. FCSN, NFR and GBD conducted MS experiments. NFR, FCSN and RM analyzed data. NFR, FCSN and RM wrote the manuscript. All authors read and approved the manuscript.

Abstract

RNA editing is a posttranscriptional process that changes nucleotide sequences of cytosine-to-uracil by a deamination reaction that can revert non-neutral codon mutations. Pentatricopeptide repeat proteins are editing *trans*-elements described to recognize specific RNA *cis*-elements and perform the deamination reaction. The PPR proteins are classified in P and PLS subfamily. Several studies have described these trans-elements and their cognate editing sites. In this paper, we have designed RNA biotinylated probes based in soybean plastid RNA editing sites to perform a *trans*-element specific protein isolation. Soybean *cis*-elements from the three different RNA probes show some differences in respect to other species. Samples from probe pulldown were submitted to mass spectrometry for protein identification. Among the detected peptides, five corresponded to PPR proteins. More than one PPR protein, with distinct functional domains, was pulled-down with a unique RNA probe. Comparison of Arabidopsis genes to the soybean PPR proteins allow identification of the closest related homologs. A differential gene expression analysis demonstrated that one PPR protein have an increased expression under salt stress. The present study represents the first identification of RNA editing *trans*-elements in soybeans. Our data also indicates that potential multiple *trans*-elements should interact to RNA *cis*-elements in order to perform the editing.

Introduction

Chloroplasts harbor the metabolic core that makes plants what they are. But the evolutionary history of these organelles was not easy. Chloroplast were once free-living prokaryotes. Several adaptations were essential to circumvent the conflicts between the host genome and the endosymbiont ones. Massive transfers of genetic information to the host genome and its functional assimilation leads to retraction in endosymbiotic genome (Timmis et al. 2004; Chen et al. 2015). A strong selective pressure acted to maintain the remaining endosymbiotic genetic information. Posttranscriptional processes were selected by promoting the maintenance of essential sequences for gene expression and functional proteins. In plastids, RNA editing is a nucleotide change from cytosine to uracil (C-to-U) and less frequently from uracil to cytosine (U-to-C) by a deamination and amination reaction respectively (Tillich et al. 2006a; Chateigner-Boutin and Small 2010; Takenaka et al. 2013).

These changes are necessary to RNA maturation, to generate start or stop codons or even to result in changes of amino acid identity (Schallenberg-Rüdinger and Knoop 2016).

Studies have extensively been performed to elucidate molecular features, mechanism and machinery of plastid RNA editing. *Cis*-acting regulatory sequence were identified and reported to be determinant to plastid RNA editing site specificity (Bock et al. 1996). In general, 20 nucleotides upstream sequence, and in some case 10 nucleotides downstream sequence of RNA editing site correspond to the *cis*-elements for RNA editing (Shikanai 2006; Vu and Tsukahara 2017). Besides that, several mutants in *Arabidopsis thaliana* have been identified with plastid RNA editing defects. These defects allowed to characterize *trans*-acting factors that perform RNA editing. The first *trans*-acting factors identified were the Pentatricopeptide Repeat proteins (PPRs) (Fisk et al. 1999; Kotera et al. 2005; Fujii and Small 2011). PPR proteins are characterized by tandem arrays of the degenerate 31 to 36-amino acid repeating units, called PPR motifs, repeated in tandem up to 30 times, that folds into a pair of antiparallel α-helices, forming a solenoid structure (Small and Peeters 2000; Ichinose and Sugita 2016).

PPR proteins constitute a large family of nuclear-encoded proteins in land plants, likely to have involved in retrotransposition, genome duplication events and retention of duplicated genes (Lurin et al. 2004; O'Toole et al. 2008; Fujii and Small 2011). This protein family have about 450 members in Arabidopsis and over 1000 in land plants and correspond the most studied RNA editing factors already recognized (Cheng et al. 2016). PPR proteins form sequence-specific associations with RNA, and these associations affect folding, processing and/or translation of the RNA, thus manipulating expression of the transcript (Fujii and Small 2011). The sequence-specific associations occurs from the interaction between protein motifs and RNA, where 1 motif corresponds to 1 base, and the amino acids at particular positions determine the nucleotide-binding specificity (Kobayashi et al. 2012; Yagi et al. 2013a).

The plastid editosome machinery have also non-PPR proteins components. The RIP/MORF family have ten members in Arabidopsis and is widespread in angiosperms. RIP1/MORF8, the first RIP/MORF member described to compose plastid editosome, was identified by immunoprecipation of an epitope-tagged RARE1, a PPR protein. The *rip1* mutant shown alterations in editing events of 14 editing sites (Bentolila et al. 2012). Other two members, RIP2/MORF2 and RIP9/MORF9, are required for almost every editing site

in plastids (Takenaka et al. 2012; Bentolila et al. 2013). RIP2/MORF2 interacts with CLB19 to promote RNA editing in *rpoA* and *clpP* transcripts (Ramos-Vega et al. 2015). RIP9/MORF9 interact with PLS-type PPRs, leading to conformational changes increasing the PPR RNA-binding activity (Hackett et al. 2017). Yeast two-hybrid assays have shown that RIP2/MORF2 and RIP9/MORF9 can form heterodimers (Takenaka et al. 2012; Zehrmann et al. 2015b).

The ORRM protein family have 20 members in Arabidopsis most of which are organelle targeted. ORRM1 is plastid-targeted and the unique that harbor a RIP/MORF motif (Sun et al. 2013). ORRM1 controls more than 60% of plastid editing sites in Arabidopsis (Sun et al. 2013; Sun et al. 2016). ORRM6 is required to *psbF* transcript editing (Hackett et al. 2017). The function of the ORRM domain in editing remains unclear, but based in OMMR1 function, is speculated that its ORRM binds nearby to editing sites, preparing additional factors or enhancing the specificity of the PPR protein to binds in the *cis*-element recognition sequences (Tillich et al. 2009; Sun et al. 2013; Sun et al. 2016). The OZ protein family, with four members in Arabidopsis, have an annotated zinc finger domain. All members are organelle targeted. OZ1 mutant and transient silencing have altered editing activity in most of plastid editing sites (Naested 2004; Sun et al. 2015). Another additional proteins, CP31, OCP3 and PPO1, have been described to affects RNA editing efficiency (Coego 2005; Tillich et al. 2009; García-Andrade et al. 2013; Zhang et al. 2014).

Plastid RNA editing was reported in most of plant lineages, with exception of some liverworts in Marchantiales (Rudinger et al. 2008). The number of editing sites varies among species. In seed plants, plastid editing sites have already been reported in rice (21), maize (26), tobacco (34), cucumber (51) and *A. thaliana* (43) (Maier et al. 1995; Corneille et al. 2000; Ruwe et al. 2013; Ichinose and Sugita 2016). The identification of editing sites and measurement of editing levels have demonstrated differences among tissues (Tseng et al. 2013) and developmental stages (Miyata and Sugita 2004). These findings can be used to evaluate the impact of different stresses on these mechanisms (Nakajima and Mulligan 2001; Van Den Bekerom et al. 2013).

Soybean is a model crop with few previously studies about plastid RNA editing. Our group has described 43 phylogenetically conserved and 5 non-conserved editing sites in *Glycine max* using sRNA, RNA-seq data (Rodrigues et al. 2017a). Besides that, we also have described the salt stress effect in soybean plastid RNA editing (Rodrigues et al. 2017b).

Based in sequencing data, three *cis*-elements were selected, all of them presenting high editing levels, where greater plastid editosome activity is expected. These sequences were used to perform a RNA-pulldown protein purification to identify plastid editosome *trans*-elements components acting in selected soybean plastid editing sites and its specificity among sites, as well as other proteins that have non-specific RNA binding activity.

Materials and Methods

Cis-elements RNA Probe design

The soybean chloroplast genome was retrieved from NC_007942.1 accession. The coding sequences were used to designing the probes. The probe design used *atpF*-92, *ndhB*-1481 and *rps14*-80 editing sites as reference to select 28 upstream and 7 downstream nucleotides, totalizing a 36 nucleotides probe from each editing site: *atpF*-92, UUUAAUACCGAUAUUUUAGCAACAAAUCCAAUAAAU; *ndhB*-1481, GUAUGUGUGAUAGCAUCUACUAUACCAGGAAUAUCA; and *rps14*-80, AAAUAUCAUUUGAUUCGUCGAUCCUCAAAAAAGGAA. The RNA probes were synthesized and biotinylated at 5' end. To analyze the recognition sequence conservation among species, chloroplast coding sequences for each transcript were retrieved from *A. thaliana* (NC_000932.1), *Eucalyptus grandis* (NC_014570.1), *Eugenia uniflora* (NC_027744.1), *Nicotiana tabacum* (NC_001879.2), *Oryza sativa* (NC_001320.1), *Panicum virgatum* (NC_015990.1), *Sorghum bicolor* (NC_008602.1) and *Zea mays* (NC_001666.2). A tree was created using Neighbor-Joining method with p-distance model performed in the Molecular Evolutionary Genetics Analysis (MEGA) 6.0 software (Tamura et al. 2013) and sequence logos were generated using WebLogo3 (Crooks 2004) at http://weblogo.threeplusone.com.

Plant material and chloroplast isolation

To chloroplast isolation, soybean (Glycine max (L.) Merrill) cultivar Conquista were cultivated until the fifth trifoliate (V5) stage. The modified high salt chloroplast isolation protocol (Shi et al. 2012; Vieira et al. 2014) was followed to obtain chloroplasts. All the following steps were carried out at 0º C. Prior to extraction, 25 g (fresh weight) of leaves without petioles were collected and kept in dark for 48 h at 4º C to decrease starch level. Fresh leaves were cleaned with distilled water and homogenized in 400 ml of isolation buffer (1.5 M NaCl, 0.25 M ascorbic acid, 12.5 mM boric acid, 50 mM Tris-HCl pH 8.0, 7 mM

EDTA, 1% PVP-40 and 1 mM DTT) for 30 s. Homogenate was filtered into Falcon tubes, using two layers of Miracloth by softly squeezing the cloth. The homogenate was centrifuged twice at 200 g for 20 min at 4º C. The pellet, containing nucleus and cell-wall debris, were discarded. The supernatant was centrifuged at 3500 g for 20 min at 4º C, resulting in a chloroplast pellet contaminated with some nuclear DNA. The pellet was gently resuspended in 40 ml per tube of wash buffer (1.5 M NaCl, 12.5 mM boric acid, 50 mM Tris-HCl pH 8.0, 25 mM EDTA, 1 % PVP-40 and 1 mM DTT), followed by centrifugation at 3500 g for 20 min at 4º C. The supernatant was discarded. The pellet was resuspended again with 40 ml wash buffer and centrifuged at 3500 g for 20 min at 4º C twice to obtain the final chloroplast pellet.

Plastid protein extraction and protein isolation by RNA probe pulldown

All the following steps were carried out at 0º C, if not otherwise stated. The final chloroplast pellet was resuspended in lysis buffer (0.2 M CH_3COOK, 30 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 2 mM DTT) and transferred to a microcentrifuge tube. The resuspended solution was pulled through a syringe (0.3 mm × 8 mm) 60 times. The homogenate was centrifuged twice at 16.000 g for 20 min at 4º C. A supernatant aliquot was transferred to a new tube, and the same volume of incubation buffer (150 mM NaCl, 20 mM Tris-HCl pH 8.0, 1 mM EDTA, 5 mM $MgCl_2$, 0.5% Triton X-100) was added. The homogenate was transferred to a new tube and added biotinylated probes (final concentration 5 µM) corresponding to each editing sites. The solution was incubated at 160 rpm for 30 min at 25º C. The homogenate was transferred to a centrifuge tube containing streptavidin-agarose resin previously washed with lysis and incubation buffer 1:1 (v/v), thrice. The washing step correspond to add the solution, gentle manual shaking and resin decantation, followed by discarding the volume above the resin. The solution was maintained on gentle manual shaking for 15 min. Two washing steps were performed with lysis and incubation buffer 1:1 (v/v), followed by three washing steps with lysis and incubation buffer (without Triton X-100) 1:1 (v/v). The final solution containing streptavidin-agarose resin, biotinylated probes and plastid proteins was maintained at -20º C before sample preparation.

Sample preparation for proteomic analysis

The resins were incubated for 5 minutes, at room temperature, with 7 M urea/2 M thiourea. Proteins extracted from resins were further reduced using 10 mM DTT for 60 min at 35° C and alkylated using 40 mM iodoacetamide for 60 min at 35º C in dark. Urea

concentration was diluted to less than 1 M using 50 mM NH_4HCO_3 pH 8.0 and proteins were digested with trypsin (Promega) overnight at 35° C. Trifluoroacetic acid (TFA) was added (final concentration 0.1%) in order to stop digestion and peptides were passed through spin columns (Harvard Apparatus) filled with C-18 material, dried under vacuum and stored at -20° C for further use. Two biological replicates were subjected to digestion for each RNA probe.

Protein identification by mass spectrometry assays

Peptides derived from the tryptic digestion (2 μg) were loaded onto a C18 reversed-phase pre-column (2 cm long, 100 μm internal diameter, with ReproSil-Pur C18-AQ 5 μm beads - Dr. Maisch GmbH) and fractionated on a New Objective PicoFrit® Self-Pack column (18 cm long, 75 μm internal diameter, with ReproSil-Pur C18-AQ 3 μm beads - Dr. Maisch GmbH). The samples were analyzed in an EASY-nLC II system (Proxeon Biosystems) coupled in sequence to a high resolution ESI-LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific). The peptides were eluted using the gradient starting from 100% phase A (0.1% formic acid, 5% acetonitrile) to 35% phase B (0.1% formic acid, 95% acetonitrile) for 107 minutes, 35-100% of phase B for 5 minutes, and 100% of phase B for 8 minutes, totaling 120 minutes in a flow of 250 mL min-1. After each run, the column was washed with 100% of phase B and re-equilibrated with phase A. The m/z spectra were obtained in positive mode with data-dependent automatic acquisition - Data Dependent Acquisition (DDA) - of the MS and MS/MS spectra. The MS spectra were obtained in high resolution in the Orbitrap analyzer with resolution from 30,000 at m/z 400, mass range of m/z 350-2000, Automatic Gain Control (AGC) of 1×10^6 and maximum injection time of 500 MS. The MS/MS spectra were obtained by higher energy collisional dissociation (HCD) in the Orbitrap for the 10 most intense ions with a charge ≥ 2; resolution of 7500 at m/z 400; signal threshold of 10,000; normalized energy of collision (NCE) of 30; and dynamic exclusion of 45 s.

Proteome Discoverer 2.1 software was used for data analysis applying the Sequest™ algorithm and a *G. max* database downloaded from Phytozome (June 2017). The parameters used were: full-tryptic search space, up to two missed cleavages allowed for trypsin, precursor mass tolerance of 10 ppm, and fragment mass tolerance of 0.1 Da.

Carbamidomethylation of cysteine was included as fixed modification, and methionine oxidation and protein N-terminal acetylation as dynamic modifications

Phylogenetic analysis of trans-acting editing factors

Sequences corresponding to complete gene, coding regions, transcripts and proteins were retrieved from Phytozome database. These protein sequences were used as queries in the BLASTP with default parameters at the Phytozome database to retrieve *A. thaliana*, and *G. max*. To determine the structural organization and motif/domain composition of the trans-factors, the protein sequences were submitted to Pfam web server (http://pfam.xfam.org/) to prediction of functional domains (Finn et al. 2016). The sequence domain found in each protein sequence were retrieved to create a fasta file. The domains protein sequences were aligned using MUSCLE (Edgar 2004). The multiple alignments were manually inspected using Molecular Evolutionary Genetics Analysis (MEGA) 6.0 software (Tamura et al. 2013). The model of protein evolution for protein matrix substitution was calculated from multiple alignment by ProtTest3 (Darriba et al. 2011). The phylogenetic tree was constructed using Bayesian method, performed in BEAST 1.8.4 software (Drummond and Rambaut 2007). The Yule tree was selected as a tree prior to Bayesian analysis and 10,000,000 generations were performed with Markov chain Monte Carlo (MCMC) algorithms. The tree was visualized and edited using FigTree v1.4.3 software.

Differential gene expression

Public mRNAs libraries of soybean leaves, deposited in NCBI GEO (http://www.ncbi.nlm.nih.gov/geo/), accession number GSE69571, were used in this study to evaluate the differential gene expression of PPRs proteins identified. SAM files were created using the bowtie alignment (Langmead et al. 2009) and default parameters, with zero mismatches. A count table containing data from all libraries was created. This table was the input file to differential expression analysis performed using the Bioconductor DESeq2 package (Love et al. 2014) with a adjusted p-value cutoff of 0.05.

Results

Editing sites cis-elements conservation

Recognition sequence from *atpF*-92, *ndhB*-1481 and *rps14*-80 editing sites were analyzed at 30 downstream and 20 upstream nucleotides in eight species (Fig. 1). The *atpF*-

90 sequence is clearly divided among monocots and dicots (Fig 1a). Monocots already have a thymine in editing site location (Fig. S1a). Other differences occur after 10 upstream and 28 downstream nucleotides. The *ndhB*-1481 recognition sequence is the most conserved among all analyzed recognition sequences. Differences could be observed only in position 27 upstream and 19 downstream from editing site (Fig 1b). The *rps14*-80 recognition sequence is the most variable sequence among all analyzed. Differences could be observed even within monocots (Fig. 1c). In total, were observed 14 positions with nucleotide differences in *rps14*-80 recognition sequence (Fig. S1c).

Protein profiles of RNA probe pulldown

Despite sequence differences in the designed RNA probes, several proteins could be identified by using RNA probes pulldown. The list containing all proteins identified in this study is found in Supplemental Table 1. The elution profile using *atpF*-92, *ndhB*-1481 and *rps14*-80 probes comprises 83, 106 and 78 proteins respectively. The non-probes elution profile comprises 160 proteins. The non-redundant set of 317 proteins were submitted to TargetP (http://www.cbs.dtu.dk/services/TargetP/) web-server to predicts subcellular location. Fortyfive proteins were predicted to localize to chloroplast, 30 proteins to localize to mitochondria and 44 to secretory pathway (Supplemental Table 2). Chloroplast RNA binding proteins, PPRs, RNA helicases and translation factors, were identified in RNA probe pulldown profiles. These proteins are involved in RNA metabolism as RNA splicing and editing, and in the translation process (Table 1).

Two RNA helicases were identified in proteins profiles of RNA probe pulldown. Glyma.02G119000 was found in all RNA probe pulldown profiles and correspond to a DEAD/DEAH box helicase, an ATP-dependent RNA helicase. It's have four domains: an AAA (ATPases associated with a variety of cellular activities) domain, a helicase conserved C-terminal domain, a helicase associated domain (HA2) and an Oligonucleotide/oligosaccharide-binding (OB)-fold domain. Glyma.18G014800 correspond to a Helicase, IBR and zinc finger protein domain-containing protein and was found in RNA pulldown of *rps14*-80 probe. Translation initiator factors IF-2 (Glyma.19G044300 and Glyma.08G174200) were identified in RNA probe pulldown. Both were predicted to localize to chloroplast by TargetP. Glyma.08G174200 and Glyma.19G044300 were identified respectively in *atpF*-90 and *ndhB*-1481 probes pulldown profiles.

Others plastid proteins that are not RNA binding were also identified. The Light-harvesting complex II chlorophyll a/b binding protein 1, LHCB1 (Glyma.16G165200), was identified in *rps14*-80 probe pulldown. The CHLOROPLAST UNUSUAL POSITIONING1, CHUP1 protein (Glyma.20G185300) was identified in protein profiles of *atpF*-90 and *ndhB*-1481 probes. The weak chloroplast movement under blue light, WEB1 protein (Glyma.18G021300 and Glyma.08G266500), were identified in protein profiles of *atpF*-90, *ndhB*-1481 and *rps14*-80 probes. The magnesium chelatase subunit H (Glyma.10G097800) was identified in *atpF*-90 probe pulldown. Non-plastid proteins were also identified. Cytosolic translation and transcription factors, kinases, metabolic enzymes and, in a fewer abundance, cytoskeleton components were the main non-plastid contaminations in the RNA probe pulldown profiles (Supplemental Table 1).

Pentatricopeptide repeat proteins

PPR proteins were identified in different RNA probe pulldown profiles. In total, five PPR proteins were identified (Table 1). Glyma.11G217500 and Glyma.19G025700 proteins were identified in the *atpF*-90 pulldown profile. Glyma.11G217500 have two Pfam domains, PPR (PF01535) and PPR_2 (PF13041). Glyma.19G025700 have three domains; PPR (PF01535), PPR_3 (PF13812) and DYW (PF14432), a cytosine-deaminase domain. Glyma.01G016100 and Glyma.11G111200 were found in the *ndhB*-1481 pulldown profile. Glyma.01G016100 have three domains, PPR (PF01535), PPR_2 (PF13041) and PPR_3 (PF13812). Glyma.11G111200 two domains, PPR (PF01535) and PPR_2 (PF13041). Glyma.02G174500 and Glyma.11G111200 proteins were identified in *rps14*-80 probe pulldown profile. Glyma.02G174500 have three domains, PPR (PF01535), PPR_2 (PF13041) and DYW (PF14432). A neighbor-joining tree was created to observe the relationship between proteins. Glyma.19G025700 grouped with Glyma.02G174500, due DYW domains. The Glyma.11G217500 and Glyma.11G111200 protein grouped in another clade. They have only PPR and PPR_2 domains, although in different amounts. The protein Glyma.01G016100 have also PPR and PPR_2 domains, but besides these, due have a PPR_1 domain, it´s take place in an intermediary local in tree (Figure 2).

To identify homologs and understand the evolutionary relationships of the identified PPRs with described PPRs that are involved in plastid RNA editing in *A. thaliana*, we conducted a phylogenetic analysis using the only domain protein sequences. The final

dataset consists in 33 sequences, the five soybean PPRs identified by RNA probe pulldown and 28 Arabidopsis PPRs proteins (Table S3). The phylogenetic analysis of the PPR amino acid sequences resulted in the formation of well-supported clades separating the different PPR types (Figure 3). Besides that, PPRs from Arabidopsis formed groups with soybean identified PPRs proteins supported by high posteriori probabilities. The Glyma.02G174500 and Glyma.19G025700 grouped respectively to AT3G13770 and AT5G15340 proteins within DYW clade. Glyma.01G016100 grouped to AT5G39710 in a PPR_1/PPR2 domain clade. The Glyma.11G111200 protein grouped to AT5G50280 in a PPR/PPR2 domain clade. The Glyma.11G217500 could not group to any Arabidopsis protein and remained as a basal protein.

Gene expression analysis of identified PPR genes

A differential gene expression analysis was conducted to evaluate the expression of respective PPRs under salt stress. Were evaluated the five identified PPR genes and another seven genes, five eukaryotic elongation factor 1-beta (Glyma.02G276600, Glyma.04G195100, Glyma.06G170900, Glyma.13G073200 and Glyma.14G039100) and two F-box (Glyma.11G126500.1 and Glyma.12G051100) genes that correspond to reference genes. These genes were already described as reference genes for normalization in soybean under salt stress (Le et al. 2012). Only two genes, Glyma.02G174500 and Glyma.11G111200, both identified in *rps14*-80 probe pulldown, demonstrated differential expression between control and salt treatment libraries. Glyma.02G174500 had a 1.09-fold change (p-value 0.0117) increase. Glyma.11G111200 decrease your expression in -0.65-fold change (p-value 0.0004) (Fig. S2).

Discussion

To date, despite the great amount of plastid genome characterization studies, some of them have focused only on description of RNA editing sites and even fewer ones on their respective *cis*-elements. Besides that, model species to RNA editing *trans*-factors studies have been restricted to Arabidopsis, maize (Nakajima et al. 2001; Sun et al. 2013), rice (Asano et al. 2013; Tan et al. 2014) and *Physcomitrella patens* (Ichinose et al. 2014). In this paper, were analyzed three soybean plastid RNA editing sites and their respective *cis* and *trans*-factors. The recognition sequence of three editing sites were compared among eight

species. Each evaluated editing site have a conservation pattern that may leads to a site-recognition protein alteration among species. In tobacco plastids, RNA editing sites with similar *cis*-elements are recognized by an identical site-recognition protein (Kobayashi et al. 2007). Also in tobacco plastids, mutations in recognition sequences were harmful to editing or resulted in the generation of a new editing target (Hayes and Hanson 2006). Therefore, few mutations in the recognition sequences may not alter the site recognition protein through the evolutionary history of the species. In this same perspective, in vitro RNA editing demonstrated that deletions, insertions and mutations events can leads to a variation in protein that recognize the editing site between plant species, without loss of RNA editing (Neuwirt et al. 2005). In this way, there is a selective pressure to RNA editing trans-factors be conserved independently of their targets editing sites (Tillich et al. 2006b).

Most studies that have identified RNA editing trans-factors and their interactions are based on co-immunoprecipitation and mutant genetic screening. The immunoprecipitation is based in the use of antibodies that recognize the "target" protein fused to the generic GFP tag. The RNA binding protein coupled to RNA sequences are isolated by an anti-GFP antibody, following the RNA identification (Terzi and Simpson 2009). In this paper, we have used specific biotinylated-RNA probes binding to streptavidin resins as isolation protein step to mass spectrometry assays. This approach allowed isolate plastid proteins, including non-RNA binding proteins. The light-harvesting chlorophyll a/b-binding (LHC) proteins constitutes the higher plant light-harvesting antenna. The LHCII complex, composed by proteins Lhcb1-6, is associated with photosystem II (PSII) and mediates the flow of excitation energy toward the reaction center (Jackowski G. et al. 2001; Mozzo et al. 2008). Light-harvesting complex II chlorophyll a/b binding protein 1 (Lhcb1) is the main constituent of LHCII (Umate 2010). The CHUP1 and WEMBL proteins are involved in chloroplast positioning and photorealocation movement (Oikawa 2003; Kodama et al. 2010). CHUP1, a actin-binding protein, is involved in the chloroplast-actin filament polymerization by binding to profilin, F-actin (in vitro) and G-actin (Schmidt von Braun and Schleiff 2008). The WEB1 protein acts together with PMI2 by suppressing J-domain protein required for chloroplast accumulation response 1 (JAC1) to regulate the chloroplast-actin filament dynamics, modulating the velocity of chloroplast photorelocation movement (Kodama et al. 2010; Kong and Wada 2011). The presence of these proteins indicates a plastid specific protein isolation, although it does not indicate isolation of specific stroma proteins.

The magnesium chelatase catalyzes the insertion of Mg^{2+} into protoporphyrin IX the first dedicated step in chlorophyll biosynthesis. In higher plants, the magnesium chelatase enzyme consists of three subunits, ChlI, ChlD, and ChlH (Jensen et al. 1996). The magnesium chelatase subunit H (ChlH) have been described in Arabidopsis as a GUN (genomes uncoupled) 5 protein, a component of plastid-to-nucleus signal transduction (Mochizuki et al. 2001). A thioredoxin was also identified in same RNA pulldown profile. The thioredoxin-assisted reduction of CHLI1 subunit activates the ATPase activity of magnesium chelatase (Ikegami et al. 2007). However, thioredoxin-assisted reduction could maintains the ChlH subunit activated (Jensen et al. 2000). Other genomes uncoupled protein, GUN1, a PPR protein, have been suggest that might transiently interact with magnesium chelatase subunit D (CHLD) (Tadini et al. 2016). So, these proteins could transiently form a "complex" involved in retrograde signaling. The identification of these enzymes was already reported on other studies (Olinares et al. 2010).

RNA binding proteins were identified in all RNA probes pulldown. Two initiation factors were identified. For Glyma.08G174200, the Arabidopsis homolog is cpIF2. cpIF2 plays a vital role in translation by binding fMet-$tRNA^{met}$ with 30S ribosomes and subsequently forms a large complex with 50S ribosomes in a GTP-dependent manner (Miura et al. 2007). eIF2B is a guanine nucleotide exchange factor (GEF) protein that controls the cpIF2 activity. After cpIF2 is therefore released from ribosomes in its inactive (GDP-bound) state in complex with eIF5, cpIF2 must be reactivated to the GTP form(Singh et al. 2006). This process is carried out by eIF2B (Jennings and Pavitt 2014; Wortham and Proud 2015). Two RNA helicases were identified. One of them, Glyma.02G119000, is present in all probe pulldown. Its Arabidopsis homolog, AT1G26370 is a DEAH-box RNA helicase involved in pre-mRNA splicing (Ohtani et al. 2013). Although the Glyma.02G119000 have been predicted to localize to chloroplast by TargetP, the AT1G26370 protein was detected in the nucleolar region by subcellular localization experiments (Ohtani et al. 2013). We could not find only one Arabidopsis homolog for Glyma.18G014800 protein. The two candidates, At4G01020 and AT5G10370 are described as a ATP-dependent DEAH-box RNA helicase, chloroplastic. Despite this, AT5G10370 have been described only highly expressed in primary root tissue using microarray expression data from various datasets in the Gene Chip platform of Genevestigator (Xu et al. 2013). Despite the identification of these proteins and

their respective homologs, the significance of the relationship with probes is not yet clear. Their identification may have been due to nonspecific RNA binding.

PPR proteins were identified in probe pulldown. Recently, a study redefined the structural motifs of PPR domains (Cheng et al. 2016). According to this definition and based in our phylogenetic analysis, Glyma.01G016100, Glyma.11G217500 and Glyma.11G111200 belong to P subfamily, while Glyma.02G174500 and Glyma.19G025700 to DYW subgroup of PLS subfamily. P-type PPR proteins are involved in two main functions: stabilization and processing of specific RNA termini and control of the translation of specific mRNAs (Barkan and Small 2014). The DYW-type PPR proteins are involved in editing their cognate editing sites, and in some cases, the DYW domain may participate in editing additional sites (Hayes et al. 2015). The distribution of PPR among probe pulldown profile suggest that multiple trans-elements are necessary to editing in analyzed editing sites. In case of *atpF*-90 and *rps14*-80, a P-type and a DYW-type can interact to promotes editing. Some studies have demonstrated the requirement of two PPR proteins to RNA editing in plastid and mitochondria (Andrés-Colás et al. 2017; Guillaumot et al. 2017). The Glyma.11G111200 protein was identified in two pulldown profiles, *ndhB*-1481, *rps14*-80. OTP82 and CRR22 have been reported to acts as a site-specificity factors at multiple RNA editing sites with unrelated cis-acting elements in plastids (Okuda and Shikanai 2012). The same can occurs with The Glyma.11G111200. In vitro experiments have demonstrated a cross-competition in plastid RNA editing, suggesting a sharing of trans-factors between different editing sites (Heller et al. 2008). Sharing of trans-factors can confer a vantage by could recognize more editing sites with a lower number of required proteins. Beside that, a unique PPR can be dual target to plastid and mitochondria, acting in different cis-element of different organelles (Yap et al. 2015; Ichinose and Sugita 2016; Andrés-Colás et al. 2017).

In a previous study (Rodrigues et al. 2017b), we demonstrated some plastid RNA editing enhancement in soybean leaves under salt stress. One of them was the *rps14*-80 editing site. Here, we evaluated the expression pattern of PPR proteins under salt stress. Interestingly, Glyma.02G174500, a DYW-type protein identified in *rps14*-80 pulldown, have an increase by about one-fold. Despite slight increase, is possible that the increase in editing rate and in the Glyma.02G174500 gene expression are related because it's their cognate trans-factor. However, other experiments to confirm the relation of identified PPRs proteins with respective editing sites are necessary.

Conclusions

RNA biotinylated probes of *atpF*-90, *ndhB*-1481 and *rps14*-80 allowed the identification of five PPR proteins. Three P-type genes; Glyma.01g016100, Glyma.11g217500 and Glyma.11g111200 and two DYW-type PPR genes; Glyma.02g174500 and Glyma.19g025700. Multiple PPR proteins can interact to promote RNA editing in all three editing sites analyzed. The expression pattern of Glyma.02g174500 correspond to RNA editing pattern of a cognate editing site under salt stress, suggesting the trans-action function of DYW-type protein.

References

Andrés-Colás N, Zhu Q, Takenaka M, et al (2017) Multiple PPR protein interactions are involved in the RNA editing system in *Arabidopsis* mitochondria and plastids. Proc Natl Acad Sci 114:201705815. doi: 10.1073/pnas.1705815114

Asano T, Miyao A, Hirochika H, et al (2013) A pentatricopeptide repeat gene of rice is required for splicing of chloroplast transcripts and RNA editing of ndhA. Plant Biotechnol 30:57–64. doi: 10.5511/plantbiotechnology.12.1217a

Barkan A, Small I (2014) Pentatricopeptide Repeat Proteins in Plants. Annu Rev Plant Biol 65:415–442. doi: 10.1146/annurev-arplant-050213-040159

Bentolila S, Heller WP, Sun T, et al (2012) RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. Proc Natl Acad Sci 109:E1453–E1461. doi: 10.1073/pnas.1121465109

Bentolila S, Oh J, Hanson MR, Bukowski R (2013) Comprehensive High-Resolution Analysis of the Role of an Arabidopsis Gene Family in RNA Editing. PLoS Genet 9:e1003584. doi: 10.1371/journal.pgen.1003584

Bock R, Hermann M, Kössel H (1996) In vivo dissection of cis-acting determinants for plastid RNA editing. EMBO J 15:5052–5059.

Chateigner-Boutin A-L, Small I (2010) Plant RNA editing. RNA Biol 7:213–219. doi: 10.4161/rna.7.2.11343

Chen H, Yu Y, Chen X, et al (2015) Plastid DNA insertions in plant nuclear genomes: the sites, abundance and ages, and a predicted promoter analysis. Funct Integr Genomics 15:131–139. doi: 10.1007/s10142-014-0422-z

Cheng S, Gutmann B, Zhong X, et al (2016) Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. Plant J 85:532–547. doi: 10.1111/tpj.13121

Coego A (2005) An Arabidopsis Homeodomain Transcription Factor, OVEREXPRESSOR OF CATIONIC PEROXIDASE 3, Mediates Resistance to Infection by Necrotrophic Pathogens. PLANT CELL ONLINE 17:2123–2137. doi: 10.1105/tpc.105.032375

Corneille S, Lutz K, Maliga P (2000) Conservation of RNA editing between rice and maize plastids: are most editing events dispensable? Mol Gen Genet MGG 264:419–424. doi: 10.1007/s004380000295

Crooks GE (2004) WebLogo: A Sequence Logo Generator. Genome Res 14:1188–1190. doi: 10.1101/gr.849004

Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest-HPC: Fast selection of best-fit models of protein evolution. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 6586 LNCS:177–184. doi: 10.1007/978-3-642-21878-1_22

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214. doi: 10.1186/1471-2148-7-214

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. doi: 10.1093/nar/gkh340

Finn RD, Coggill P, Eberhardt RY, et al (2016) The Pfam protein families database: Towards a more sustainable future. Nucleic Acids Res 44:D279–D285. doi: 10.1093/nar/gkv1344

Fisk DG, Walker MB, Barkan A (1999) Molecular cloning of the maize gene crp1 reveals similarity between regulators of mitochondrial and chloroplast gene expression. EMBO J 18:2621–2630. doi: 10.1093/emboj/18.9.2621

Fujii S, Small I (2011) The evolution of RNA editing and pentatricopeptide repeat genes. New Phytol 191:37–47. doi: 10.1111/j.1469-8137.2011.03746.x

García-Andrade J, Ramírez V, López A, Vera P (2013) Mediated Plastid RNA Editing in Plant Immunity. PLoS Pathog 9:e1003713. doi: 10.1371/journal.ppat.1003713

Guillaumot D, Lopez-Obando M, Baudry K, et al (2017) Two interacting PPR proteins are major Arabidopsis editing factors in plastid and mitochondria. Proc Natl Acad Sci 114:8877–8882. doi: 10.1073/pnas.1705780114

Hackett JB, Shi X, Kobylarz AT, et al (2017) An Organelle RNA Recognition Motif Protein is Required for Photosynthetic Subunit psbF Transcript Editing. Plant Physiol 173:pp.01623.2016. doi: 10.1104/pp.16.01623

Hayes ML, Dang KN, Diaz MF, Mulligan RM (2015) A Conserved Glutamate Residue in the C-terminal Deaminase Domain of Pentatricopeptide Repeat Proteins Is Required for RNA Editing Activity. J Biol Chem 290:10136–10142. doi: 10.1074/jbc.M114.631630

Hayes ML, Hanson MR (2006) Identification of a sequence motif critical for editing of a tobacco chloroplast transcript. RNA 13:281–288. doi: 10.1261/rna.295607

Heller WP, Hayes ML, Hanson MR (2008) Cross-competition in Editing of Chloroplast RNA Transcripts in Vitro Implicates Sharing of Trans-factors between Different C Targets. J Biol Chem 283:7314–7319. doi: 10.1074/jbc.M709595200

Ichinose M, Sugita M (2016) RNA Editing and Its Molecular Mechanism in Plant Organelles. Genes (Basel) 8:5. doi: 10.3390/genes8010005

Ichinose M, Uchida M, Sugita M (2014) Identification of a pentatricopeptide repeat RNA editing factor in Physcomitrella patens chloroplasts. FEBS Lett 588:4060–4064. doi: 10.1016/j.febslet.2014.09.031

Ikegami A, Yoshimura N, Motohashi K, et al (2007) The CHLI1 subunit of Arabidopsis thaliana magnesium chelatase is a target protein of the chloroplast thioredoxin. J Biol Chem 282:19282–19291. doi: 10.1074/jbc.M703324200

Jackowski G., Kacprzak K., Jansson S. (2001) Identification of Lhcb1 / Lhcb2 / Lhcb3 heterotrimers of the main light- harvesting chlorophyll a / b - protein complex of Photosystem II (LHC II). Biochim Biophys Acta 1504:340–345. doi: 10.1016/S0005-2728(00)00262-0

Jennings MD, Pavitt GD (2014) A new function and complexity for protein translation initiation factor eIF2B. Cell Cycle 13:2660–2665. doi: 10.4161/15384101.2014.948797

Jensen PE, Reid JD, Hunter CN (2000) Modification of cysteine residues in the ChlI and ChlH subunits of magnesium chelatase results in enzyme inactivation. Biochem J 352:435. doi: 10.1042/0264-6021:3520435

Jensen PE, Willows RD, Petersen BL, et al (1996) Structural genes for Mg-chelatase subunits in barley: Xantha-f, -g and -h. Mol Gen Genet 250:383–394. doi:

10.1007/s004380050090

Kobayashi K, Kawabata M, Hisano K, et al (2012) Identification and characterization of the RNA binding surface of the pentatricopeptide repeat protein. Nucleic Acids Res 40:2712–2723. doi: 10.1093/nar/gkr1084

Kobayashi Y, Matsuo M, Sakamoto K, et al (2007) Two RNA editing sites with cis-acting elements of moderate sequence identity are recognized by an identical site-recognition protein in tobacco chloroplasts. Nucleic Acids Res 36:311–318. doi: 10.1093/nar/gkm1026

Kodama Y, Suetsugu N, Kong S-G, Wada M (2010) Two interacting coiled-coil proteins, WEB1 and PMI2, maintain the chloroplast photorelocation movement velocity in Arabidopsis. Proc Natl Acad Sci 107:19591–19596. doi: 10.1073/pnas.1007836107

Kong SG, Wada M (2011) New insights into dynamic actin-based chloroplast photorelocation movement. Mol Plant 4:771–781. doi: 10.1093/mp/ssr061

Kotera E, Tasaka M, Shikanai T (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. Nature 433:326–330. doi: 10.1038/nature03229

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. doi: gb-2009-10-3-r25 [pii]\n10.1186/gb-2009-10-3-r25

Le DT, Aldrich DL, Valliyodan B, et al (2012) Evaluation of Candidate Reference Genes for Normalization of Quantitative RT-PCR in Soybean Tissues under Various Abiotic Stress Conditions. PLoS One. doi: 10.1371/journal.pone.0046487

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. doi: 10.1186/s13059-014-0550-8

Lurin C, Andrés C, Aubourg S, et al (2004) Genome-Wide Analysis of Arabidopsis Pentatricopeptide Repeat Proteins Reveals Their Essential Role in Organelle Biogenesis. Plant Cell 16:2089–2103. doi: 10.1105/tpc.104.022236.discover

Maier RM, Neckermann K, Igloi GL, Kössel H (1995) Complete Sequence of the Maize Chloroplast Genome: Gene Content, Hotspots of Divergence and Fine Tuning of Genetic Information by Transcript Editing. J Mol Biol 251:614–628. doi: 10.1006/jmbi.1995.0460

Miura E, Kato Y, Matsushima R, et al (2007) The Balance between Protein Synthesis and

Degradation in Chloroplasts Determines Leaf Variegation in Arabidopsis yellow variegated Mutants. Plant Cell Online 19:1313–1328. doi: 10.1105/tpc.106.049270

Miyata Y, Sugita M (2004) Tissue- and stage-specific RNA editing of rps14 transcripts in moss (Physcomitrella patens) chloroplasts. J Plant Physiol 161:113–115. doi: 10.1078/0176-1617-01220

Mochizuki N, Brusslan JA, Larkin R, et al (2001) Arabidopsis genomes uncoupled 5 (GUN5) mutant reveals the involvement of Mg-chelatase H subunit in plastid-to-nucleus signal transduction. Proc Natl Acad Sci 98:2053–2058. doi: 10.1073/pnas.98.4.2053

Mozzo M, Passarini F, Bassi R, et al (2008) Photoprotection in higher plants: The putative quenching site is conserved in all outer light-harvesting complexes of Photosystem II. Biochim Biophys Acta - Bioenerg 1777:1263–1267. doi: 10.1016/j.bbabio.2008.04.036

Naested H (2004) Arabidopsis VARIEGATED 3 encodes a chloroplast-targeted, zinc-finger protein required for chloroplast and palisade cell development. J Cell Sci 117:4807–4818. doi: 10.1242/jcs.01360

Nakajima Y, Mulligan M (2001) Heat stress results in incomplete C-to-U editing of maize chloroplast mRNAs and correlates with changes in chloroplast transcription rate. Curr Genet 40:209–213. doi: 10.1007/s002940100249

Nakajima Y, Mulligan M, Mulligan RM (2001) Heat stress results in incomplete C-to-U editing of maize chloroplast mRNAs and correlates with changes in chloroplast transcription rate. Curr Genet 40:209–213. doi: 10.1007/s002940100249

Neuwirt J, Takenaka M, Merwe JA van der, Brennicke A (2005) An in vitro RNA editing system from cauliflower mitochondria: Editing site recognition parameters can vary in different plant species. RNA 11:1563–1570. doi: 10.1261/rna.2740905

O'Toole N, Hattori M, Andres C, et al (2008) On the expansion of the pentatricopeptide repeat gene family in plants. Mol Biol Evol 25:1120–1128. doi: 10.1093/molbev/msn057

Ohtani M, Demura T, Sugiyama M (2013) Arabidopsis ROOT INITIATION DEFECTIVE1, a DEAH-Box RNA Helicase Involved in Pre-mRNA Splicing, Is Essential for Plant Development. Plant Cell 25:2056–2069. doi: 10.1105/tpc.113.111922

Oikawa K (2003) CHLOROPLAST UNUSUAL POSITIONING1 Is Essential for Proper Chloroplast Positioning. PLANT CELL ONLINE 15:2805–2815. doi: 10.1105/tpc.016428

Okuda K, Shikanai T (2012) A pentatricopeptide repeat protein acts as a site-specificity factor at multiple RNA editing sites with unrelated cis-acting elements in plastids. Nucleic Acids Res 40:5052–5064. doi: 10.1093/nar/gks164

Olinares PDB, Ponnala L, van Wijk KJ (2010) Megadalton Complexes in the Chloroplast Stroma of Arabidopsis thaliana Characterized by Size Exclusion Chromatography, Mass Spectrometry, and Hierarchical Clustering. Mol Cell Proteomics 9:1594–1615. doi: 10.1074/mcp.M000038-MCP201

Ramos-Vega M, Guevara-García A, Llamas E, et al (2015) Functional analysis of the Arabidopsis thaliana CHLOROPLAST BIOGENESIS 19 pentatricopeptide repeat editing protein. New Phytol 208:430–441. doi: 10.1111/nph.13468

Rodrigues NF, Christoff AP, da Fonseca GC, et al (2017a) Unveiling Chloroplast RNA Editing Events Using Next Generation Small RNA Sequencing Data. Front Plant Sci 8:1686. doi: 10.3389/fpls.2017.01686

Rodrigues NF, Fonseca GC da, Kulcheski FR, Margis R (2017b) Salt stress affects mRNA editing in soybean chloroplasts. Genet Mol Biol 40:200–208. doi: 10.1590/1678-4685-gmb-2016-0055

Rudinger M, Polsakiewicz M, Knoop V (2008) Organellar RNA Editing and Plant-Specific Extensions of Pentatricopeptide Repeat Proteins in Jungermanniid but not in Marchantiid Liverworts. Mol Biol Evol 25:1405–1414. doi: 10.1093/molbev/msn084

Ruwe H, Castandet B, Schmitz-Linneweber C, Stern DB (2013) Arabidopsis chloroplast quantitative editotype. FEBS Lett 587:1429–1433. doi: 10.1016/j.febslet.2013.03.022

Schallenberg-Rüdinger M, Knoop V (2016) Coevolution of Organelle RNA Editing and Nuclear Specificity Factors in Early Land Plants. Adv Bot Res 78:37–93. doi: 10.1016/bs.abr.2016.01.002

Schmidt von Braun S, Schleiff E (2008) The chloroplast outer membrane protein CHUP1 interacts with actin and profilin. Planta 227:1151–1159. doi: 10.1007/s00425-007-0688-7

Shi C, Hu N, Huang H, et al (2012) An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. PLoS One 7:1–7. doi:

10.1371/journal.pone.0031468

Shikanai T (2006) RNA editing in plant organelles: machinery, physiological function and evolution. Cell Mol Life Sci 63:698–708. doi: 10.1007/s00018-005-5449-9

Singh CR, Lee B, Udagawa T, et al (2006) An eIF5/eIF2 complex antagonizes guanine nucleotide exchange by eIF2B during translation initiation. EMBO J 25:4537–4546. doi: 10.1038/sj.emboj.7601339

Small ID, Peeters N (2000) The PPR motif - a TPR-related motif prevalent in plant organellar proteins. Trends Biochem Sci 25:45–47. doi: 10.1016/S0968-0004(99)01520-0

Sun T, Bentolila S, Hanson MR (2016) The Unexpected Diversity of Plant Organelle RNA Editosomes. Trends Plant Sci 21:962–973. doi: 10.1016/j.tplants.2016.07.005

Sun T, Germain A, Giloteaux L, et al (2013) An RNA recognition motif-containing protein is required for plastid RNA editing in Arabidopsis and maize. Proc Natl Acad Sci 110:E1169–E1178. doi: 10.1073/pnas.1220162110

Sun T, Shi X, Friso G, et al (2015) A Zinc Finger Motif-Containing Protein Is Essential for Chloroplast RNA Editing. PLOS Genet 11:e1005028. doi: 10.1371/journal.pgen.1005028

Tadini L, Pesaresi P, Kleine T, et al (2016) GUN1 Controls Accumulation of the Plastid Ribosomal Protein S1 at the Protein Level and Interacts with Proteins Involved in Plastid Protein Homeostasis. Plant Physiol 170:1817–30. doi: 10.1104/pp.15.02033

Takenaka M, Zehrmann A, Verbitskiy D, et al (2012) Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. Proc Natl Acad Sci 109:5104–5109. doi: 10.1073/pnas.1202452109

Takenaka M, Zehrmann A, Verbitskiy D, et al (2013) RNA Editing in Plants and Its Evolution. Annu Rev Genet 47:335–352. doi: 10.1146/annurev-genet-111212-133519

Tamura K, Stecher G, Peterson D, et al (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729. doi: 10.1093/molbev/mst197

Tan J, Tan Z, Wu F, et al (2014) A novel chloroplast-localized pentatricopeptide repeat protein involved in splicing affects chloroplast development and abiotic stress response in rice. Mol Plant 7:1329–1349. doi: 10.1093/mp/ssu054

Terzi LC, Simpson GG (2009) Arabidopsis RNA immunoprecipitation. Plant J 59:163–168. doi: 10.1111/j.1365-313X.2009.03859.x

Tillich M, Hardel SL, Kupsch C, et al (2009) Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. Proc Natl Acad Sci 106:6002–6007. doi: 10.1073/pnas.0808529106

Tillich M, Lehwark P, Morton BR, Maier UG (2006a) The evolution of chloroplast RNA editing. Mol Biol Evol 23:1912–1921. doi: 10.1093/molbev/msl054

Tillich M, Poltnigg P, Kushnir S, Schmitz-Linneweber C (2006b) Maintenance of plastid RNA editing activities independently of their target sites. EMBO Rep 7:308–313. doi: 10.1038/sj.embor.7400619

Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet 5:123–135. doi: 10.1038/nrg1271

Tseng CC, Lee CJ, Chung YT, et al (2013) Differential regulation of Arabidopsis plastid gene expression and RNA editing in non-photosynthetic tissues. Plant Mol Biol 82:375–392. doi: 10.1007/s11103-013-0069-5

Umate P (2010) Genome-wide analysis of the family of light-harvesting chlorophyll a/b-binding proteins in Arabidopsis and rice. Plant Signal Behav 5:1537–1542. doi: 10.4161/psb.5.12.13410

Van Den Bekerom RJM, Dix PJ, Diekmann K, Barth S (2013) Variations in efficiency of plastidial RNA editing within ndh transcripts of perennial ryegrass (Lolium perenne) are not linked to differences in drought tolerance. AoB Plants 5:plt035-plt035. doi: 10.1093/aobpla/plt035

Vieira LDN, Faoro H, De Freitas Fraga HP, et al (2014) An improved protocol for intact chloroplasts and cpDNA isolation in conifers. PLoS One 9:1–8. doi: 10.1371/journal.pone.0084792

Vu LT, Tsukahara T (2017) C-to-U editing and site-directed RNA editing for the correction of genetic mutations. Biosci Trends 11:243–253. doi: 10.5582/bst.2017.01049

Wortham NC, Proud CG (2015) eIF2B: recent structural and functional insights into a key regulator of translation. Biochem Soc Trans 43:1234–40. doi: 10.1042/BST20150164

Xu R, Zhang S, Huang J, Zheng C (2013) Genome-wide comparative in silico analysis of the RNA helicase gene family in Zea mays and Glycine max: A comparison with Arabidopsis and Oryza sativa. PLoS One 8:1–13. doi: 10.1371/journal.pone.0078982

Yagi Y, Hayashi S, Kobayashi K, et al (2013) Elucidation of the RNA Recognition Code for Pentatricopeptide Repeat Proteins Involved in Organelle RNA Editing in Plants. PLoS One 8:e57286. doi: 10.1371/journal.pone.0057286

Yap A, Kindgren P, Colas des Francs-Small C, et al (2015) AEF1/MPR25 is implicated in RNA editing of plastid atpF and mitochondrial nad5 , and also promotes atpF splicing in Arabidopsis and rice. Plant J 81:661–669. doi: 10.1111/tpj.12756

Zehrmann A, Härtel B, Glass F, et al (2015) Selective homo- and heteromer interactions between the multiple organellar RNA editing factor (MORF) proteins in Arabidopsis thaliana. J Biol Chem 290:6445–6456. doi: 10.1074/jbc.M114.602086

Zhang F, Tang W, Hedtke B, et al (2014) Tetrapyrrole biosynthetic enzyme protoporphyrinogen IX oxidase 1 is required for plastid RNA editing. Proc Natl Acad Sci 111:2023–2028. doi: 10.1073/pnas.1316183111

Table 1. RNA-interacting proteins identified in mass spectrometry assays and their respective probes.

Protein	Accession	RNA probe
Pentatricopeptide repeat proteins		
PPR	Glyma.11g217500	*atpF*-90
PPR	Glyma.19g025700	*atpF*-90
PPR	Glyma.01g016100	*ndhB*-1481
PPR	Glyma.11g111200	*ndhB*-1481, *rps14*-80
PPR	Glyma.02g174500	*rps14*-80
RNA helicases		
DEAD/DEAH box helicase	Glyma.02g119000	*atpF*-90, *ndhB*-1481, *rps14*-80
Helicase, IBR and zinc finger protein domain-containing protein	Glyma.18g014800	*rps14*-80
Translation factors		
Initiation factor (IF-2)	Glyma.08g174200	*atpF*-90
Initiation factor eIF-2B subunit delta (EIF2B4)	Glyma.19g044300	*ndhB*-1481

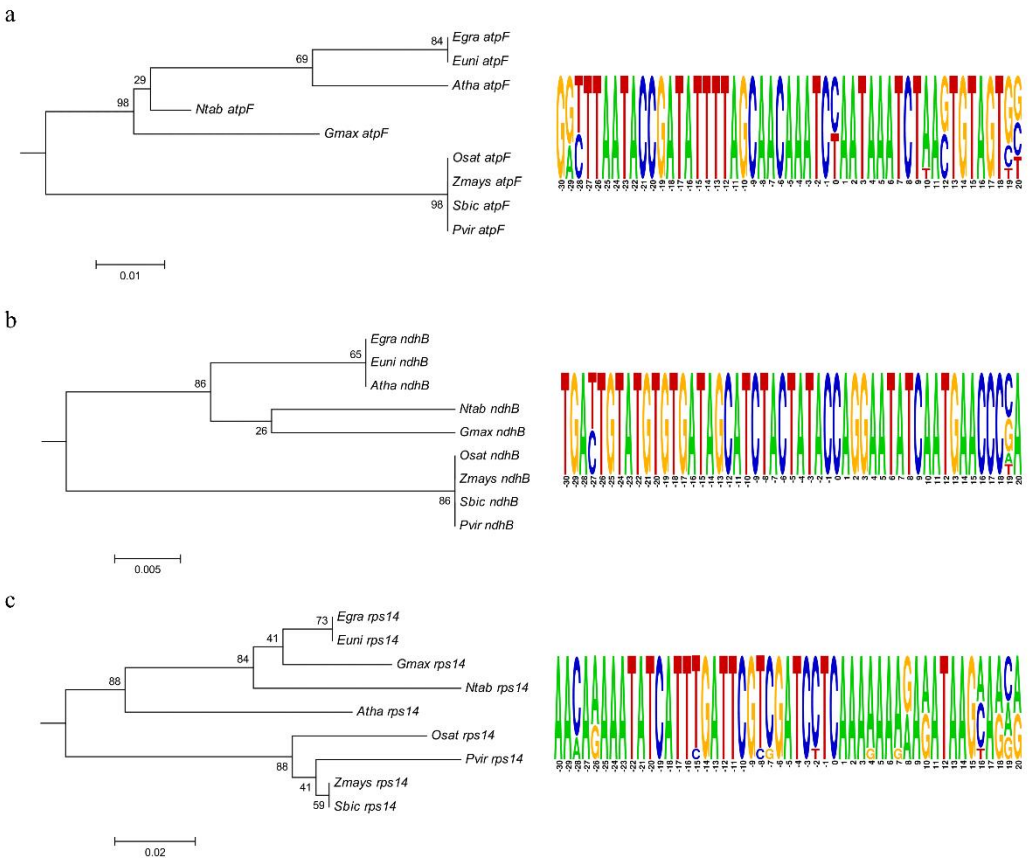

Fig. 1 – Sequence analysis of *cis*-elements. A neighbor-joining tree was created using the p-distance method and the sequence alignment of the region surrounding the (a) *atpF*-90, (b) *ndhB*-1481 and (c) *rps14*-80 editing sites, from -30 to +20 around the edited C (position zero) of *A. thaliana* (*Atha*), *E. uniflora* (*Euni*), *G. max* (*Gmax*), *N. tabacum* (*Ntab*), *O. sativa* (*Osat*), *P. virgatum* (*Pvir*), *S. bicolor* (*Sbic*) and *Z. mays* (*Zmays*). A consensus logo is showed from each alignment.

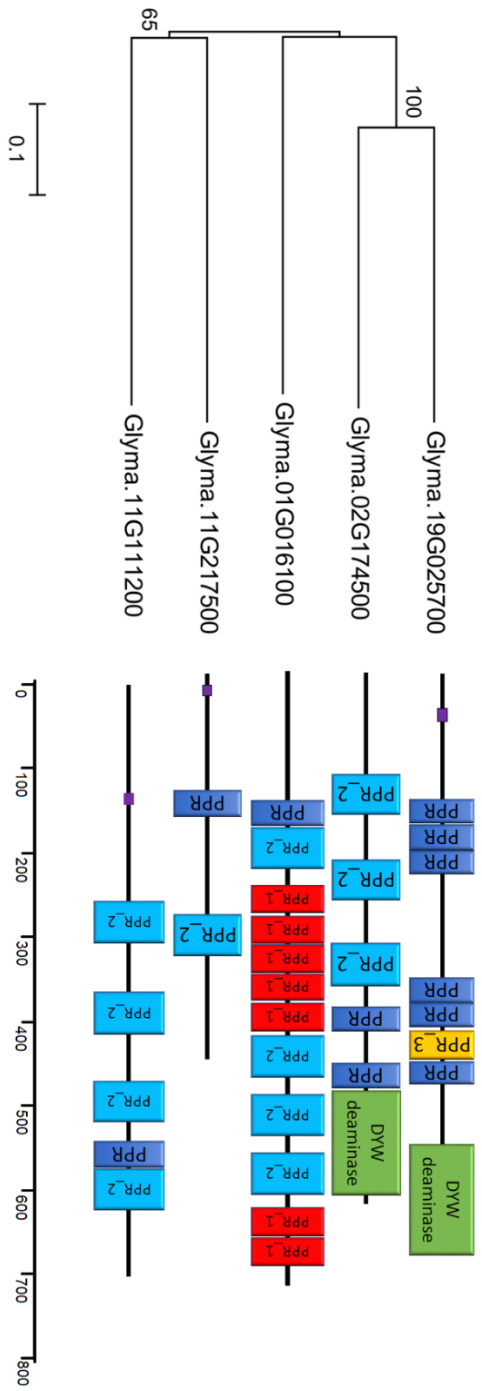

Fig. 2 – Phylogenetic relationship and structural analysis of *G. max* PPR proteins. The phylogenetic tree was inferred using the Neighbor-Joining method. The evolutionary distances were computed using the p-distance method. The protein structures were designed based in PFAM prediction.

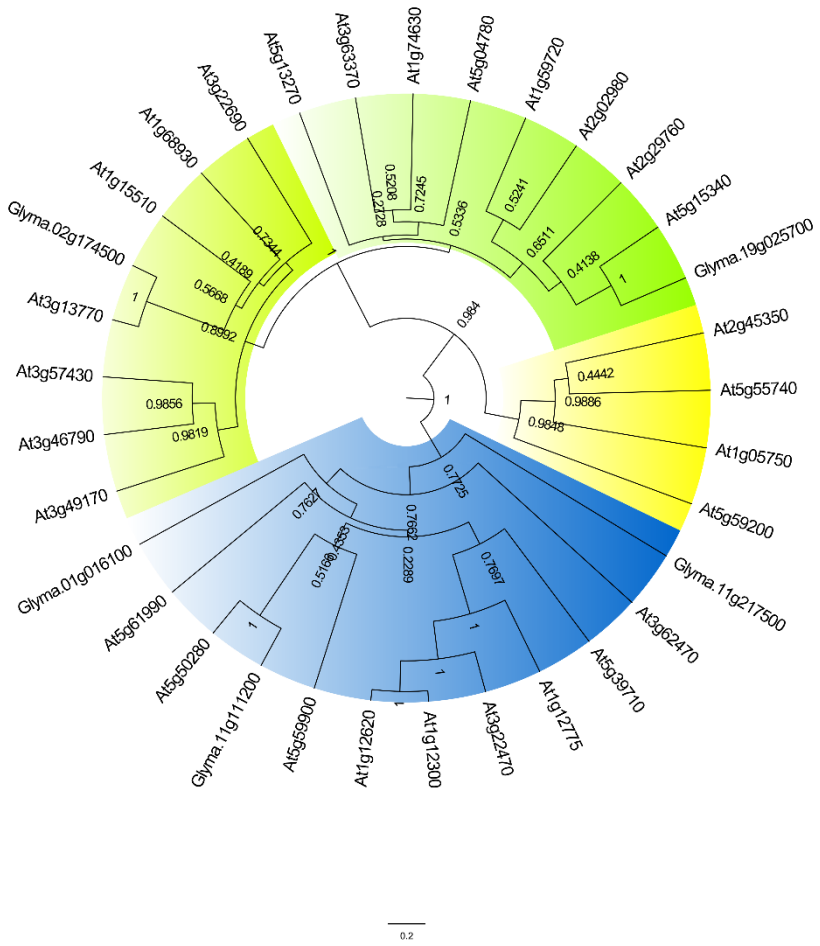

Fig. 3 - Phylogenetic relationship among PPR protein sequences. The phylogenetic analysis was performed with PPR protein sequences from *Arabidopsis thaliana* and *Glycine max*. Posteriori probabilities are labeled above the branches. In blue, PPR P-type proteins; in yellow PPR E-type proteins; in green, PPR DYW-type proteins.

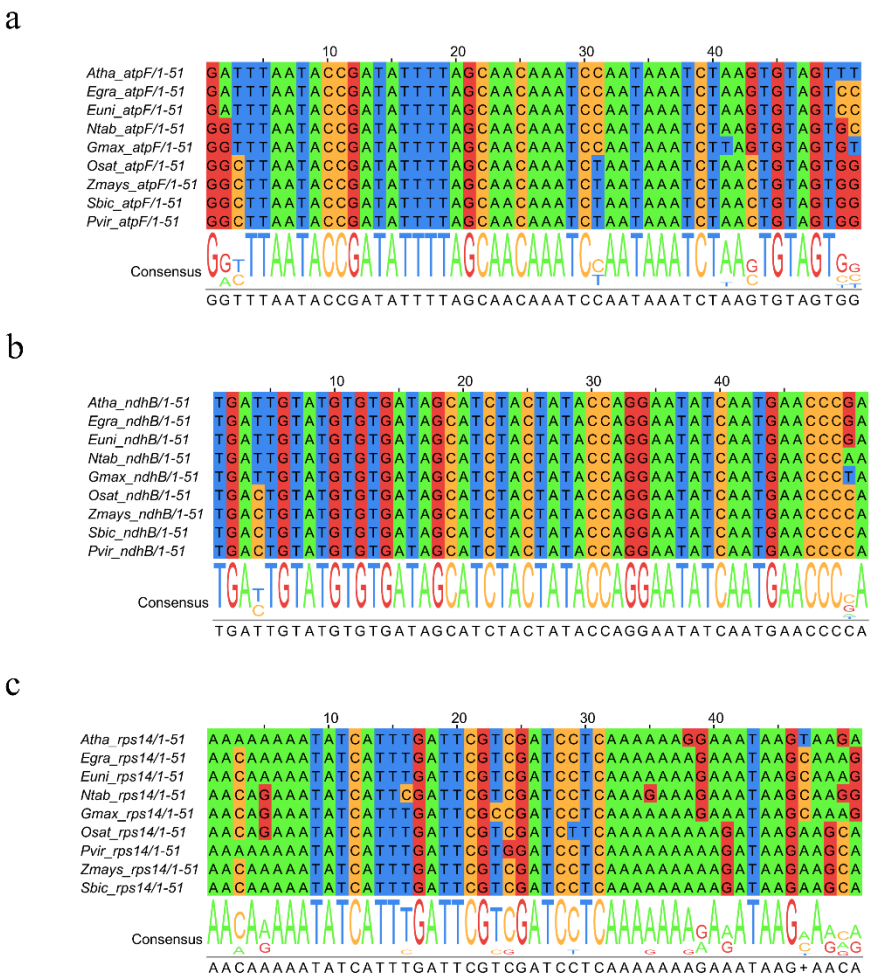

Fig. S1 – Alignment of analyzed *cis*-elements. Sequence alignment of the region surrounding the (a) *atpF*-90, (b) *ndhB*-1481 and (c) *rps14*-80 editing sites. The alignment includes the sequence from -30 to +20 around the edited C (position 31) of *A. thaliana* (*Atha*), *E. uniflora* (*Euni*), *G. max* (*Gmax*), *N. tabacum* (*Ntab*), *O. sativa* (*Osat*), *P. virgatum* (*Pvir*), *S. bicolor* (*Sbic*) and *Z. mays* (*Zmays*). Above each alignment a consensus logo is showed.

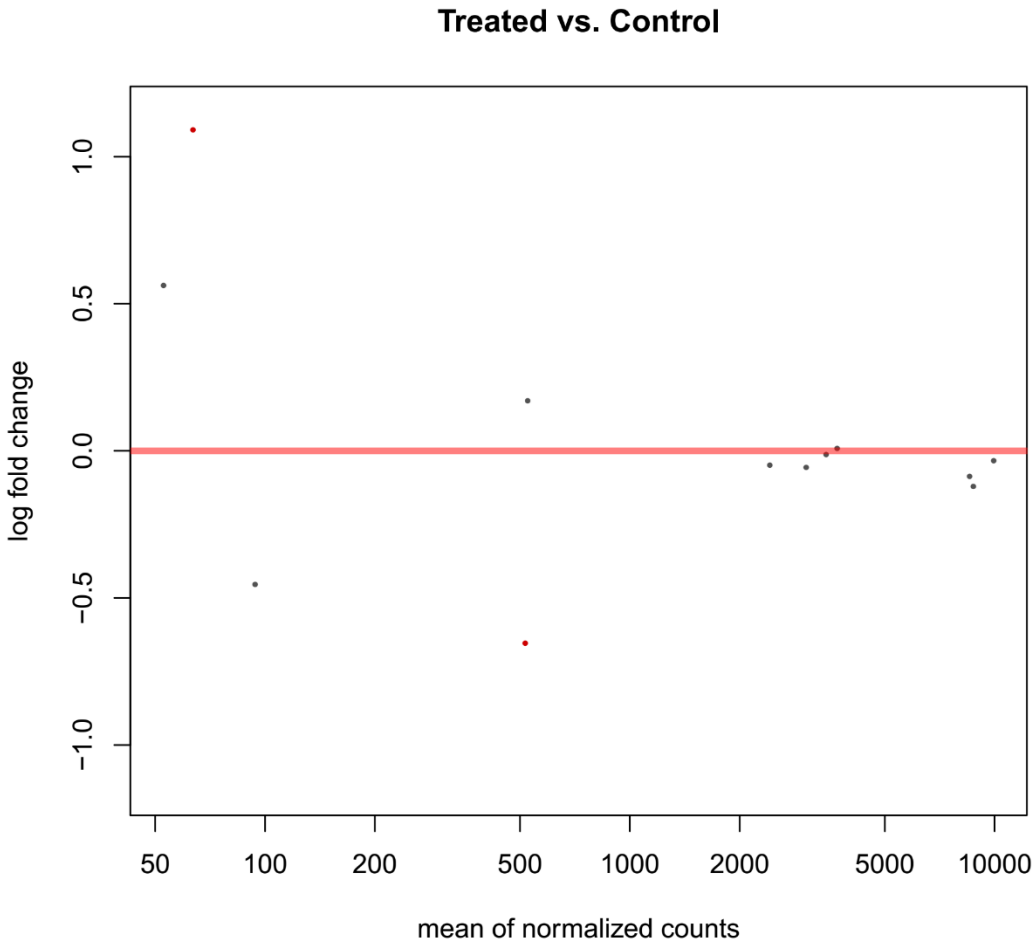

Fig. S2 - Mean difference (M) vs. average expression (A) plot of differential gene expression in salt treated versus control soybean leaves. The red dots indicate differentially expressed genes.

6. DISCUSSÃO E CONSIDERAÇÕES FINAIS

Atualmente, alguns estudos que objetivam a caracterização de genomas de cloroplastos têm também utilizado predições *in silico* dos sítios de edição de RNA, visando entender relações filogenéticas (Silva et al. 2016; Huang et al. 2017). Softwares que realizam essa predição, como PREP-Suite (Mower 2009) e PREPACT (Lenz et al. 2010) têm sido atualizados e aperfeiçoados. Todavia a predição é baseada em relações filogenéticas já descritas, não permitindo a uma predição de novos sítios. Ainda assim, após a predição, tais sítios precisam de confirmação através de outros experimentos. ChloroSeq, um programa que avalia a taxa de edição utilizando bibliotecas de sequenciamento de RNA foi recentemente disponibilizado (Castandet et al. 2016); todavia, os sítios de edição analisados são pré-estabelecidos, não sendo permitida a predição de sítios conservados ou novos (Smith and Sanitá Lima 2016). Portanto, programas ou pipelines para predição de novos sítios de edição que utilizem dados das plataformas de sequenciamento já disponíveis e que permitam uma análise quantitativa da edição nesses sítios não tinham sido desenvolvidos.

No primeiro artigo desenvolvemos e demonstramos um método que permite, utilizando-se dados de sequenciamento de sRNAs, a predição de novos sítios de edição e confirmação de sítios de edição preditos por outros programas. A confirmação desses sítios por meio de experimentos de RT-qPCR demonstraram a confiabilidade do nosso método de identificação. Nesse trabalho, apresentamos pela primeira vez, um amplo conjunto de sítios de edição em soja, incluindo sítios de edição espécie-específicos. O método descrito também permite utilizar sequenciamento de RNA (RNA-seq) para identificação e quantificação de sítios de edição.

O método desenvolvido nesse trabalho foi utilizado para avaliar os efeitos do estresse salino na edição de RNA de cloroplastos. O impacto de estresses abióticos no processo de edição tem sido demonstrado em alguns transcritos de genes específicos. Transcritos do gene *ndhB*, uma subunidade da NAD(P)H desidrogenase de cloroplasto, quando sob estresse causado por calor, apresentam edição incompleta dos sítios, resultando em defeitos de splicing (Nakajima et al. 2001). A edição de RNA em outros transcritos de cloroplastos também respondem sensivelmente ao calor (Karcher and Bock 2002). Além de uma redução global na eficiência da edição e splicing, uma maior abundância de transcritos de cloroplasto, incluindo intergênicos e antisensos é verificada, provavelmente resultado uma redução na atividade das proteínas de metabolismo de RNA.

O segundo artigo demonstrou efeitos da salinidade na edição de RNA em cloroplastos de soja. O estresse salino levou a um aumento da edição de RNA em alguns transcritos dos componentes da cadeia de transferência de elétrons, fotossistemas e complexos de tradução. Já tem sido descrito que a salinidade reduz a atividade fotossintética pela inibição do PSII (Parida and Das 2005; Zheng et al. 2009; Allu et al. 2014). Estudos proteômicos demonstraram o aumento da tradução em diversos genes, incluindo fatores de tradução e genes de cloroplastos, como *psaB* e *rps12* durante estresses abióticos (Salekdeh et al. 2002; Jiang et al. 2007; Hashiguchi et al. 2009). Portanto, o aumento na edição dos transcritos identificados podem ser uma resposta para manutenção da homeostase através da atividade funcional das proteínas em resposta ao estresse salino. Entender alterações adaptativas que otimizam funções básicas como fotossíntese, metabolismo de RNA ou tradução de transcritos plastidiais pode contribuir na geração de cultivares que sejam tolerantes à estresses abióticos (Tonti-Filippini et al. 2017).

No terceiro artigo, fatores associados a *cis*-elementos de três sítios de edição em soja, *atpF*-92, *ndhB*-1481 e *rps14*-80, foram isolados por coprecipitação com sondas de RNA biotinilados e identificados utilizando-se espectrometria de massas. No total, cinco PPRs foram identificadas, além de outras proteínas de ligação ao RNA. A predição de endereçamento indicou cloroplastos e mitocôndrias como alvo dessas PPRs. Ensaios de localização são necessários para confirmar as predições.

Estudos que identificam fatores de edição usaram screening de mutantes e imunoprecipitação dos alvos marcados com epítopos. Com mutantes disponíveis em Arabidopsis, a caracterização dos níveis de edição permitiram identificar os sítios de edição regulados pelos fatores em questão (Kotera et al. 2005; Sun et al. 2013). Sendo soja uma espécie modelo com métodos de transformação mais laboriosos (Homrich et al. 2012), o silenciamento das PPRs identificadas em um sistema transiente e a caracterização do nível de expressão nos respectivos sítios de edição podem confirmar a associação identificada pela espectrometria de massas de forma mais rápida. No segundo artigo, demonstramos o aumento da edição no sítio *rps14*-80. Um dos fatores identificados que está associado a esse sítio, Glyma.02g174500, uma PPR-DYW, apresentou um aumento da expressão nas bibliotecas de folhas tratadas com sal. É possível que o aumento da taxa de edição e da expressão do gene Glyma.02g174500 estejam relacionados por ser seu fator de edição

cognato. Dessa forma, esse gene poderia ser um alvo para estudos de tolerância a estresses abióticos.

Proteínas da família PPR são ótimas candidatas a fatores de tolerância a estresses; algumas PPRs têm sido demonstradas como participantes da edição de RNA em organelas e também necessárias para respostas a estresses abióticos (Liu et al. 2016). SLG1 é uma PPR pertence à subclasse E+ endereçada para mitocôndria. O mutante *slg1*, tem um defeito na edição de RNA do sítio *nad3*-250, da NADH desidrogenase do complexo I de mitocôndrias, exibindo crescimento lento e fenótipo de desenvolvimento atrasado. *slg1* também demonstra uma maior sensibilidade a vários estresses abióticos (Yuan and Liu 2012). SLO2, também uma PPR-E+, participa da edição de RNA de sete sítios em mitocôndrias (Zhu et al. 2012). Os mutantes *slo2* são hipersensíveis ao estresse salino e osmótico durante o estágio de germinação, enquanto plantas adultas mostram aumento da tolerância à seca e ao sal (Zhu et al. 2014).

WSL é uma PPR direcionada para cloroplastos em arroz, que está envolvida com splicing de transcritos do gene *rpl2*. O mutante *wsl* mostra sensibilidade aumentada à salinidade e acumula mais H_2O_2 do que o tipo selvagem. Dessa forma, a redução de eficiência da tradução pode afetar a resposta do mutante ao estresse abiótico, o que é corroborado pelos nossos dados pelo aumento de edição para manutenção da resposta ao estresse (Tan et al. 2014). SOAR1, uma PPR duplamente marcada para endereçamento para o núcleo e citoplasma, regula negativamente a sinalização de ABA, é um regulador positivo da resposta da planta aos estresses abióticos. A superexpressão de SOAR1 resulta na resistência da germinação das sementes a uma salinidade extremamente alta e na insensibilidade ao sal em plantas maduras, em contraste com a hipersensibilidade ao sal do mutante *soar1*. Alterações na expressão SOAR1 alteram a expressão de um subconjunto de genes envolvidos em respostas a estresse osmótico, salino e de frio (Jiang et al. 2015). Portanto, proteínas da família PPR são ótimos alvos em estudos de tolerância a estresses abióticos, estando envolvidas em edição, ou não.

Além de todo o conhecimento gerado sobre edição de RNA de cloroplastos em soja, a influência de estresses abióticos e a identificação dos primeiros fatores de edição nessa espécie, os métodos e dados do presente trabalho também contribuirão para a descoberta de novos fatores de edição em plantas não modelos. A prospecção desses novos fatores pode ser de grande importância permitindo a identificação de proteínas de resposta a tolerância a

estresses abióticos que podem ser utilizadas em programas de melhoramento de cultivares de importância econômica.

7. REFERÊNCIAS

Albertin CB, Simakov O, Mitros T, Wang ZY, Pungor JR, Edsinger-Gonzales E, Brenner S, Ragsdale CW and Rokhsar DS (2015) The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature 524:220–224. doi: 10.1038/nature14668

Allu AD, Soja AM, Wu A, Szymanski J and Balazadeh S (2014) Salt stress and senescence: identification of cross-talk regulatory components. J Exp Bot 65:3993–4008. doi: 10.1093/jxb/eru173

Alvarez-Pizarro JC, Gomes-Filho E, de Lacerda CF, Alencar NLM and Prisco JT (2009) Salt-induced changes on H+-ATPase activity, sterol and phospholipid content and lipid peroxidation of root plasma membrane from dwarf-cashew (Anacardium occidentale L.) seedlings. Plant Growth Regul 59:125–135. doi: 10.1007/s10725-009-9395-7

Andrés-Colás N, Zhu Q, Takenaka M, De Rybel B, Weijers D and Van Der Straeten D (2017) Multiple PPR protein interactions are involved in the RNA editing system in *Arabidopsis* mitochondria and plastids. Proc Natl Acad Sci 114:201705815. doi: 10.1073/pnas.1705815114

Asano T, Miyao A, Hirochika H, Kikuchi S and Kadowaki K (2013) A pentatricopeptide repeat gene of rice is required for splicing of chloroplast transcripts and RNA editing of ndhA. Plant Biotechnol 30:57–64. doi: 10.5511/plantbiotechnology.12.1217a

Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, DePamphilis C, Albert V a, Aono N, Aoyama T, Ambrose BA et al. (2011) The Selaginella Genome Identifies Genetic Changes Associated with the Evolution of Vascular Plants. Science (80-) 332:960–963. doi: 10.1126/science.1203810

Barkan A, Rojas M, Fujii S, Yap A, Chong YS, Bond CS and Small I (2012) A Combinatorial Amino Acid Code for RNA Recognition by Pentatricopeptide Repeat Proteins. PLoS Genet 8:e1002910. doi: 10.1371/journal.pgen.1002910

Barkan A and Small I (2014) Pentatricopeptide Repeat Proteins in Plants. Annu Rev Plant Biol 65:415–442. doi: 10.1146/annurev-arplant-050213-040159

Bartel DP (2004) MicroRNAs. Cell 116:281–297. doi: 10.1016/S0092-8674(04)00045-5

Bayer-Császár E, Haag S, Jörg A, Glass F, Härtel B, Obata T, Meyer EH, Brennicke A and Takenaka M (2017) The conserved domain in MORF proteins has distinct affinities to the PPR and E elements in PPR RNA editing factors. Biochim Biophys Acta - Gene Regul Mech 1860:813–828. doi: 10.1016/j.bbagrm.2017.05.004

Benne R, Van Den Burg J, Brakenhoff JPJ, Sloof P, Van Boom JH and Tromp MC (1986) Major transcript of the frameshifted coxll gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell 46:819–826. doi: 10.1016/0092-8674(86)90063-2

Bentolila S, Heller WP, Sun T, Babina a. M, Friso G, van Wijk KJ and Hanson MR (2012)

RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. Proc Natl Acad Sci 109:E1453–E1461. doi: 10.1073/pnas.1121465109

Bentolila S, Oh J, Hanson MR and Bukowski R (2013) Comprehensive High-Resolution Analysis of the Role of an Arabidopsis Gene Family in RNA Editing. PLoS Genet 9:e1003584. doi: 10.1371/journal.pgen.1003584

Blanc V and Davidson NO (2010) APOBEC-1-mediated RNA editing. Wiley Interdiscip Rev Syst Biol Med 2:594–602. doi: 10.1002/wsbm.82

Bock R, Hermann M and Kössel H (1996) In vivo dissection of cis-acting determinants for plastid RNA editing. EMBO J 15:5052–5059.

Bock R, Kössel H and Maliga P (1994) Introduction of a heterologous editing site into the tobacco plastid genome: the lack of RNA editing leads to a mutant phenotype. EMBO J 13:4623–8.

Borges F and Martienssen RA (2015) The expanding world of small RNAs in plants. Nat Rev Mol Cell Biol 16:727–741. doi: 10.1038/nrm4085

Boussardon C, Avon a, Kindgren P, Bond CS, Challenor M, Lurin C and Small I (2014) The cytidine deaminase signature HxE(x)(n)CxxC of DYW1 binds zinc and is necessary for RNA editing of ndhD-1. New Phytol 203:1090–1095. doi: 10.1111/nph.12928

Brehme N, Zehrmann A, Verbitskiy D, HÃ¤rtel B and Takenaka M (2014) Mitochondrial RNA editing PPR proteins can tolerate protein tags at E as well as at DYW domain termini. Front Plant Sci 5:3–6. doi: 10.3389/fpls.2014.00127

Broich SL and Palmer RG (1980) A cluster analysis of wild and domesticated soybean phenotypes. Euphytica 29:23–32. doi: 10.1007/BF00037246

Cannon SB and Shoemaker RC (2012) Evolutionary and comparative analyses of the soybean genome. Breed Sci 61:437–444. doi: 10.1270/jsbbs.61.437

Carrillo C (1997) RNA editing status of nad7 intron domains in wheat mitochondria. Nucleic Acids Res 25:403–409. doi: 10.1093/nar/25.2.403

Castandet B, Choury D, Bégu D, Jordana X and Araya A (2010) Intron RNA editing is essential for splicing in plant mitochondria. Nucleic Acids Res 38:7112–7121. doi: 10.1093/nar/gkq591

Castandet B, Hotto AM, Strickler SR and Stern DB (2016) ChloroSeq, an Optimized Chloroplast RNA-Seq Bioinformatic Pipeline, Reveals Remodeling of the Organellar Transcriptome Under Heat Stress. G3 (Bethesda) 6:2817–27. doi: 10.1534/g3.116.030783

Chateigner-Boutin A-L, Colas des Francs-Small C, Fujii S, Okuda K, Tanz SK and Small I (2013) The E domains of pentatricopeptide repeat proteins from different organelles are not

functionally equivalent for RNA editing. Plant J 74:935–945. doi: 10.1111/tpj.12180

Chateigner-Boutin A-L and Small I (2010) Plant RNA editing. RNA Biol 7:213–219. doi: 10.4161/rna.7.2.11343

Chaves MM, Flexas J and Pinheiro C (2009) Photosynthesis under drought and salt stress: Regulation mechanisms from whole plant to cell. Ann Bot 103:551–560. doi: 10.1093/aob/mcn125

Chawla G and Sokol NS (2014) ADAR mediates differential expression of polycistronic microRNAs. Nucleic Acids Res 42:5245–5255. doi: 10.1093/nar/gku145

Chen H, Yu Y, Chen X, Zhang Z, Gong C, Li J and Wang A (2015) Plastid DNA insertions in plant nuclear genomes: the sites, abundance and ages, and a predicted promoter analysis. Funct Integr Genomics 15:131–139. doi: 10.1007/s10142-014-0422-z

Cheng S, Gutmann B, Zhong X, Ye Y, Fisher MF, Bai F, Castleden I, Song Y, Song B, Huang J et al. (2016) Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. Plant J 85:532–547. doi: 10.1111/tpj.13121

Choi H-K, Mun J-H, Kim D-J, Zhu H, Baek J-M, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB et al. (2004) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci 101:15289–15294. doi: 10.1073/pnas.0402251101

Coego A (2005) An Arabidopsis Homeodomain Transcription Factor, OVEREXPRESSOR OF CATIONIC PEROXIDASE 3, Mediates Resistance to Infection by Necrotrophic Pathogens. PLANT CELL ONLINE 17:2123–2137. doi: 10.1105/tpc.105.032375

Corneille S, Lutz K and Maliga P (2000) Conservation of RNA editing between rice and maize plastids: are most editing events dispensable? Mol Gen Genet MGG 264:419–424. doi: 10.1007/s004380000295

Cramer GR, Urano K, Delrot S, Pezzotti M and Shinozaki K (2011) Effects of abiotic stress on plants: a systems biology perspective. BMC Plant Biol 11:163. doi: 10.1186/1471-2229-11-163

Crooks GE (2004) WebLogo: A Sequence Logo Generator. Genome Res 14:1188–1190. doi: 10.1101/gr.849004

Darriba D, Taboada GL, Doallo R and Posada D (2011) ProtTest-HPC: Fast selection of best-fit models of protein evolution. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 6586 LNCS:177–184. doi: 10.1007/978-3-642-21878-1_22

Delannoy E, Le Ret M, Faivre-Nitschke E, Estavillo GM, Bergdoll M, Taylor NL, Pogson BJ, Small I, Imbault P and Gualberto JM (2009) Arabidopsis tRNA Adenosine Deaminase Arginine Edits the Wobble Nucleotide of Chloroplast tRNAArg(ACG) and Is Essential for

Efficient Chloroplast Translation. PLANT CELL ONLINE 21:2058–2071. doi: 10.1105/tpc.109.066654

Doyle JJ, Doyle JL, Rauscher JT and Brown AHD (2004) Evolution of the perennial soybean polyploid complex (Glycine subgenus Glycine): A study of contrasts. Biol J Linn Soc 82:583–597. doi: 10.1111/j.1095-8312.2004.00343.x

Drummond AJ and Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214. doi: 10.1186/1471-2148-7-214

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. doi: 10.1093/nar/gkh340

FAO and ITPS (2015) Global soil status, processes and trends. Status of the World's Soil Resources (SWSR) – Main Report. Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils, Rome, Italy, pp 98–168

Feagin JE, Abraham JM and Stuart K (1988) Extensive editing of the cytochrome c oxidase III transcript in Trypanosoma brucei. Cell 53:413–422. doi: 10.1016/0092-8674(88)90161-4

Feller U, Anders I and Demirevska K (2008) Degradation of Rubisco and other Chloroplast Proteins Under Abiotic Stress. Plant Physiol 34:5–18.

Fiebig A, Stegemann S and Bock R (2004) Rapid evolution of RNA editing sites in a small non-essential plastid gene. Nucleic Acids Res 32:3615–22. doi: 10.1093/nar/gkh695

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A et al. (2016) The Pfam protein families database: Towards a more sustainable future. Nucleic Acids Res 44:D279–D285. doi: 10.1093/nar/gkv1344

Fisk DG, Walker MB and Barkan A (1999) Molecular cloning of the maize gene crp1 reveals similarity between regulators of mitochondrial and chloroplast gene expression. EMBO J 18:2621–2630. doi: 10.1093/emboj/18.9.2621

Fujii S and Small I (2011) The evolution of RNA editing and pentatricopeptide repeat genes. New Phytol 191:37–47. doi: 10.1111/j.1469-8137.2011.03746.x

García-Andrade J, Ramírez V, López A and Vera P (2013) Mediated Plastid RNA Editing in Plant Immunity. PLoS Pathog 9:e1003713. doi: 10.1371/journal.ppat.1003713

Gill SS and Tuteja N (2010) Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. Plant Physiol Biochem 48:909–930. doi: 10.1016/j.plaphy.2010.08.016

Glass F, Härtel B, Zehrmann A, Verbitskiy D and Takenaka M (2015) MEF13 Requires MORF3 and MORF8 for RNA Editing at Eight Targets in Mitochondrial mRNAs in Arabidopsis thaliana. Mol Plant 8:1466–1477. doi: 10.1016/j.molp.2015.05.008

Gomez JM (2003) Location and effects of long-term NaCl stress on superoxide dismutase and ascorbate peroxidase isoenzymes of pea (Pisum sativum cv. Puget) chloroplasts. J Exp Bot 55:119–130. doi: 10.1093/jxb/erh013

Grewe F, Herres S, Viehöver P, Polsakiewicz M, Weisshaar B and Knoop V (2011) A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte Isoetes engelmannii. Nucleic Acids Res 39:2890–2902. doi: 10.1093/nar/gkq1227

Guevara J, Sukerman M and Velasco M (2015) OECD-FAO Agricultural Outlook 2015. doi: 10.1787/agr_outlook-2015-en

Guillaumot D, Lopez-Obando M, Baudry K, Avon A, Rigaill G, Falcon de Longevialle A, Broche B, Takenaka M, Berthomé R, De Jaeger G et al. (2017) Two interacting PPR proteins are major Arabidopsis editing factors in plastid and mitochondria. Proc Natl Acad Sci 114:8877–8882. doi: 10.1073/pnas.1705780114

Guo W, Grewe F and Mower JP (2015) Variable Frequency of Plastid RNA Editing among Ferns and Repeated Loss of Uridine-to-Cytidine Editing from Vascular Plants. PLoS One 10:e0117075. doi: 10.1371/journal.pone.0117075

Haag S, Schindler M, Berndt L, Brennicke A, Takenaka M and Weber G (2017) Crystal structures of the Arabidopsis thaliana organellar RNA editing factors MORF1 and MORF9. Nucleic Acids Res 45:4915–4928. doi: 10.1093/nar/gkx099

Hackett JB, Shi X, Kobylarz AT, Lucas MK, Wessendorf RL, Hines KM, Bentolila S, Hanson MR and Lu Y (2017) An Organelle RNA Recognition Motif Protein is Required for Photosynthetic Subunit psbF Transcript Editing. Plant Physiol 173:pp.01623.2016. doi: 10.1104/pp.16.01623

Hammani K, Bonnard G, Bouchoucha A, Gobert A, Pinker F, Salinas T and Giegé P (2014) Helical repeats modular proteins are major players for organelle gene expression. Biochimie 100:141–150. doi: 10.1016/j.biochi.2013.08.031

Hashiguchi A, Sakata K and Komatsu S (2009) Proteome analysis of early-stage soybean seedlings under flooding stress. J Proteome Res 8:2058–2069. doi: 10.1021/pr801051m

Hayes ML, Dang KN, Diaz MF and Mulligan RM (2015) A Conserved Glutamate Residue in the C-terminal Deaminase Domain of Pentatricopeptide Repeat Proteins Is Required for RNA Editing Activity. J Biol Chem 290:10136–10142. doi: 10.1074/jbc.M114.631630

Hayes ML and Hanson MR (2006) Identification of a sequence motif critical for editing of a tobacco chloroplast transcript. RNA 13:281–288. doi: 10.1261/rna.295607

He Y, Fu J, Yu C, Wang X, Jiang Q, Hong J, Lu K, Xue G, Yan C, James A et al. (2015) Increasing cyclic electron flow is related to Na+ sequestration into vacuoles for salt tolerance in soybean. J Exp Bot 66:6877–6889. doi: 10.1093/jxb/erv392

He Y, Yu C, Zhou L, Chen Y, Liu A, Jin J, Hong J, Qi Y and Jiang D (2014) Rubisco decrease is involved in chloroplast protrusion and Rubisco-containing body formation in soybean (Glycine max.) under salt stress. Plant Physiol Biochem 74:118–124. doi: 10.1016/j.plaphy.2013.11.008

Heller WP, Hayes ML and Hanson MR (2008) Cross-competition in Editing of Chloroplast RNA Transcripts in Vitro Implicates Sharing of Trans-factors between Different C Targets. J Biol Chem 283:7314–7319. doi: 10.1074/jbc.M709595200

Hirose T and Sugiura M (1996) Cis-acting elements and trans-acting factors for accurate translation of chloroplast psbA mRNAs: development of an in vitro translation system from tobacco chloroplasts. EMBO J 15:1687–1695.

Homrich MS, Wiebke-Strohm B, Weber RLM and Bodanese-Zanettini MH (2012) Soybean genetic transformation: a valuable tool for the functional study of genes and the production of agronomically improved plants. Genet Mol Biol 35:998–1010. doi: 10.1590/S1415-47572012000600015

Huang YY, Cho ST, Haryono M and Kuo CH (2017) Complete chloroplast genome sequence of common bermudagrass (Cynodon dactylon (L.) Pers.) and comparative analysis within the family Poaceae. PLoS One 12:1–16. doi: 10.1371/journal.pone.0179055

IBGE (2015) Produção agrícola municipal: culturas temporárias e permanentes, 1st ed. Instituto Brasileiro de Geografia e Estatística - IBGE, Rio de Janeiro

IBGE (2016) Produção agrícola municipal: Culturas temporárias e permanentes, 1st ed. Instituto Brasileiro de Geografia e Estatística - IBGE, Rio de Janeiro

Ichinose M and Sugita M (2016) RNA Editing and Its Molecular Mechanism in Plant Organelles. Genes (Basel) 8:5. doi: 10.3390/genes8010005

Ichinose M, Uchida M and Sugita M (2014) Identification of a pentatricopeptide repeat RNA editing factor in Physcomitrella patens chloroplasts. FEBS Lett 588:4060–4064. doi: 10.1016/j.febslet.2014.09.031

Ikegami A, Yoshimura N, Motohashi K, Takahashi S, Romano PGN, Hisabori T, Takamiya KI and Masuda T (2007) The CHLI1 subunit of Arabidopsis thaliana magnesium chelatase is a target protein of the chloroplast thioredoxin. J Biol Chem 282:19282–19291. doi: 10.1074/jbc.M703324200

Jackowski G., Kacprzak K. and Jansson S. (2001) Identification of Lhcb1 / Lhcb2 / Lhcb3 heterotrimers of the main light- harvesting chlorophyll a / b - protein complex of Photosystem II (LHC II). Biochim Biophys Acta 1504:340–345. doi: 10.1016/S0005-2728(00)00262-0

Jennings MD and Pavitt GD (2014) A new function and complexity for protein translation initiation factor eIF2B. Cell Cycle 13:2660–2665. doi: 10.4161/15384101.2014.948797

Jensen PE, Reid JD and Hunter CN (2000) Modification of cysteine residues in the ChlI and ChlH subunits of magnesium chelatase results in enzyme inactivation. Biochem J 352:435. doi: 10.1042/0264-6021:3520435

Jensen PE, Willows RD, Petersen BL, Vothknecht UC, Stummann BM, Kannangara CG, Von Wettstein D and Henningsen KW (1996) Structural genes for Mg-chelatase subunits in barley: Xantha-f, -g and -h. Mol Gen Genet 250:383–394. doi: 10.1007/s004380050090

Jiang S-C, Mei C, Liang S, Yu Y-T, Lu K, Wu Z, Wang X-F and Zhang D-P (2015) Crucial roles of the pentatricopeptide repeat protein SOAR1 in Arabidopsis response to drought, salt and cold stresses. Plant Mol Biol 88:369–385. doi: 10.1007/s11103-015-0327-9

Jiang Y, Yang B, Harris NS and Deyholos MK (2007) Comparative proteomic analysis of NaCl stress-responsive proteins in Arabidopsis roots. J Exp Bot 58:3591–607. doi: 10.1093/jxb/erm207

Kalaji HM, Govindjee, Bosa K, Kościelniak J and Żuk-Gołaszewska K (2011) Effects of salt stress on photosystem II efficiency and CO2 assimilation of two Syrian barley landraces. Environ Exp Bot 73:64–72. doi: 10.1016/j.envexpbot.2010.10.009

Kamthan A, Chaudhuri A, Kamthan M and Datta A (2015) Small RNAs in plants: recent development and application for crop improvement. Front Plant Sci 6:1–17. doi: 10.3389/fpls.2015.00208

Karcher D and Bock R (2002) Temperature sensitivity of RNA editing and intron splicing reactions in the plastid ndhB transcript. Curr Genet 41:48–52. doi: 10.1007/s00294-002-0278-y

Kim DDY, Kim TTY, Walsh T, Kobayashi Y, Matise TC, Buyske S and Gabriel A (2004) Widespread RNA editing of embedded Alu elements in the human transcriptome. Genome Res 14:1719–25. doi: 10.1101/gr.2855504

Kim MY, Lee S, Van K, Kim T-H, Jeong S-C, Choi I-Y, Kim D-S, Lee Y-S, Park D, Ma J et al. (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (Glycine soja Sieb. and Zucc.) genome. Proc Natl Acad Sci 107:22032–22037. doi: 10.1073/pnas.1009526107

Kim MY, Van K, Kang YJ, Kim KH and Lee S-H (2012) Tracing soybean domestication history: From nucleotide to genome. Breed Sci 61:445–452. doi: 10.1270/jsbbs.61.445

Kindgren P, Yap A, Bond CS and Small I (2015) Predictable Alteration of Sequence Recognition by RNA Editing Factors from Arabidopsis. Plant Cell Online 27:403–416. doi: 10.1105/tpc.114.134189

Kobayashi K, Kawabata M, Hisano K, Kazama T, Matsuoka K, Sugita M and Nakamura T (2012) Identification and characterization of the RNA binding surface of the pentatricopeptide repeat protein. Nucleic Acids Res 40:2712–2723. doi:

10.1093/nar/gkr1084

Kobayashi Y, Matsuo M, Sakamoto K, Wakasugi T, Yamada K and Obokata J (2007) Two RNA editing sites with cis-acting elements of moderate sequence identity are recognized by an identical site-recognition protein in tobacco chloroplasts. Nucleic Acids Res 36:311–318. doi: 10.1093/nar/gkm1026

Koch M, Breithaupt C, Kiefersauer R, Freigang J, Huber R and Messerschmidt A (2004) Crystal structure of protoporphyrinogen IX oxidase: a key enzyme in haem and chlorophyll biosynthesis. EMBO J 23:1720–1728. doi: 10.1038/sj.emboj.7600189

Kodama Y, Suetsugu N, Kong S-G and Wada M (2010) Two interacting coiled-coil proteins, WEB1 and PMI2, maintain the chloroplast photorelocation movement velocity in Arabidopsis. Proc Natl Acad Sci 107:19591–19596. doi: 10.1073/pnas.1007836107

Kollipara KP, Singh RJ and Hymowitz T (1997) Phylogenetic and genomic relationships in the genus Glycine Willd. based on sequences from the ITS region of nuclear rDNA. Genome 40:57–68. doi: 10.1139/g97-008

Komatsu S and Ahsan N (2009) Soybean proteomics and its application to functional analysis. J Proteomics 72:325–336. doi: 10.1016/j.jprot.2008.10.001

Kong SG and Wada M (2011) New insights into dynamic actin-based chloroplast photorelocation movement. Mol Plant 4:771–781. doi: 10.1093/mp/ssr061

Kotera E, Tasaka M and Shikanai T (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. Nature 433:326–330. doi: 10.1038/nature03229

Koyro H-W (2006) Effect of salinity on growth, photosynthesis, water relations and solute composition of the potential cash crop halophyte Plantago coronopus (L.). Environ Exp Bot 56:136–146. doi: 10.1016/j.envexpbot.2005.02.001

Kugita M (2003) RNA editing in hornwort chloroplasts makes more than half the genes functional. Nucleic Acids Res 31:2417–2423. doi: 10.1093/nar/gkg327

Kulcheski FR, de Oliveira LF, Molina LG, Almerão MP, Rodrigues F a, Marcolino J, Barbosa JF, Stolf-Moreira R, Nepomuceno AL, Marcelino-Guimarães FC et al. (2011) Identification of novel soybean microRNAs involved in abiotic and biotic stresses. BMC Genomics 12:307. doi: 10.1186/1471-2164-12-307

Kupsch C, Ruwe H, Gusewski S, Tillich M, Small I and Schmitz-Linneweber C (2012) Arabidopsis Chloroplast RNA Binding Proteins CP31A and CP29A Associate with Large Transcript Pools and Confer Cold Stress Tolerance by Influencing Multiple Chloroplast RNA Processing Steps. Plant Cell 24:4266–4280. doi: 10.1105/tpc.112.103002

Langmead B, Trapnell C, Pop M and Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. doi: gb-2009-10-3-r25 [pii]\n10.1186/gb-2009-10-3-r25

Le DT, Aldrich DL, Valliyodan B, Watanabe Y, van Ha C, Nishiyama R, Guttikonda SK, Quach TN, Gutierrez-Gonzalez JJ, Tran LSP et al. (2012) Evaluation of Candidate Reference Genes for Normalization of Quantitative RT-PCR in Soybean Tissues under Various Abiotic Stress Conditions. PLoS One. doi: 10.1371/journal.pone.0046487

Lee G-A, Crawford GW, Liu L, Sasaki Y and Chen X (2011) Archaeological Soybean (Glycine max) in East Asia: Does Size Matter? PLoS One 6:e26720. doi: 10.1371/journal.pone.0026720

Lenz H, Rüdinger M, Volkmar U, Fischer S, Herres S, Grewe F and Knoop V (2010) Introducing the plant RNA editing prediction and analysis computer tool PREPACT and an update on RNA editing site nomenclature. Curr Genet 56:189–201. doi: 10.1007/s00294-009-0283-5

Li H, Dong Y, Yin H, Wang N, Yang J, Liu X, Wang Y, Wu J and Li X (2011) Characterization of the stress associated microRNAs in Glycine max by deep sequencing. BMC Plant Biol 11:170. doi: 10.1186/1471-2229-11-170

Liu J-M, Zhao J-Y, Lu P-P, Chen M, Guo C-H, Xu Z-S and Ma Y-Z (2016) The E-Subgroup Pentatricopeptide Repeat Protein Family in Arabidopsis thaliana and Confirmation of the Responsiveness PPR96 to Abiotic Stresses. Front Plant Sci 7:1825. doi: 10.3389/fpls.2016.01825

Liu YX, Han YP, Chang W, Zou Q, Guo MZ and Li W Bin (2010) Genomic analysis of MicroRNA promoters and their cis-acting elements in soybean. Agric Sci China 9:1561–1570. doi: 10.1016/S1671-2927(09)60252-2

Long RC, Li MN, Kang JM, Zhang TJ, Sun Y and Yang QC (2015) Small RNA deep sequencing identifies novel and salt-stress-regulated microRNAs from roots of Medicago sativa and Medicago truncatula. Physiol Plant 154:13–27. doi: 10.1111/ppl.12266

Love MI, Huber W and Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. doi: 10.1186/s13059-014-0550-8

Lu C, Qiu N, Lu Q, Wang B and Kuang T (2002) Does salt stress lead to increased susceptibility of photosystem II to photoinhibition and changes in photosynthetic pigment composition in halophyte Suaeda salsa grown outdoors? Plant Sci 163:1063–1068. doi: 10.1016/S0168-9452(02)00281-9

Lu KX, Cao BH, Feng XP, He Y and Jiang DA (2009) Photosynthetic response of salt-tolerant and sensitive soybean varieties. Photosynthetica 47:381–387. doi: 10.1007/s11099-009-0059-7

Lu KX, Yang Y, He Y and Jiang DA (2008) Induction of cyclic electron flow around photosystem 1 and state transition are correlated with salt tolerance in soybean. Photosynthetica 46:10–16. doi: 10.1007/s11099-008-0003-2

Lurin C, Andrés C, Aubourg S, Bellaoui M, Bitton F, Bruyère C, Caboche M, Debast C, Gualberto J, Hoffmann B et al. (2004) Genome-Wide Analysis of Arabidopsis Pentatricopeptide Repeat Proteins Reveals Their Essential Role in Organelle Biogenesis. Plant Cell 16:2089–2103. doi: 10.1105/tpc.104.022236.discover

Ma Y, Qin F and Tran LSP (2012) Contribution of genomics to gene discovery in plant abiotic stress responses. Mol Plant 5:1176–1178. doi: 10.1093/mp/sss085

Mahajan S and Tuteja N (2005) Cold, salinity and drought stresses: An overview. Arch Biochem Biophys 444:139–158. doi: 10.1016/j.abb.2005.10.018

Maier RM, Neckermann K, Igloi GL and Kössel H (1995) Complete Sequence of the Maize Chloroplast Genome: Gene Content, Hotspots of Divergence and Fine Tuning of Genetic Information by Transcript Editing. J Mol Biol 251:614–628. doi: 10.1006/jmbi.1995.0460

Maréchal-Drouard L, Kumar R, Remacle C and Small I (1996) RNA editing of larch mitochondrial tRNA(His) precursors is a prerequisite for processing. Nucleic Acids Res 24:3229–34. doi: 10.1093/nar/24.16.3229

Mareéchal-Drouard L, Ramamonjisoa D, Cosset A, Weil JH and Dietrich A (1993) Editing corrects mispairing in the acceptor stem of bean and potato mitochondrial phenylalanine transfer RNAs. Nucleic Acids Res 21:4909–4914. doi: 10.1093/nar/21.21.4909

Mathesius U, Djordjevic MA, Oakes M, Goffard N, Haerizadeh F, Weiller GF, Singh MB and Bhalla PL (2011) Comparative proteomic profiles of the soybean (Glycine max) root apex and differentiated root zone. Proteomics 11:1707–1719. doi: 10.1002/pmic.201000619

Meng Y, Chen D, Jin Y, Mao C, Wu P and Chen M (2010) RNA editing of nuclear transcripts in Arabidopsis thaliana. BMC Genomics 11:S12. doi: 10.1186/1471-2164-11-S4-S12

Miura E, Kato Y, Matsushima R, Albrecht V, Laalami S and Sakamoto W (2007) The Balance between Protein Synthesis and Degradation in Chloroplasts Determines Leaf Variegation in Arabidopsis yellow variegated Mutants. Plant Cell Online 19:1313–1328. doi: 10.1105/tpc.106.049270

Miyata Y, Sugita C, Maruyama K and Sugita M (2008) RNA editing in the anticodon of tRNA Leu (CAA) occurs before group I intron splicing in plastids of a moss Takakia lepidozioides S. Hatt. & Inoue. Plant Biol 10:250–255. doi: 10.1111/j.1438-8677.2007.00027.x

Miyata Y and Sugita M (2004) Tissue- and stage-specific RNA editing of rps14 transcripts in moss (Physcomitrella patens) chloroplasts. J Plant Physiol 161:113–115. doi: 10.1078/0176-1617-01220

Mochizuki N, Brusslan JA, Larkin R, Nagatani A and Chory J (2001) Arabidopsis genomes uncoupled 5 (GUN5) mutant reveals the involvement of Mg-chelatase H subunit in plastid-to-nucleus signal transduction. Proc Natl Acad Sci 98:2053–2058. doi: 10.1073/pnas.98.4.2053

Mower JP (2009) The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucleic Acids Res 37:W253–W259. doi: 10.1093/nar/gkp337

Mozzo M, Passarini F, Bassi R, van Amerongen H and Croce R (2008) Photoprotection in higher plants: The putative quenching site is conserved in all outer light-harvesting complexes of Photosystem II. Biochim Biophys Acta - Bioenerg 1777:1263–1267. doi: 10.1016/j.bbabio.2008.04.036

Munekage Y, Hashimoto M, Miyake C, Tomizawa K, Endo T, Tasaka M and Shikanai T (2004) Cyclic electron flow around photosystem I is essential for photosynthesis. Nature 429:579–582. doi: 10.1038/nature02598

Naested H (2004) Arabidopsis VARIEGATED 3 encodes a chloroplast-targeted, zinc-finger protein required for chloroplast and palisade cell development. J Cell Sci 117:4807–4818. doi: 10.1242/jcs.01360

Nakajima Y and Mulligan M (2001) Heat stress results in incomplete C-to-U editing of maize chloroplast mRNAs and correlates with changes in chloroplast transcription rate. Curr Genet 40:209–213. doi: 10.1007/s002940100249

Nakajima Y, Mulligan M and Mulligan RM (2001) Heat stress results in incomplete C-to-U editing of maize chloroplast mRNAs and correlates with changes in chloroplast transcription rate. Curr Genet 40:209–213. doi: 10.1007/s002940100249

Neuwirt J, Takenaka M, Merwe JA van der and Brennicke A (2005) An in vitro RNA editing system from cauliflower mitochondria: Editing site recognition parameters can vary in different plant species. RNA 11:1563–1570. doi: 10.1261/rna.2740905

Nishikura K (2010) Functions and Regulation of RNA Editing by ADAR Deaminases. Annu Rev Biochem 79:321–349. doi: 10.1146/annurev-biochem-060208-105251

Nwokolo E (1996) Soybean (Glycine max (L.) Merr.). Food and Feed from Legumes and Oilseeds. Springer US, Boston, MA, pp 90–102

O'Toole N, Hattori M, Andres C, Iida K, Lurin C, Schmitz-Linneweber C, Sugita M and Small I (2008) On the expansion of the pentatricopeptide repeat gene family in plants. Mol Biol Evol 25:1120–1128. doi: 10.1093/molbev/msn057

Ohtani M, Demura T and Sugiyama M (2013) Arabidopsis ROOT INITIATION DEFECTIVE1, a DEAH-Box RNA Helicase Involved in Pre-mRNA Splicing, Is Essential for Plant Development. Plant Cell 25:2056–2069. doi: 10.1105/tpc.113.111922

Ohtani S, Ichinose M, Tasaki E, Aoki Y, Komura Y and Sugita M (2010) Targeted Gene Disruption Identifies Three PPR-DYW Proteins Involved in RNA Editing for Five Editing Sites of the Moss Mitochondrial Transcripts. Plant Cell Physiol 51:1942–1949. doi: 10.1093/pcp/pcq142

Oikawa K (2003) CHLOROPLAST UNUSUAL POSITIONING1 Is Essential for Proper Chloroplast Positioning. PLANT CELL ONLINE 15:2805–2815. doi: 10.1105/tpc.016428

Okuda K, Chateigner-Boutin A-L, Nakamura T, Delannoy E, Sugita M, Myouga F, Motohashi R, Shinozaki K, Small I and Shikanai T (2009) Pentatricopeptide Repeat Proteins with the DYW Motif Have Distinct Molecular Functions in RNA Editing and RNA Cleavage in Arabidopsis Chloroplasts. Plant Cell Online 21:146–156. doi: 10.1105/tpc.108.064667

Okuda K, Nakamura T, Sugita M, Shimizu T and Shikanai T (2006) A Pentatricopeptide Repeat Protein Is a Site Recognition Factor in Chloroplast RNA Editing. J Biol Chem 281:37661–37667. doi: 10.1074/jbc.M608184200

Okuda K and Shikanai T (2012) A pentatricopeptide repeat protein acts as a site-specificity factor at multiple RNA editing sites with unrelated cis-acting elements in plastids. Nucleic Acids Res 40:5052–5064. doi: 10.1093/nar/gks164

Okuda K, Shoki H, Arai M, Shikanai T, Small I and Nakamura T (2014) Quantitative analysis of motifs contributing to the interaction between PLS-subfamily members and their target RNA sequences in plastid RNA editing. Plant J 80:870–882. doi: 10.1111/tpj.12687

Oldenkott B, Yamaguchi K, Tsuji-Tsukinoki S, Knie N and Knoop V (2014) Chloroplast RNA editing going extreme: more than 3400 events of C-to-U editing in the chloroplast transcriptome of the lycophyte *Selaginella uncinata*. Rna 20:1499–1506. doi: 10.1261/rna.045575.114

Olinares PDB, Ponnala L and van Wijk KJ (2010) Megadalton Complexes in the Chloroplast Stroma of Arabidopsis thaliana Characterized by Size Exclusion Chromatography, Mass Spectrometry, and Hierarchical Clustering. Mol Cell Proteomics 9:1594–1615. doi: 10.1074/mcp.M000038-MCP201

Parida AK and Das AB (2005) Salt tolerance and salinity effects on plants: a review. Ecotoxicol Environ Saf 60:324–349. doi: 10.1016/j.ecoenv.2004.06.010

Porath HT, Knisbacher BA, Eisenberg E and Levanon EY (2017) Massive A-to-I RNA editing is common across the Metazoa and correlates with dsRNA abundance. Genome Biol 18:185. doi: 10.1186/s13059-017-1315-y

Pyo YJ, Kwon K-C, Kim A and Cho MH (2013) Seedling Lethal1, a Pentatricopeptide Repeat Protein Lacking an E/E+ or DYW Domain in Arabidopsis, Is Involved in Plastid Gene Expression and Early Chloroplast Development. Plant Physiol 163:1844–1858. doi: 10.1104/pp.113.227199

Ramos-Vega M, Guevara-García A, Llamas E, Sánchez-León N, Olmedo-Monfil V, Vielle-Calzada JP and León P (2015) Functional analysis of the Arabidopsis thaliana CHLOROPLAST BIOGENESIS 19 pentatricopeptide repeat editing protein. New Phytol 208:430–441. doi: 10.1111/nph.13468

Robbins JC, Heller WP and Hanson MR (2009) A comparative genomics approach identifies a PPR-DYW protein that is essential for C-to-U editing of the Arabidopsis chloroplast accD transcript. RNA 15:1142–1153. doi: 10.1261/rna.1533909

Rodrigues NF, Christoff AP, da Fonseca GC, Kulcheski FR and Margis R (2017a) Unveiling Chloroplast RNA Editing Events Using Next Generation Small RNA Sequencing Data. Front Plant Sci 8:1686. doi: 10.3389/fpls.2017.01686

Rodrigues NF, Fonseca GC da, Kulcheski FR and Margis R (2017b) Salt stress affects mRNA editing in soybean chloroplasts. Genet Mol Biol 40:200–208. doi: 10.1590/1678-4685-gmb-2016-0055

Rudinger M, Polsakiewicz M and Knoop V (2008) Organellar RNA Editing and Plant-Specific Extensions of Pentatricopeptide Repeat Proteins in Jungermanniid but not in Marchantiid Liverworts. Mol Biol Evol 25:1405–1414. doi: 10.1093/molbev/msn084
Ruf S and Bock R (2011) In Vivo Analysis of RNA Editing in Plastids. Methods Mol Biol. pp 137–150

Ruiz-Ferrer V and Voinnet O (2009) Roles of Plant Small RNAs in Biotic Stress Responses. Annu Rev Plant Biol 60:485–510. doi: 10.1146/annurev.arplant.043008.092111

Ruwe H, Castandet B, Schmitz-Linneweber C and Stern DB (2013) Arabidopsis chloroplast quantitative editotype. FEBS Lett 587:1429–1433. doi: 10.1016/j.febslet.2013.03.022

Salekdeh GH, Siopongco J, Wade LJ, Ghareyazie B and Bennett J (2002) Proteomic analysis of rice leaves during drought stress and recovery. Proteomics 2:1131–45. doi: 10.1002/1615-9861(200209)2:9<1131::AID-PROT1131>3.0.CO;2-1

Salone V, Rüdinger M, Polsakiewicz M, Hoffmann B, Groth-Malonek M, Szurek B, Small I, Knoop V and Lurin C (2007) A hypothesis on the identification of the editing enzyme in plant organelles. FEBS Lett 581:4132–4138. doi: 10.1016/j.febslet.2007.07.075

Sasaki Y, Kozaki A, Ohmori A, Iguchi H and Nagano Y (2001) Chloroplast RNA Editing Required for Functional Acetyl-CoA Carboxylase in Plants. J Biol Chem 276:3937–3940. doi: 10.1074/jbc.M008166200

Schallenberg-Rüdinger M and Knoop V (2016) Coevolution of Organelle RNA Editing and Nuclear Specificity Factors in Early Land Plants. Adv Bot Res 78:37–93. doi: 10.1016/bs.abr.2016.01.002

Schaub M and Keller W (2002) RNA editing by adenosine deaminases generates RNA and protein diversity. Biochimie 84:791–803. doi: 10.1016/S0300-9084(02)01446-3

Schmidt von Braun S and Schleiff E (2008) The chloroplast outer membrane protein CHUP1 interacts with actin and profilin. Planta 227:1151–1159. doi: 10.1007/s00425-007-0688-7

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J et al. (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–

183. doi: 10.1038/nature08670

Severin AJ, Woody JL, Bolon Y-T, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE et al. (2010) RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. BMC Plant Biol 10:160. doi: 10.1186/1471-2229-10-160

Shi C, Hu N, Huang H, Gao J, Zhao YJ and Gao LZ (2012) An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. PLoS One 7:1–7. doi: 10.1371/journal.pone.0031468

Shi X, Bentolila S and Hanson MR (2016a) Organelle RNA recognition motif-containing (ORRM) proteins are plastid and mitochondrial editing factors in Arabidopsis. Plant Signal Behav 11:1–4. doi: 10.1080/15592324.2016.1167299

Shi X, Castandet B, Germain A, Hanson MR and Bentolila S (2017a) ORRM5, an RNA recognition motif-containing protein, has a unique effect on mitochondrial RNA editing. J Exp Bot 68:2833–2847. doi: 10.1093/jxb/erx139

Shi X, Germain A, Hanson MR and Bentolila S (2016b) RNA Recognition Motif-Containing Protein ORRM4 Broadly Affects Mitochondrial RNA Editing and Impacts Plant Development and Flowering. Plant Physiol 170:294–309. doi: 10.1104/pp.15.01280

Shi X, Hanson MR and Bentolila S (2017b) Functional diversity of Arabidopsis organelle-localized RNA-recognition motif-containing proteins. Wiley Interdiscip Rev RNA 8:1–14. doi: 10.1002/wrna.1420

Shi X, Hanson MR and Bentolila S (2015) Two RNA recognition motif-containing proteins are plant mitochondrial editing factors. Nucleic Acids Res 43:3814–3825. doi: 10.1093/nar/gkv245

Shikanai T (2006) RNA editing in plant organelles: machinery, physiological function and evolution. Cell Mol Life Sci 63:698–708. doi: 10.1007/s00018-005-5449-9

Shikanai T (2015) RNA editing in plants: Machinery and flexibility of site recognition. Biochim Biophys Acta - Bioenerg 1847:779–785. doi: 10.1016/j.bbabio.2014.12.010

Silva SR, Diaz YCA, Penha HA, Pinheiro DG, Fernandes CC, Miranda VFO, Michael TP and Varani AM (2016) The chloroplast genome of Utricularia reniformis sheds light on the evolution of the ndh gene complex of terrestrial carnivorous plants from the lentibulariaceae family. PLoS One 11:1–29. doi: 10.1371/journal.pone.0165176

Silveira JA., Melo AR., Viégas R. and Oliveira JT. (2001) Salinity-induced effects on nitrogen assimilation related to growth in cowpea plants. Environ Exp Bot 46:171–179. doi: 10.1016/S0098-8472(01)00095-8

Simon SA, Zhai J, Nandety RS, McCormick KP, Zeng J, Mejia D and Meyers BC (2009) Short-Read Sequencing Technologies for Transcriptional Analyses. Annu Rev Plant Biol 60:305–333. doi: 10.1146/annurev.arplant.043008.092032

Singh CR, Lee B, Udagawa T, Mohammad-Qureshi SS, Yamamoto Y, Pavitt GD and Asano K (2006) An eIF5/eIF2 complex antagonizes guanine nucleotide exchange by eIF2B during translation initiation. EMBO J 25:4537–4546. doi: 10.1038/sj.emboj.7601339

Small ID and Peeters N (2000) The PPR motif - a TPR-related motif prevalent in plant organellar proteins. Trends Biochem Sci 25:45–47. doi: 10.1016/S0968-0004(99)01520-0

Smith DR and Sanitá Lima M (2016) Unraveling chloroplast transcriptomes with ChloroSeq, an organelle RNA-Seq bioinformatics pipeline. Brief Bioinform 18:bbw088. doi: 10.1093/bib/bbw088

Sommer B, Köhler M, Sprengel R and Seeburg PH (1991) RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. Cell 67:11–19. doi: 10.1016/0092-8674(91)90568-J

Sun T, Bentolila S and Hanson MR (2016) The Unexpected Diversity of Plant Organelle RNA Editosomes. Trends Plant Sci 21:962–973. doi: 10.1016/j.tplants.2016.07.005

Sun T, Germain A, Giloteaux L, Hammani K, Barkan A, Hanson MR and Bentolila S (2013) An RNA recognition motif-containing protein is required for plastid RNA editing in Arabidopsis and maize. Proc Natl Acad Sci 110:E1169–E1178. doi: 10.1073/pnas.1220162110

Sun T, Shi X, Friso G, Van Wijk K, Bentolila S and Hanson MR (2015) A Zinc Finger Motif-Containing Protein Is Essential for Chloroplast RNA Editing. PLOS Genet 11:e1005028. doi: 10.1371/journal.pgen.1005028

Tadini L, Pesaresi P, Kleine T, Rossi F, Guljamow A, Sommer F, Mühlhaus T, Schroda M, Masiero S, Pribil M et al. (2016) GUN1 Controls Accumulation of the Plastid Ribosomal Protein S1 at the Protein Level and Interacts with Proteins Involved in Plastid Protein Homeostasis. Plant Physiol 170:1817–30. doi: 10.1104/pp.15.02033

Takenaka M, Verbitskiy D, van der Merwe JA, Zehrmann A and Brennicke A (2008) The process of RNA editing in plant mitochondria. Mitochondrion 8:35–46. doi: 10.1016/j.mito.2007.09.004

Takenaka M, Zehrmann A, Verbitskiy D, Härtel B and Brennicke A (2013) RNA Editing in Plants and Its Evolution. Annu Rev Genet 47:335–352. doi: 10.1146/annurev-genet-111212-133519

Takenaka M, Zehrmann A, Verbitskiy D, Kugelmann M, Hartel B and Brennicke A (2012) Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. Proc Natl Acad Sci 109:5104–5109. doi: 10.1073/pnas.1202452109

Tamura K, Stecher G, Peterson D, Filipski A and Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729. doi:

10.1093/molbev/mst197

Tan J, Tan Z, Wu F, Sheng P, Heng Y, Wang X, Ren Y, Wang J, Guo X, Zhang X et al. (2014) A novel chloroplast-localized pentatricopeptide repeat protein involved in splicing affects chloroplast development and abiotic stress response in rice. Mol Plant 7:1329–1349. doi: 10.1093/mp/ssu054

Terzi LC and Simpson GG (2009) Arabidopsis RNA immunoprecipitation. Plant J 59:163–168. doi: 10.1111/j.1365-313X.2009.03859.x

Tillich M, Hardel SL, Kupsch C, Armbruster U, Delannoy E, Gualberto JM, Lehwark P, Leister D, Small ID and Schmitz-Linneweber C (2009) Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. Proc Natl Acad Sci 106:6002–6007. doi: 10.1073/pnas.0808529106

Tillich M, Lehwark P, Morton BR and Maier UG (2006a) The evolution of chloroplast RNA editing. Mol Biol Evol 23:1912–1921. doi: 10.1093/molbev/msl054

Tillich M, Poltnigg P, Kushnir S and Schmitz-Linneweber C (2006b) Maintenance of plastid RNA editing activities independently of their target sites. EMBO Rep 7:308–313. doi: 10.1038/sj.embor.7400619

Timmis JN, Ayliffe MA, Huang CY and Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet 5:123–135. doi: 10.1038/nrg1271

Tonti-Filippini J, Nevill PG, Dixon K and Small I (2017) What can we do with 1000 plastid genomes? Plant J 90:808–818. doi: 10.1111/tpj.13491

Tseng CC, Lee CJ, Chung YT, Sung TY and Hsieh MH (2013) Differential regulation of Arabidopsis plastid gene expression and RNA editing in non-photosynthetic tissues. Plant Mol Biol 82:375–392. doi: 10.1007/s11103-013-0069-5

Tseng CC, Sung TY, Li YC, Hsu SJ, Lin CL and Hsieh MH (2010) Editing of accD and ndhF chloroplast transcripts is partially affected in the Arabidopsis vanilla cream1 mutant. Plant Mol Biol 73:309–323. doi: 10.1007/s11103-010-9616-5

Turner M, Yu O and Subramanian S (2012) Genome organization and characteristics of soybean microRNAs. BMC Genomics 13:169. doi: 10.1186/1471-2164-13-169

Umate P (2010) Genome-wide analysis of the family of light-harvesting chlorophyll a/b-binding proteins in Arabidopsis and rice. Plant Signal Behav 5:1537–1542. doi: 10.4161/psb.5.12.13410

Van Den Bekerom RJM, Dix PJ, Diekmann K and Barth S (2013) Variations in efficiency of plastidial RNA editing within ndh transcripts of perennial ryegrass (Lolium perenne) are not linked to differences in drought tolerance. AoB Plants 5:plt035-plt035. doi: 10.1093/aobpla/plt035

Van Der Merwe JA, Takenaka M, Neuwirt J, Verbitskiy D and Brennicke A (2006) RNA editing sites in plant mitochondria can share cis-elements. FEBS Lett 580:268–272. doi: 10.1016/j.febslet.2005.12.011

Verbitskiy D, Zehrmann A, Härtel B, Brennicke A and Takenaka M (2012) Two related RNA-editing proteins target the same sites in mitochondria of Arabidopsis thaliana. J Biol Chem 287:38064–38072. doi: 10.1074/jbc.M112.397992

Vermel M, Guermann B, Delage L, Grienenberger JM, Marechal-Drouard L and Gualberto JM (2002) A family of RRM-type RNA-binding proteins specific to plant mitochondria. Proc Natl Acad Sci U S A 99:5866–5871. doi: 10.1073/pnas.092019599

Vieira LDN, Faoro H, De Freitas Fraga HP, Rogalski M, De Souza EM, De Oliveira Pedrosa F, Nodari RO and Guerra MP (2014) An improved protocol for intact chloroplasts and cpDNA isolation in conifers. PLoS One 9:1–8. doi: 10.1371/journal.pone.0084792

Vogel J, Hübschmann T, Börner T and Hess WR (1997) Splicing and intron-internal RNA editing of trnK-matK transcripts in barley plastids: support for MatK as an essential splice factor. J Mol Biol 270:179–187. doi: 10.1006/jmbi.1997.1115

Vu LT and Tsukahara T (2017) C-to-U editing and site-directed RNA editing for the correction of genetic mutations. Biosci Trends 11:243–253. doi: 10.5582/bst.2017.01049

Wagoner J a, Sun T, Lin L and Hanson MR (2015) Cytidine Deaminase Motifs within the DYW Domain of Two Pentatricopeptide Repeat-containing Proteins Are Required for Site-specific Chloroplast RNA Editing. J Biol Chem 290:2957–2968. doi: 10.1074/jbc.M114.622084

Wang D, Shannon MC and Grieve CM (2001) Salinity reduces radiation absorption and use efficiency in soybean. F Crop Res 69:267–277. doi: 10.1016/S0378-4290(00)00154-4

Wang G, Zhong M, Shuai B, Song J, Zhang J, Han L, Ling H, Tang Y, Wang G and Song R (2017) E+ subgroup PPR protein defective kernel 36 is required for multiple mitochondrial transcripts editing and seed development in maize and Arabidopsis. New Phytol 214:1563–1578. doi: 10.1111/nph.14507

Wang W, Vinocur B and Altman A (2003) Plant responses to drought, salinity and extreme temperatures: Towards genetic engineering for stress tolerance. Planta 218:1–14. doi: 10.1007/s00425-003-1105-5

Wolf PG, Rowe CA and Hasebe M (2004) High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern Adiantum capillus-veneris. Gene 339:89–97. doi: 10.1016/j.gene.2004.06.018

Wortham NC and Proud CG (2015) eIF2B: recent structural and functional insights into a key regulator of translation. Biochem Soc Trans 43:1234–40. doi: 10.1042/BST20150164

Xu R, Zhang S, Huang J and Zheng C (2013) Genome-wide comparative in silico analysis of the RNA helicase gene family in Zea mays and Glycine max: A comparison with Arabidopsis and Oryza sativa. PLoS One 8:1–13. doi: 10.1371/journal.pone.0078982

Xu T, Wang Y, Liu X, Lv S, Feng C, Qi M and Li T (2015) Small RNA and degradome sequencing reveals microRNAs and their targets involved in tomato pedicel abscission. Planta 242:963–984. doi: 10.1007/s00425-015-2318-0

Yagi Y, Hayashi S, Kobayashi K, Hirayama T and Nakamura T (2013a) Elucidation of the RNA Recognition Code for Pentatricopeptide Repeat Proteins Involved in Organelle RNA Editing in Plants. PLoS One 8:e57286. doi: 10.1371/journal.pone.0057286

Yagi Y, Tachikawa M, Noguchi H, Satoh S, Obokata J and Nakamura T (2013b) Pentatricopeptide repeat proteins involved in plant organellar RNA editing. RNA Biol 10:1419–1425. doi: 10.4161/rna.24908

Yan J, Zhang Q, Guan Z, Wang Q, Li L, Ruan F, Lin R, Zou T and Yin P (2017) MORF9 increases the RNA-binding activity of PLS-type pentatricopeptide repeat protein in plastid RNA editing. Nat Plants 3:17037. doi: 10.1038/nplants.2017.37

Yap A, Kindgren P, Colas des Francs-Small C, Kazama T, Tanz SK, Toriyama K and Small I (2015) AEF1/MPR25 is implicated in RNA editing of plastid atpF and mitochondrial nad5 , and also promotes atpF splicing in Arabidopsis and rice. Plant J 81:661–669. doi: 10.1111/tpj.12756

Yuan H and Liu D (2012) Functional disruption of the pentatricopeptide protein SLG1 affects mitochondrial RNA editing, plant development, and responses to abiotic stresses in Arabidopsis. Plant J 70:432–444. doi: 10.1111/j.1365-313X.2011.04883.x

Zehrmann A, Härtel B, Glass F, Bayer-Császár E, Obata T, Meyer E, Brennicke A and Takenaka M (2015a) Selective Homo- and Heteromer Interactions between the Multiple Organellar RNA Editing Factor (MORF) Proteins in Arabidopsis thaliana. J Biol Chem 290:6445–6456. doi: 10.1074/jbc.M114.602086

Zehrmann A, Härtel B, Glass F, Bayer-Császár E, Obata T, Meyer E, Brennicke A and Takenaka M (2015b) Selective homo- and heteromer interactions between the multiple organellar RNA editing factor (MORF) proteins in Arabidopsis thaliana. J Biol Chem 290:6445–6456. doi: 10.1074/jbc.M114.602086

Zehrmann A, Verbitskiy D, van der Merwe JA, Brennicke A and Takenaka M (2009) A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of Arabidopsis thaliana. Plant Cell 21:558–67. doi: 10.1105/tpc.108.064535

Zeng HQ, Zhu YY, Huang SQ and Yang ZM (2010) Analysis of phosphorus-deficient responsive miRNAs and cis-elements from soybean (Glycine max L.). J Plant Physiol 167:1289–1297. doi: 10.1016/j.jplph.2010.04.017

Zhang DY, Ali Z, Wang CB, Xu L, Yi JX, Xu ZL, Liu XQ, He XL, Huang YH, Khan IA et al. (2013) Genome-Wide Sequence Characterization and Expression Analysis of Major Intrinsic Proteins in Soybean (Glycine max L.). PLoS One 8:e56312. doi: 10.1371/journal.pone.0056312

Zhang F, Tang W, Hedtke B, Zhong L, Liu L, Peng L, Lu C, Grimm B and Lin R (2014) Tetrapyrrole biosynthetic enzyme protoporphyrinogen IX oxidase 1 is required for plastid RNA editing. Proc Natl Acad Sci 111:2023–2028. doi: 10.1073/pnas.1316183111

Zhang J, Tan W, Yang XH and Zhang HX (2008) Plastid-expressed choline monooxygenase gene improves salt and drought tolerance through accumulation of glycine betaine in tobacco. Plant Cell Rep 27:1113–1124. doi: 10.1007/s00299-008-0549-2

Zheng C, Jiang D, Liu F, Dai T, Jing Q and Cao W (2009) Effects of salt and waterlogging stresses and their combination on leaf photosynthesis, chloroplast ATP synthesis, and antioxidant capacity in wheat. Plant Sci 176:575–582. doi: 10.1016/j.plantsci.2009.01.015

Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, Yu Y, Shu L, Zhao Y, Ma Y et al. (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat Biotechnol 33:408–414. doi: 10.1038/nbt.3096

Zhu Q, Dugardeyn J, Zhang C, Mühlenbock P, Eastmond PJ, Valcke R, De Coninck B, Öden S, Karampelias M, Cammue BPA et al. (2014) The Arabidopsis thaliana RNA Editing Factor SLO2, which Affects the Mitochondrial Electron Transport Chain, Participates in Multiple Stress and Hormone Responses. Mol Plant 7:290–310. doi: 10.1093/mp/sst102

Zhu Q, Dugardeyn J, Zhang C, Takenaka M, Kühn K, Craddock C, Smalle J, Karampelias M, Denecke J, Peters J et al. (2012) SLO2, a mitochondrial pentatricopeptide repeat protein affecting several RNA editing sites, is required for energy metabolism. Plant J 71:836–849. doi: 10.1111/j.1365-313X.2012.05036.x

8. ANEXOS

Outras produções científicas relacionadas no período

Genetics and Molecular Biology, 40, 4, 871-876 (2017)

DOI: http://dx.doi.org/10.1590/1678-4685-GMB-2017-0096

Genome Insight

Complete sequence and comparative analysis of the chloroplast genome of *Plinia trunciflora*

Maria Eguiluz[1], Priscila Mary Yuyama[2], Frank Guzman[2], Nureyev Ferreira Rodrigues[1] and Rogerio Margis[1,2]

[1]*Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil.*
[2]*Departamento de Biofísica, Centro de Biotecnologia, Laboratório de Genomas e Populações de Plantas, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil.*

Abstract

Plinia trunciflora is a Brazilian native fruit tree from the Myrtaceae family, also known as jaboticaba. This species has great potential by its fruit production. Due to the high content of essential oils in their leaves and of anthocyanins in the fruits, there is also an increasing interest by the pharmaceutical industry. Nevertheless, there are few studies focusing on its molecular biology and genetic characterization. We herein report the complete chloroplast (cp) genome of *P. trunciflora* using high-throughput sequencing and compare it to other previously sequenced Myrtaceae genomes. The cp genome of *P. trunciflora* is 159,512 bp in size, comprising inverted repeats of 26,414 bp and single-copy regions of 88,097 bp (LSC) and 18,587 bp (SSC). The genome contains 111 single-copy genes (77 protein-coding, 30 tRNA and four rRNA genes). Phylogenetic analysis using 57 cp protein-coding genes demonstrated that *P. trunciflora, Eugenia uniflora* and *Acca sellowiana* form a cluster with closer relationship to *Syzygium cumini* than with *Eucalyptus*. The complete cp sequence reported here can be used in evolutionary and population genetics studies, contributing to resolve the complex taxonomy of this species and fill the gap in genetic characterization.

Keywords: Jaboticaba, Myrtaceae, chloroplast genome, next-generation sequencing.

Received: April 18, 2017; Accepted: July 13, 2017.

Plinia trunciflora (O.Berg) Kausel, synonym *Myrciaria trunciflora* O.Berg, is a native Brazilian tree that belongs to the Myrtaceae family and is widely distributed in the southern and southeastern areas of Brazil (Sobral *et al.*, 2012). Among all identified *Plinia* sp. species, *P. cauliflora* (DC.) Berg (synonym *M. cauliflora* (Mart.) O.Berg), *P. jaboticaba* (Vell.) Berg (synonym *M. jaboticaba* O.Berg) and *P. trunciflora* are endemic to Brazil. All of these species produce a similar grape-like edible fruit, known as jaboticaba, which presents a sweet jelly-like white pulp covered by a purple peel. Jaboticaba (*P. trunciflora*) has attracted attention because of its significant levels of phenolic compounds associated with health benefits, such as antidepressant and antioxidant effects and the prevention of neurodegenerative diseases and diabetes (Stasi and Hiruma-Lima, 2002; Sacchet *et al.*, 2015). These benefits have largely been attributed to the capacity of these compounds to prevent or reduce oxidative stress. Additionally, jaboticaba (*P. trunciflora*) is largely consumed fresh or used to make jellies, juices, wines, spirits and vinegar (Balerdi *et al.*, 2006).

Send correspondence to Rogerio Margis. Departamento de Biofísica, Centro de Biotecnologia, Laboratório de Genomas e Populações de Plantas, Universidade Federal do Rio Grande do Sul (UFRGS), Avenida Bento Gonçalves 9500, Prédio 43432, Sala 206, Porto Alegre, RS, CEP 91501-970 Brazil. E-mail: rogerio.margis@ufrgs.br.

Despite the nutritional and productive recognized importance of this species, the taxonomic classification is still controversial. This is mostly so because it is based on morphological evaluation of the trees, fruits and seeds, regarding physical, chemical, physicochemical, and germinal characters that have shown the existence of variability (Guedes *et al.*, 2014). Therefore, molecular studies are needed to better clarify the phylogenetic relationships among the species from this genus.

The chloroplast (cp) genome is a circular molecule of double-stranded DNA that consists of four distinct regions, a large and a small single copy region (LSC and SSC, respectively) separated by two inverted repeat regions (IRa and IRb). Despite the high degree of conservation in its structure, gene content and organization, the presence of mutations, duplications and rearrangements of genes make it an attractive option for phylogenetic studies (Costa *et al.*, 2016). In the case of Myrtaceae, there are only few phylogenetic and evolutionary studies based on cp genes (Craven and Biffin 2005; Payn *et al.*, 2007; Biffin *et al.*, 2010; Bayly

et al., 2013; Eguiluz *et al.*, 2017; Machado *et al.*, 2017), and there are even less that include the *Plinia* genus (Vasconcelos *et al.*, 2017).

In this study, young leaves from a *Plinia trunciflora* tree harvested in Gravataí, RS, Brazil (latitude (S): 29°51'52"; longitude (W): 50°53'53") were used to extract total DNA by the CTAB method (Doyle and Doyle, 1990). DNA quality was evaluated by electrophoresis in a 1% agarose gel, and DNA quantity was determined using a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). One genomic paired-end library of 100 nt length was generated by Fasteris SA (Plan-les-Ouates, Switzerland) using an Illumina HiSeq2000 platform (Illumina Inc., San Diego, CA, USA). The paired-end sequence reads were filtered against 42 Myrtaceae cp genomes (Table S1) using BWA software with two mismatches allowed (Li and Durbin, 2009). The obtained reads were assembled *de novo* with ABySS software (Simpson *et al.*, 2009). The cp genome scaffolds were orientated using cp genome sequences of *Eucalyptus globulus*, *Eucalyptus grandis* and *Eugenia uniflora* L. using BLASTN (Camacho *et al.*, 2009). A gap region was filled in by Sanger sequencing using primers F: 5' GGGTTATCCTGCACTTGGAA and R: 3' TGCTGTCGAAGCTCCATCTA. Genes were annotated using DOGMA (Wyman *et al.*, 2004) and BLAST homology searches. tRNAs (transfer RNA) were predicted using tRNAscan-SE program (Schattner *et al.*, 2005) and confirmed by comparison with the appropriate homologs in *E. globulus*. The circular cp genome map was drawn using OGDRAW online program (Lohse *et al.*, 2007). For the phylogenetic analysis, a set of 57 cp protein-coding sequences (Table S2) from 56 species belonging to Malvids (Eurosids II) (Table S3) were used with *Vitis vinifera* serving as outgroup. Nucleotide sequences were aligned using MUSCLE available in MEGA version 6.0 (Tamura *et al.*, 2013), and a Bayesian tree was generated using MrBayes version 3.1.2 (Ronquist and Huelsenbeck, 2003) with 5,000,000 generations sampled every 100 generations and discarding the first 25% of trees as burn-in, with posterior probability (PP) values for each node. The GTR+I+G nucleotide substitution model determined by MODELTEST version 3.7 (Posada and Crandall, 1998) was used. The phylogenetic tree was rooted and visualized using FigTree software (http://tree.bio.ed.ac.uk/software/figtree/).

A total of 148,824,244 raw Illumina paired-end reads from the *P. truncliflora* nuclear genome were filtered against 42 Myrtaceae cp genomes. The 8,912,157 obtained reads were *de novo* assembled into non-redundant contigs and singletons covering about 99% of the genome (minimum coverage=144 reads, maximum coverage=18,789 reads). Two final large scaffolds were obtained and joined into a cp circular genome using Sanger sequencing. The complete cp genome of *P. trunciflora* is 159,512 bp in size and was submitted to GenBank (accession number: KU318111). The size is similar to that of other Myrtaceae species (Eguiluz *et al.*, 2017; Machado *et al.*, 2017). The cp genome included an LSC region of 88,097 bp, an SSC region of 18,587 bp and a pair of inverted repeats (IRa and IRb) of 26,414 bp each (Figure 1). Coding regions comprise 47.2%, 13.3% correspond to rRNAs and tRNAs, and 39.5% of the genome comprises non-coding regions, including introns, pseudogenes and intergenic spacers (Table 1). In general, all genomic features showed similarity in structure and gene abundance with other Myrtaceae species (Bayly *et al.*, 2013; Eguiluz *et al.*, 2017; Machado *et al.*, 2017). The genome contained 131 genes in total, which includes 111 single-copy genes corresponding to 77 protein-coding genes, 30 transfer RNA (tRNA) genes and four ribosomal genes (rRNA) (Figure 1, Table 1). The *ycf1*, *ycf2* and *ycf15* sequences were annotated as pseudogenes based on the presence of many stop codons in their coding sequences and by comparison with sequences of *E. globulus* and *S. cumini*. Of the 131 genes in *P. trunciflora*, seven of the tRNAs genes and all four rRNA genes occurred within the IR regions and consequently were duplicated (Table 1). The cp genome has 20 intron-containing genes: 12 protein coding genes and six tRNA genes which contain one intron, and the *clpP* and *ycf3* genes that contain two introns each. The *rps12* gene is a trans-spliced gene with the 5'end located in the LSC region and the duplicated 3'end in the IR

Table 1 - Summary of the *Plinia trunciflora* chloroplast genome characteristics.

Feature	*Plinia trunciflora*
Total cpDNA size	159,512 bp
LSC size (bp)	88,097 bp
SSC size (bp)	18,586 bp
IR size (bp)	26,414 bp
Protein coding regions (%)	60.48%
rRNA and tRNA (%)	13.3%
Introns size (% total)	10.65%
Intergenic sequences and pseudogenes size (%)	28.9%

Feature	*Plinia trunciflora*
Number of genes	131 genes
Number of different protein coding genes	77
Number of different tRNA genes	30
Number of different rRNA genes	4
Number of different duplicated genes	16
Pseudogenes	3
GC content (%)	37%

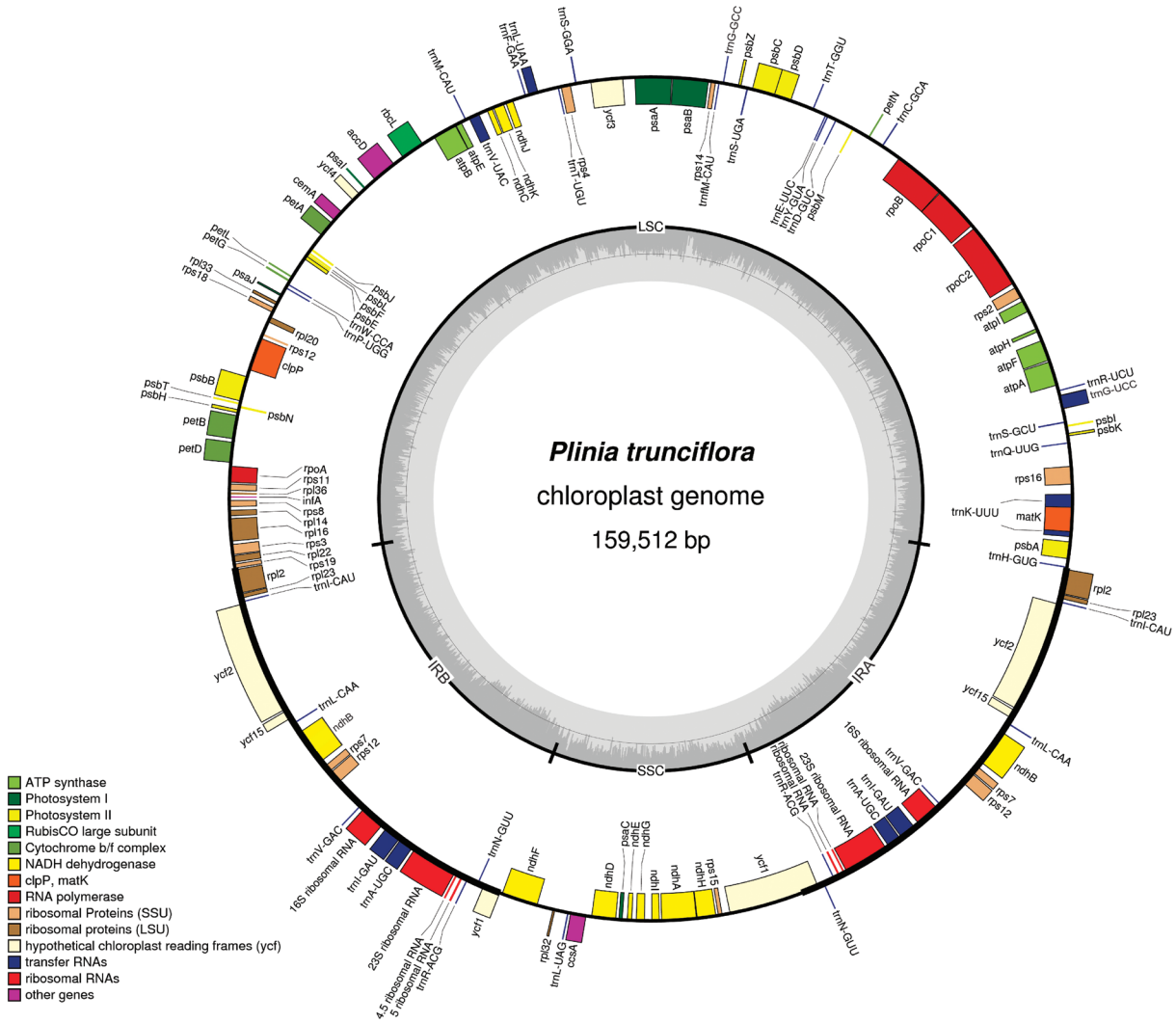

Figure 1 - Gene map of the *Plinia trunciflora* chloroplast genome. The structure of the cp genome consists of one large and small single copy (LSC and SSC, respectively) and a pair of inverted repeats (IRa and IRb). Genes drawn inside the circle are transcribed counterclockwise and those outsides are clockwise. Genes belonging to different functional groups are indicated by different tonalities. The darker gray in the inner circle corresponds to GC content, while the lighter gray corresponds to AT content.

regions. The *trnK-UUU* has 2,529 bp, with the largest intron encompassing also the *matK* gene.

The whole cp genome analysis revealed that the cp genomes of *P. trunciflora* and *E. uniflora* are shorter in comparison to other Myrtaceae, such as *E. globulus*, *E. grandis, E. uniflora* and *S. cumini*, (Figure 2). Despite its size, the total length of introns in *P. trunciflora* (16,972 pb) is the largest in Myrtaceae, e.g. *S. cumini* presents 14,469 bp and the same is observed in *E. globulus* and *E. grandis*. The size of the intergenic spacer located between the IRa/LSC border and the first gene of LSC in *P. trunciflora* is more similar to *Eucalyptus* species than its closer species *E. uniflora* (Figure 2). The comparison of the *ndhK* gene of *P. trunciflora*, with 678 bp, indicated a smaller gene size than that in other plants, such as *E. uniflora* (858 pb), *S. cumini* (855 bp), *E. globulus* (855 bp) and *E. grandis* (853 bp). The same size (678 bp) for this gene is found in *Arabidopsis thaliana*. The effective size of the coding sequence is confirmed by the presence of a thymine in position 53,811 bp in the cp genome from *P. trunciflora* that creates a stop codon and makes this gene shorter than in other Myrtaceae.

Our phylogeny includes the sister relationship of the orders Brassicales, Malvales and Sapindales and the orders Geraniales and Myrtales. All these results agree with previous studies based on multiple genes or complete cp genomes (Ruhfel *et al.*, 2014). By analyzing the Myrtaceae family clade we showed that *P. trunciflora*, *E. uniflora* and *Acca sellowiana* form a single cluster of Neotropical Myrtaceae, and that this clade has a shorter genetic distance with *S. cumini* than to the Australian Myrtaceae clade (Figure 3). Additionally, our analysis corroborates that *Corymbia gummifera* is paraphyletic in respect to *Angophora*. A previous phylogenetic analysis using certain cp

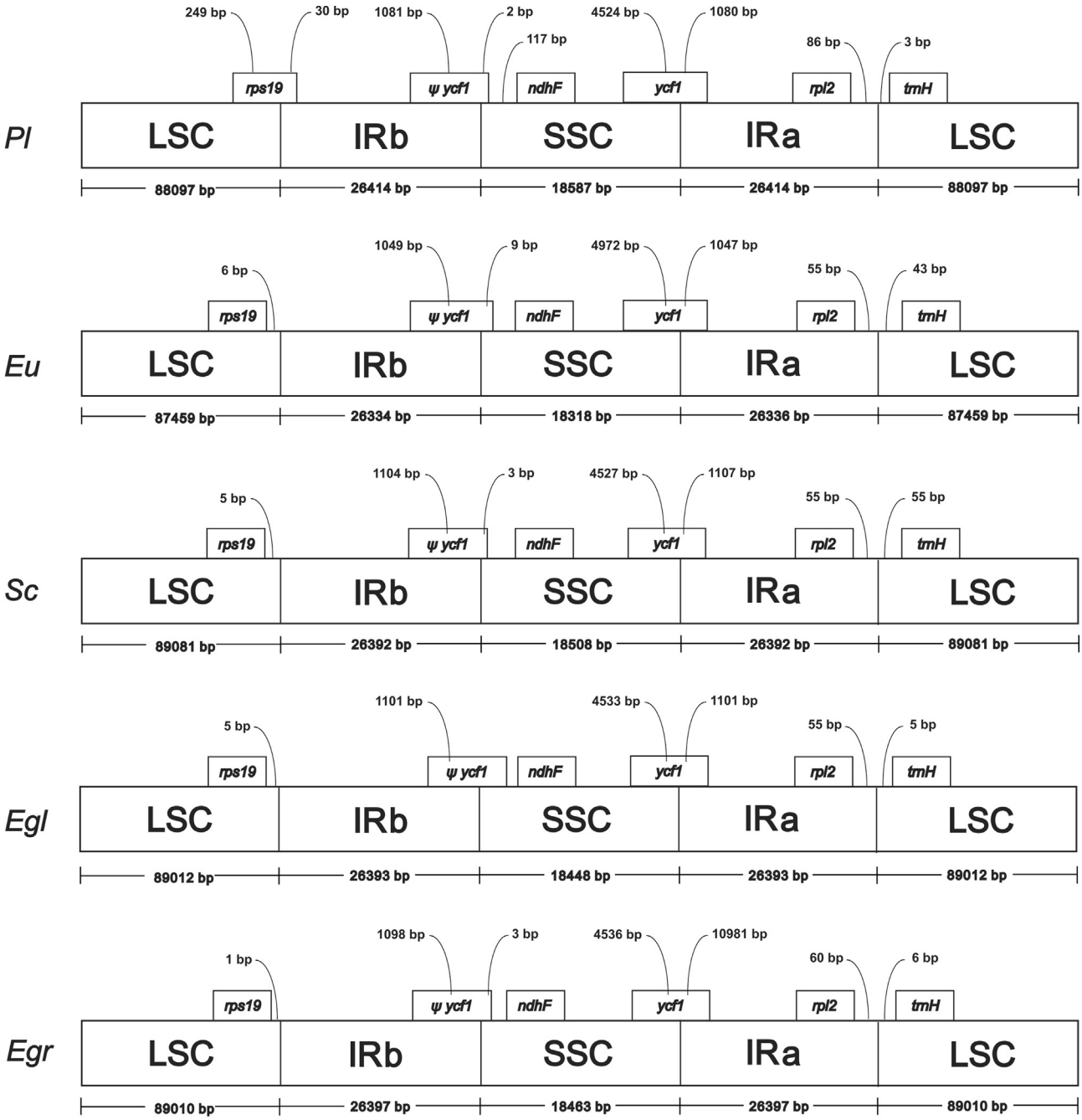

Figure 2 - Comparison of the borders of LSC, SSC and IR regions among five chloroplast genomes. Boxes above the main line indicate the predicted genes, while pseudogenes at the borders are shown by Ψ. Variation in *rps19* gene length is displayed at the IRb/LSC borders of *Plinia trunciflora*, *Eugenia uniflora*, *Syzygium cumini*, *Eucalyptus globulus* and *Eucalyptus grandis*, but only in *P. trunciflora*, this gene is located at IRb and LSC regions. This figure is not drawn to scale.

genes (ITS, *matK* and *ndhF*) of Myrtaceae species showed that *Eucalyptus*, *Syzygium*, *Eugenia* and *Myrciaria* (synonym of *Plinia*) form a distinct clade that is consistent with characteristics of the pollen (Thornhill *et al.*, 2012). As can be observed in the Bayesian tree (Figure 3), *Plinia* could be paraphyletic in relation to *Eugenia* and *Acca*, in agreement with the embryo morphology and studies using cp regions that placed *Plinia*, *Myrciaria* and *Siphoneugena* as the emerging "Plinia group" (Lucas *et al.*, 2007). Taxon sampling and phylogenetic methodology could affect the different results. Therefore, additional complete cp genome sequences will help in the comprehension of the relationship among Myrtaceae species.

The *Plinia trunciflora* genome represents the first complete cp genome sequence for the genus *Plinia* and shows a set of features that could be further explored for

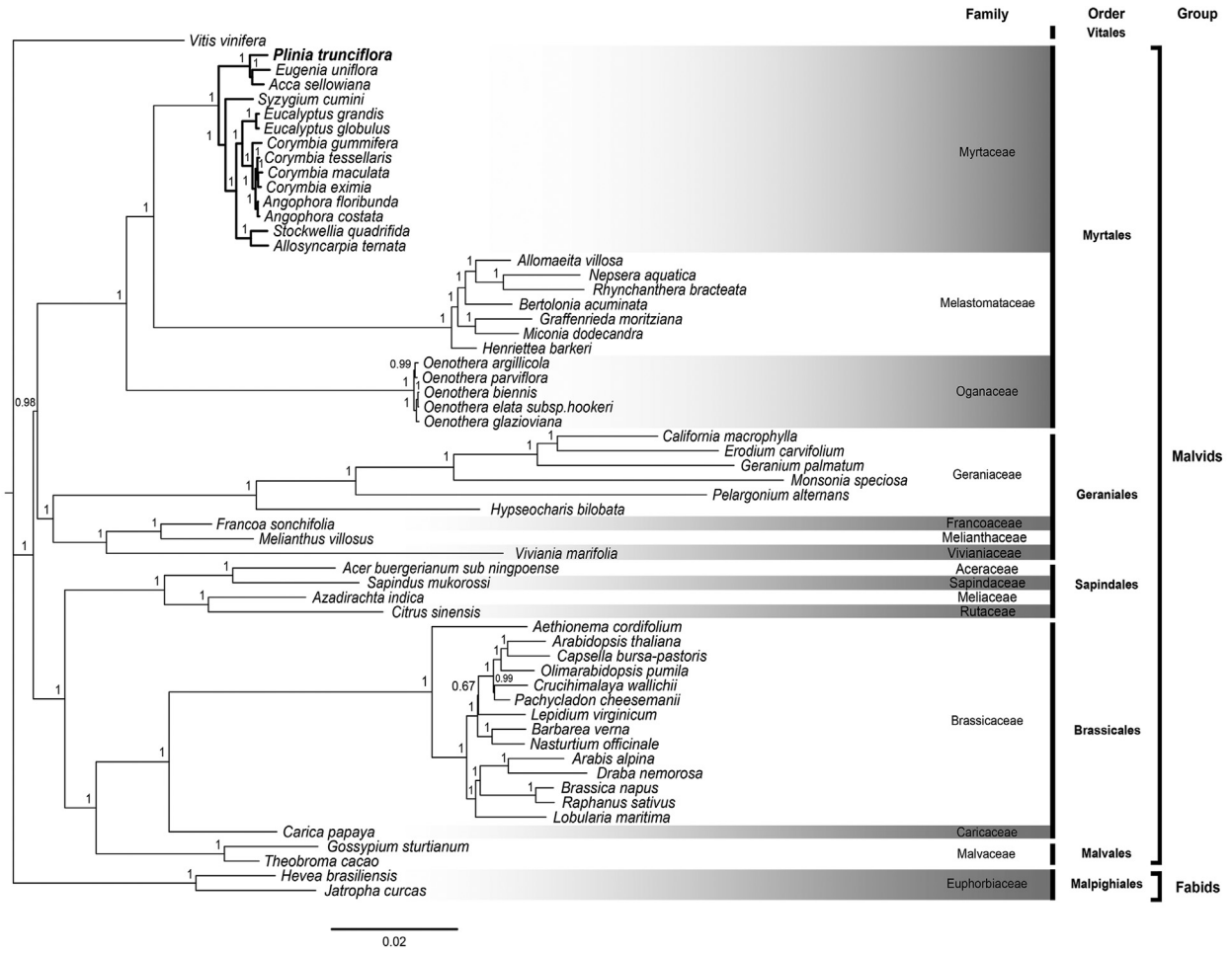

Figure 3 - Phylogenetic tree of Eurosids II based on 57 cp protein-coding genes generated by Bayesian method from 56 species. Bold branches indicate the Myrtaceae species. Numbers above each node are posterior probability values. Family, order and clade are also indicated. *Vitis vinifera* was considered as outgroup.

population and phylogenetic studies within this group. Moreover, these data increase the genetic and genomic resources available in Myrtaceae by adding a new strategy of organelle genome assembly.

Acknowledgments

This study was carried out with financial support from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Rio Grande do Sul (FAPERGS).

References

Bayly MJ, Rigault P, Spokevicius A, Ladiges PY, Ades PK, Anderson C, Bossinger G, Merchant A, Udovicic F, Woodrow IE, *et al.* (2013) Chloroplast genome analysis of Australian eucalypts - Eucalyptus, Corymbia, Angophora, Allosyncarpia and Stockwellia (Myrtaceae). Mol Phylogenet Evol 69:704-716.

Balerdi CF, Rafie R and Crane J (2006) Jaboticaba (*Myrciaria cauliflora*, Berg.) a delicious fruit with an excellent market potential. Proc Florida State Hortic Soc 119:66-68.

Biffin E, Lucas EJ, Craven L, Da Costa IR, Harrington MG and Crisp MD (2010) Evolution of exceptional species richness among lineages of fleshy-fruited Myrtaceae. Ann Bot 106:79-93.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K and Madden TL (2009) BLAST+: Architecture and applications. BMC Bioinformatics 10:421.

Costa JF, Lin SM, Macaya EC, Fernández-García C and Verbruggenet H (2016) Chloroplast genomes as a tool to resolve red algal phylogenies: A case study in the Nemaliales. BMC Evol Biol 16:205.

Craven LA and Biffin E (2005) *Anetholea anisata* transferred to, and two new Australian taxa of *Syzygium* (Myrtaceae). Blumea 50:157-162.

Doyle JJ and Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13-15.

Eguiluz M, Rodrigues FN, Guzman F, Yuyama P and Margis R (2017) The chloroplast genome sequence from *Eugenia uniflora*, a Myrtaceae from Neotropics. Plant Syst Evol doi: 10.1007/s00606-017-1431-x.

Guedes MNS, Rufini JCM, Azevedo AM and Pinto NAVD (2014) Fruit quality of jabuticaba progenies cultivated in a tropical climate of altitude. Fruits 69:449-458.

Li H and Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760.

Lohse M, Drechsel O and Bock R (2007) Organellar Genome DRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr Genet 52:267-274.

Lucas EJ, Harris SA, Mazine FF, Belsham SR, Lughadha EMN, Telford A, Gasson PE and Chase MW (2007) Suprageneric phylogenetics of Myrteae, the generically richest tribe in Myrtaceae (Myrtales). Taxon 56:1105-1128.

Machado LO, Vieira LD, Stefenon VM, Pedrosa OF, De Souza EM, Guerra MP and Nodari RO (2017) Phylogenomic relationship of feijoa (*Acca sellowiana* (O.Berg) Burret) with other Myrtaceae based on complete chloroplast genome sequences. Genetica 145:1-12.

Payn KG, Dvorak WS and Myburg AA (2007) Chloroplast DNA phylogeography reveals the island colonisation route of *Eucalyptus urophylla* (Myrtaceae). Aust J Bot 55:673-683.

Posada D and Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. Bioinformatics 14:817-818.

Ronquist F and Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE and Burleigh JG (2014) From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol Biol 14:23.

Sacchet C, Mocelin R, Sachett A, Bevilaqua F, Chitolina R, Kuhn F, Boligon AA, Athayde ML, Roman Junior WA, Rosemberg DB, *et al.* (2015) Antidepressant-like and antioxidant effects of *Plinia trunciflora* in mice. Evid Based Complement Alternat Med 2015:601503.

Stasi LC and Hiruma-Lima CA (2002) Myrtales medicinais. In: Stasi LC and Hiruma-Lima CA (eds) Plantas Medicinais na Amazônia e na Mata Atlântica. 2nd edition. Editora UNESP, São Paulo, pp 321-330.

Schattner P, Brooks AN and Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 33:W686-W689.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM and Birol I (2009) ABySS: A parallel assembler for short read sequence data. Genome Res 19:1117-1123.

Tamura K, Stecher G, Peterson D, Filipski A and Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725-2729.

Thornhill AH, Hope GS, Craven LA and Crisp MD (2012) Pollen morphology of the Myrtaceae. Part 4: Tribes Kanieae, Myrteae and Tristanieae. Aust J Bot 60:260-289.

Wyman SK, Jansen RK and Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252-3255.

Vasconcelos NCT, Proença EBC, Ahmad B, Aguilar SD, Aguilar R, Amorim SB, Campbell K, Costa RI, De-Carvalho SP, Faria EQJ, *et al.* (2017) Myrteae phylogeny, calibration, biogeography and diversification patterns: Increased understanding in the most species rich tribe of Myrtaceae. Mol Phylogenet Evol 109:113-137.

Internet Resources

Sobral M, Proença C, Souza M, Mazine F and Lucas E (2012) Myrtaceae in lista de espécies da flora do Brasil. Jardim Botânico do Rio de Janeiro [online], http://floradobrasil.jbrj.gov.br (accessed 16 September 2015).

Supplementary material

The following online material is available for this article:

Table S1 - List of 42 Myrtaceae chloroplast genomes used in chloroplast genome assembling of *Plinia trunciflora.*

Table S2 - List of 57 chloroplast protein coding genes used in the phylogenetic analysis.

Table S3 - List of 56 plastome sequences of Rosids included in the Bayesian phylogenetic analysis.

Associate Editor: Guilherme Corrêa de Oliveira

Plant Syst Evol
DOI 10.1007/s00606-017-1431-x

ORIGINAL ARTICLE

The chloroplast genome sequence from *Eugenia uniflora*, a Myrtaceae from Neotropics

Maria Eguiluz[1] · Nureyev F. Rodrigues[1] · Frank Guzman[2] · Priscila Yuyama[1] · Rogerio Margis[1,2,3]

Received: 28 September 2016 / Accepted: 28 May 2017

Abstract *Eugenia uniflora* is a plant native to tropical America that holds great ecological and economic importance. The complete chloroplast (cp) genome sequence of *Eugenia uniflora*, a member of the Neotropical Myrtaceae family, is reported here. The genome is 158,445 bp in length and exhibits a typical quadripartite structure of the large (LSC, 87,459 bp) and small (SSC, 18,318 bp) single-copy regions, separated by a pair of inverted repeats (IRs, 26,334 bp). It contains 111 unique genes, including 77 protein-coding genes, 30 tRNAs and 4 rRNAs. The genome structure, gene order, GC content and codon usage are similar to the typical angiosperm cp genomes. Comparison of the entire cp genomes of *E. uniflora* L. and three other Myrtaceae revealed an expansion of 43 bp in the intergenic spacer located between the IRA/large single-copy (LSC) border and the first gene of LSC region. Simple sequence repeat (SSR) analysis revealed that most SSRs are AT rich, which contribute to the overall AT richness of the cp genome. Additionally, fewer SSRs are distributed in the protein-coding sequences compared to the noncoding regions. Phylogenetic analysis among 58 species based on 57 cp genes demonstrated a closer relationship between *E. uniflora* L. and *Syzygium cumini* (L). Skeels compared to the Eucalyptus clade in the Myrtaceae family. The complete cp genome sequence of *E. uniflora* reported here has importance for population genetics, as well as phylogenetic and evolutionary studies in this species and other Myrtaceae species from Neotropical regions.

Keywords cpDNA · Fruit tree · Genome sequencing · NGS · Pitanga · Plant evolution

Introduction

Chloroplasts are multifunctional organelles, which possess their own genetic material and are believed to have originated from ancient endosymbiotic cyanobacteria (Ravi et al. 2008). The chloroplast (cp) genome in angiosperms usually varies between 115 and 165 kb in size and maintains highly conserved organization in most land plants. The lack of recombination, low rates of nucleotide substitutions (Wolfe et al. 1987) and primarily uniparental inheritance make plant cpDNA a valuable genetic source for phylogenetic relationship studies (Bayly et al. 2013). Sequence data from the plastid genome have transformed plant systematics and contributed greatly to unravel deep-level evolutionary relationships of taxonomically unresolved plant taxa (Jansen et al. 2007; Moore et al. 2010; Ruhfel et al. 2014).

The Myrtaceae (Myrtle, Eucalyptus, clove or guava family) is the eighth largest flowering plant family, and it is dominant among several vegetation types in South America through a variety of ecotypes (Pennington et al. 2009). *Eugenia* is the largest genus of Neotropical Myrtaceae

Handling editor: Marcus Koch.

Electronic supplementary material The online version of this article (doi:10.1007/s00606-017-1431-x) contains supplementary material, which is available to authorized users.

✉ Rogerio Margis
rogerio.margis@ufrgs.br

[1] PPGBM, Departamento de Genética, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brazil

[2] PPGBCM, Centro de Biotecnologia, sala 213, prédio 43431, Universidade Federal do Rio Grande do Sul - UFRGS, PO Box 15005, Porto Alegre, RS CEP 91501-970, Brazil

[3] Departamento de Biofísica, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brazil

Published online: 19 June 2017

family, encompassing about 5600 species, two-thirds of which are present in Brazilian ecosystems (Govaerts et al. 2015). *Eugenia* can be distinguished from the other genera of tribe Myrteae DC. by the generally 4-merous flowers, which have free calyx lobes that are separate in the flower bud, a non-tubular hypanthium that usually not extend beyond the tip of the bilocular multiovulate ovary, and finally by their embryo with cotyledons fused in a solid homogeneous mass (Mazine et al. 2014).

Eugenia uniflora L. is a fruit tree native to South America that serves as a good model for ecological studies because it grows in several different vegetation types, including forests, restingas, and arid and semiarid environments in the Brazilian northeast. This species is very versatile in terms of adaptability and plays a fundamental role in the maintenance of the shrubby coastal vegetation. Ecologically, it is an important food source for a variety of birds and mammals, and it can survive in disturbed sites within restinga habitats, especially near the beach (Almeida et al. 2012). Besides its ecologic importance, *E. uniflora* L. produces edible cherry-like fruits characterized by a low lipid and caloric content and by high amounts of polyphenols, carotenoids, and other antioxidant compounds (Spada et al. 2008) being traditionally used in folk medicine as antipyretic, stomachic, hypoglycemic, and to lower blood pressure (Lim 2012).

Despite the importance of the family, the phylogenetic relationships and delimitation of some genera are still debatable, especially in the fleshy fruit members. Many studies have provided insights into Myrtaceae phylogeny using nuclear ribosomal DNA and cp markers (Wilson et al. 2005; Lucas et al. 2007; Biffin et al. 2010; Thornhill et al. 2015; Berger et al. 2016). Although it has been recently published a phylogenetic work based on complete cp genome sequences from Myrteae tribe (Machado et al. 2017), most of these studies have been performed mainly on Eucalyptus and related genera (Steane 2005; Asif et al. 2013; Bayly et al. 2013; Reginato et al. 2016). Therefore, the availability of complete cp genomes exhibiting new variable and informative regions would help to reconstruct a more accurate phylogeny.

In this study, we present the cp genome of the fleshy fruit, *Eugenia uniflora*, obtained from whole genome sequencing and *de novo* assembly. This represents a solid resource for phylogenetic studies in the Myrtaceae family. We analyzed the genome features of *E. uniflora* and compared them with cp genomes from other Myrtaceae tribes. In addition, we performed a phylogenomic approach using 57 cp genes to reconstruct the phylogeny of Malvidae/Eurosid II group, which includes the Myrtales order.

Materials and methods

Plant material

Young leaves from *Eugenia uniflora* tree were collected from Porto Alegre, RS, Brazil (latitude (S): 30°4′2.71″; longitude (W): 51°7′11.88″). Voucher specimen was deposited at the Herbário do Instituto de Ciências Naturais (ICN 193277).

DNA sequencing and genome assembly

Total DNA was extracted from 1 g of fresh leaves using a CTAB method (Doyle and Doyle 1990). DNA quality was evaluated by electrophoresis on a 1% agarose gel, and quantification was determined using a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA).

Total DNA (10 µg) was sent to Fasteris SA (Plan-les-Ouates, Switzerland) for processing. One genomic paired-end library of 100-nt-long reads was generated using Illumina HiSeq 2000 platform (Illumina Inc., San Diego, CA, USA). To filter reads from the cp genome, the obtained paired-end sequence reads were aligned using Bowtie (Langmead 2010), against *Arabidopsis thaliana* Schur., *Glycine max* Merr., and 40 other Myrtaceae cp genomes (Online Resource 1) with a maximum of two mismatches per read. The filtered reads were assembled with ABYSS software (Simpson et al. 2009). The cp genome scaffolds were orientated by BLAST using the cp genome sequences of *Eucalyptus globulus* Labill and *Eucalyptus grandis* W.Hill as reference genomes (Altschul et al. 1990). Gap regions were filled in after Sanger sequencing using primers F: CATCCGCCAGGAGAGTTTAT, R: AAAGGG CCCTGCTATGAAAA and F: TCGGGTTGTGAGACAC ATTC, R: AACCCGCGTCTTCTCCTT. PCR was carried out in total volume of 20 µl containing 10 ng of DNA, 1× PCR buffer, 1.5 mM $MgCl_2$, 0.25 mM dNTP mix, 0.05 U of Platinum Taq DNA polymerase and 0.5 µM each of forward and reverse primers. The PCR cycle had an initial hot-start step at 94 °C for 5 min, followed by 35 cycles of 94 °C for 45 s, 60 °C for 1 min, 72 °C for 2 min and a final extension step at 72 °C for 5 min. Sanger sequencing reactions were performed using BigDye Terminator v3.1 Cycle sequencing kit and were resolved on ABI 3700 DNA Analyzer.

Genome analysis, codon usage, and repeat structure

Coding sequences (cds), rRNA, and tRNA were annotated using the automatic annotator DOGMA (Dual Organellar GenoME Annotator) (Wyman et al. 2004), verified using

BLAST searches against other plant cp genomes, and finally manually curated. tRNA genes were confirmed by comparison with the appropriate homologs in *Eucalyptus globulus* Labill cp genome and folding-verified with the tRNA scan-SE online program (http://lowelab.ucsc.edu/tRNAscan-SE). The codon usage frequency was analyzed by using MEGA (Tamura et al. 2007). A circular map of the genome was designed using the online OGDRAW program (Lohse et al. 2013). Whole chloroplast gene distribution was performed and visualized between *E. globulus* and *Syzygium cumini* (L.) Skeels. with mVISTA software using *E. uniflora* as the reference genome (Frazer et al. 2004).

The positions and type of simple sequence repeats (SSRs) were detected using MISA (http://pgrc.ipk-gatersleben.de/misa/), with thresholds of eight repeat units for mononucleotide SSRs, four repeat units for di- and trinucleotide SSRs, and three repeat units for tetra-, penta- and hexanucleotide SSRs. All of the repeats found were manually verified, and redundant results were removed. Tandem repeats were analyzed using Tandem Repeats Finder (TRF) v4.07b (Benson 1999) with the prior mentioned parameter settings. REPuter was used to identify and locate direct and inverted repeats in the cp genome of *E. uniflora* (Kurtz et al. 2001). The minimal repeat size was set to 30 bp, and the identity of repeats was no less than 90% (hamming distance equal to 3).

Phylogenetic analysis

Fifty-seven common cp protein-coding genes (PCGs) (Online Resource 2) were used to infer the phylogenetic relationships among 58 species belonging to the Malvids (Eurosids II) group available in GenBank (Online Resource 3). *Vitis vinifera* L. was set as out-group. Nucleotide sequences were aligned by MUSCLE available in MEGA version 6.0 (Tamura et al. 2007). Phylogenetic trees were generated by the maximum likelihood (ML) method, using the GTR+I+G nucleotide substitution determined by Modeltest ver. 3.7 (Posada and Crandall 1998), using RAxML v8.2.4 (Stamatakis 2014). The stability of each tree node was tested by bootstrap analysis with 1000 replicates. Bayesian analysis on the same dataset was also performed using MrBayes version 3.1.2 (Ronquist and Huelsenbeck 2003). We used the same evolutionary model with 5,000,000 generations sampled every 100 generations. The first 25% of trees were discarded as burn-into produce a consensus phylogram, with posterior probability (PP) values for each node. The phylogenetic trees were rooted and visualized using FigTree software (http://tree.bio.ed.ac.uk/software/figtree/).

Results

Genome assembly

Reads from Illumina sequencing of the *Eugenia uniflora* nuclear genome were used to assemble the cp genome. The total reads (75,127,218) were filtered and assembled *de novo* into non-redundant contigs and singletons joined into 10 scaffolds. This first draft of the cp genome resulted in mapped reads covering about 99.9% of the genome (coverage 10,938 reads, minimum coverage = 757 reads, maximum coverage = 26 327 reads).

After running BLAST with Eucalyptus genomes, the cp genome sequences resulted in two large scaffolds whose ends were finally closed using PCR and Sanger sequencing. The four junctions between IRs and SSC/LSC were determined by aligning the *E. uniflora* cp genome versus *E. globulus* and *Syzygium cumini* genomes. The final cp genome was then submitted to GenBank (accession number NC_027744).

The overall structure and general features of the *Eugenia uniflora* cp genome

The complete length of the *Eugenia* cp genome is 158,445 bp, and it includes the canonical quadripartite structure consisting of one LSC (87,459 bp), one SSC (18,318 bp) and a pair of IRs (26,334 bp) (Fig. 1). Coding regions (92,848 bp; 58.93%) account for over half of the cp genome, with the peptide-coding regions forming the largest group (81,462 bp; 51.41%), followed by ribosomal RNA genes (9050 bp; 5.71%) and transfer RNA genes (2863 bp; 1.81%). The remaining 41.07% is covered by intergenic regions, introns or pseudogenes (Table 1). The average total AT content is 63% with the IRs having lowest amount (57.2%). A total of 111 different genes, including 30 tRNAs, 4 rRNAs and 77 predicted protein-coding genes, were annotated (Table 2). Among these, seven tRNAs, four rRNAs and six protein-coding genes (*ycf*15, *rps*7, *ndh*B, *ycf*2, *rpl*23, *rpl*2) were present in duplicate in the IR regions. Three pseudogenes, *ycf*1, *ycf*15 *and inf*A, were identified and located in the boundary IRb/SSC, IRb and LSC region, respectively. In the *Eugenia* cp genome, there are 18 gene containing introns, the majority of them (12 genes) are located in the LSC region (four tRNAs and eight protein-coding genes) and the rest are distributed in IRs (two tRNA and three protein-coding genes) and SSC (1 protein-coding gene) region (Table 3). Most of the genes have only one intron, but *clp*P and *ycf*3 have two introns each. The *trn*K$^{(UUU)}$ gene has the largest intron (2530 bp) containing within it the *mat*K gene. The *rps*12 gene sequence is a trans-spliced gene with the 5′end located in

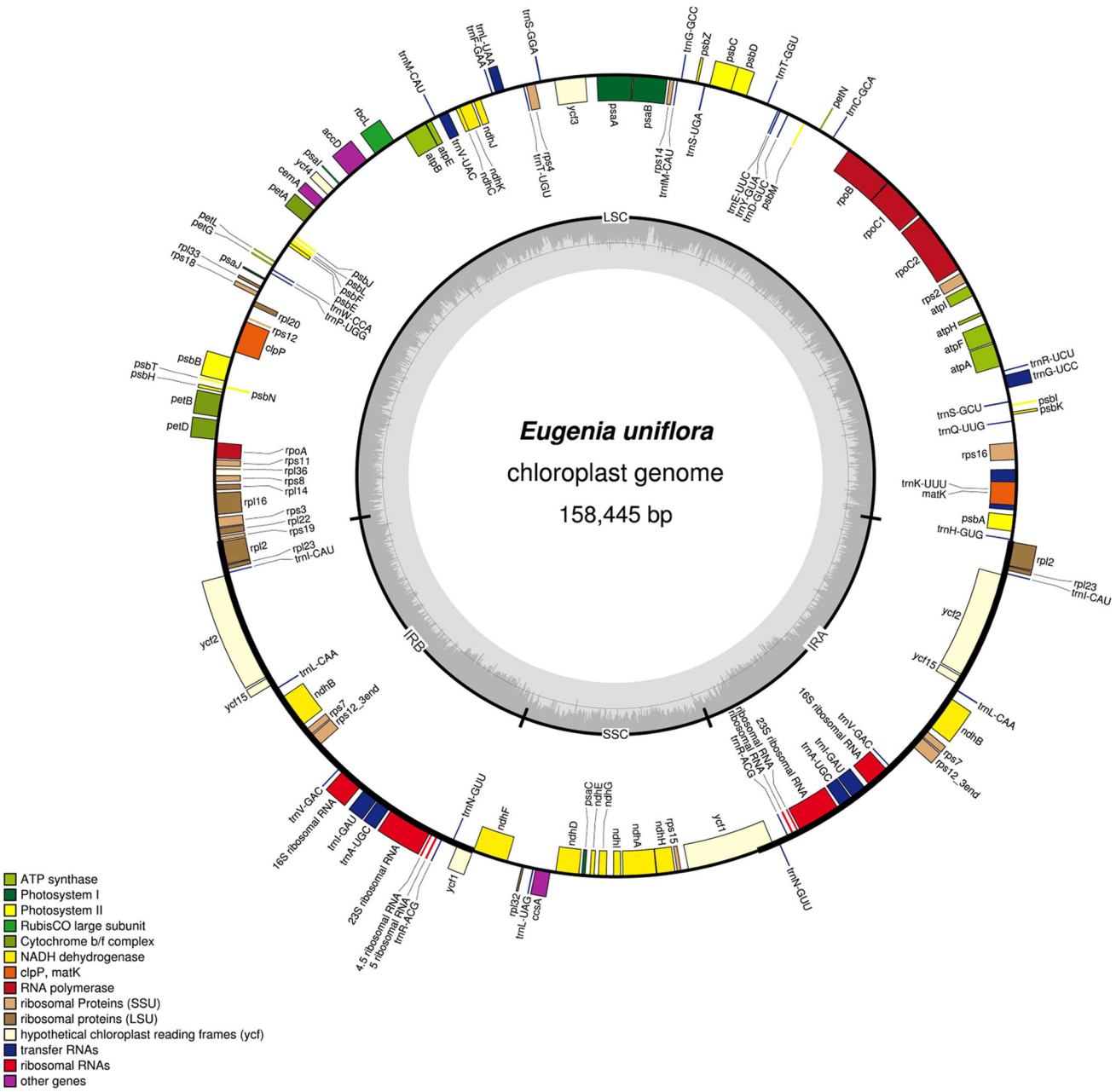

Fig. 1 *Eugenia uniflora* chloroplast genome map. The *thick lines* indicate the extent of the inverted repeats (IRa and IRb), which separate the genome into small and large single-copy regions. Genes on the outside of the map are transcribed clockwise and those on the inside of the map are transcribed counterclockwise. GC content is shown. Gene function or identifiers are displayed by different *colors* as it is indicated by inner legend

the LSC region and the duplicated 3′end in the IR regions. Based on the sequences of protein-coding genes and tRNA genes, the frequency of codon usage was deduced for the cp genome and is summarized in Table 4. The codon usage was biased toward a high representation of A and U at the third codon position, as observed in most land plant cp genes (Ravi et al. 2008).

Comparison of *Eugenia uniflora* to other Myrtaceae cp genomes

The overall sequence alignment of *E. globulus* and *S. cumini* cp genomes was compared using the annotation of *Eugenia uniflora* as a reference. The same order of genes was confirmed because order variations in cp genomes are

Table 1 Summary of the characteristics of *Eugenia uniflora* chloroplast genome

Feature	*E. uniflora*
Total cpDNA size (bp)	158,445
LSC size (bp)	87,459
SSC size (bp)	18,318
IR size (bp)	26,334
Protein-coding regions (%)	58.6%
rRNA and tRNA (%)	7.52%
Introns size (% total)	12.05%
Intergenic sequences and pseudogenes (%)	29.02%
Number of genes	131
Number of different protein-coding genes	77
Number of different tRNA genes	30
Number of different rRNA genes	4
Number of different duplicated genes	17
Pseudogenes	3
GC content	37%

relatively uncommon. The two IRs from the three cp genomes show high similarity in sequence (Fig. 2), on the other hand, the most divergent regions were those localized in the intergenic spacers in the noncoding genes. The coding region sequences show a high level of conservation. Slightly more sequence variation was observed between *E. uniflora* and *E. globulus* cp genomes in the *psa*A, *psa*B and *ycf*2 genes, compared with *S. cumini*.

IR contraction and expansion

In general, *E. uniflora* has the smallest cp genome compared to *E. globulus*, *E. grandis* and *S. cumini* and shows an expansion of the IR over the LSC region (Fig. 3). This also explains the presence of pseudogenes in the border regions, like *ycf*1 in which length variation depends upon if the IR has extended into the SSC region. In the case of *E. uniflora*, a shorter *ycf*1 pseudogene and a larger *ndh*F gene cause a reduction in the intergenic sequence. This last gene is relatively highly variable in the 3′ region (Dong et al.

Table 2 Genes present in *Eugenia uniflora* chloroplast genome

Category	Group of genes	Name of genes
Self-replication	Large subunit of ribosomal proteins	*rpl*2[b,c], 14, 16[b], 20, 22, 23[c], 32, 33, 36
	Small subunit of ribosomal proteins	*rps*2, 3, 4, 7[c], 8, 11, 12[b–d], 14, 15, 16[b], 18, 19
	rRNA genes	*rrn*4.5, 5, 16, 23
	tRNA genes	*trn*A[(UGC)b,c], C[(GCA)], D[(GUC)], E[(UUC)], F[(GAA)], G[(UCC)b,c], G[(GCC)], H[(GUG)], I[(CAU)c], I[(GAU)b,c], K[(UUU)b], L[(UAG)], L[(CAA)c], L[(UAA)b], M[(CAU)], fM[(CAU)], N[(GUU)c], Q[(UUG)], R[(ACG)c], R[(UCU)], S[(GGA)], S[(GCU)], S[(UGA)], T[(GGU)], T[(UGU)], V[(UAC)b], V[(GAC)c], W[(CCA)], Y[(GUA)], P[(UGG)]
Photosynthesis	Photosystem I	*psa*A, B, C, I, J, *ycf*3[a], *ycf*4
	Photosystem II	*psb*A, B, C, D, E, F, H, I, J, K, L, M, N, T, Z
	NADH oxidoreductase	*ndh*A[b], B[b,c], C, D, E, F, G, H, I, J, K
	Cytochrome b6/f complex	*pet*A, B[b], D[b], G, L, N
	ATP synthase	*atp*A, B, E, F[b], H, I, L
	Rubisco	*rbc*L
Other gene	Maturase	*mat*K
	Protease	*clp*P[a]
	Envelop membrane protein	*cem*A
	Subunit Acetyl-CoA carboxylase	*acc*D
	c-type cytochrome synthesis gene	*ccs*A
Unknown gene	Conserved open reading frames	*ycf*1, *ycf*2[c], *ycf*15[c]

[a] Genes containing two introns

[b] Genes containing a single intron

[c] Genes with two copies

[d] Genes split into two independent transcription units

Table 3 Genes with introns in the *Eugenia uniflora* chloroplast genome and the length of the exons and introns

Gene	Location	exon I (bp)	intron I (bp)	exon II (bp)	intron II (bp)	exon III (bp)
*trn*K$^{(UUU)}$	LSC	37	2568	35		
*rps*16	LSC	39	867	204		
*trn*G$^{(UCC)}$	LSC	23	755	49		
*atp*F	LSC	147	742	408		
*rpo*C1	LSC	453	729	1614		
*ycf*3	LSC	126	758	228	727	148
*trn*L$^{(UAA)}$	LSC	37	502	46		
*trn*V$^{(UAC)}$	LSC	39	600	37		
*clp*P	LSC	69	866	291	619	223
*pet*B	LSC	6	771	639		
*pet*D	LSC	9	752	471		
*rpl*16	LSC	9	1000	396		
*rpl*2	IR	390	664	432		
*ndh*B	IR	777	681	753		
*rps*12	IR	210	567	27		
*trn*I$^{(GAU)}$	IR	37	957	35		
*trn*A$^{(UGC)}$	IR	38	803	35		
*ndh*A	SSC	549	1067	537		

2012). The intergenic spacer located between the IRA/LSC border and the *trn*H gene of the LSC region established differences between the cp genomes. This region is 43 bp in *E. uniflora,* similar to that of *S. cumini* (55 bp), but different from other dicots where it ranges in size of 2–12 bp (Shinozaki et al. 1986; Ibrahim et al. 2006).

Repeat structure and SSR analysis

For repeat structure analysis, eleven forward, one inverted, and twelve tandem repeats were detected in the *E. uniflora* cp genome (Table 5). Most of these repeats (67%) exhibited lengths between 20 and 50 bp and were located in intergenic spacers regions and introns. The coding regions of *psa*A, *psa*B, *ycf*1 and *ycf*2 genes showed some repeated sequences. Although the number of repeats was variable respect to *Syzygium* and *Eucalyptus*, they were identified in the same genes. Most of the repeated regions identified in this work have already been compared in *S. cumini, Eugenia grandis, E. globulus, Nicotiana tabacum* L., *Gossypium barbadense* L. and show a high degree of conservation (Asif et al. 2013). It appears that dispersed repeats are very common in angiosperm cp genomes, but future comparative studies are needed to determine the functional and evolutionary role of these repeats.

SSRs are repeated DNA sequences consisting of direct tandem repeats of short (1–10 bp) nucleotide motifs. In this study, a total of 215 SSR loci were identified, most of them (76.25%) were A and T mononucleotide repeats (Table 6) similar to other Myrtaceae cp genomes (Asif et al. 2013). Most SSRs are located in intergenic regions, but some were found in *ndh*F, *pet*A, *ycf*2, *rpo*C2, *psa*J, *psb*B, *ycf*1, *ccs*A, *ycf*4 and *rps*19 coding genes (Table 6).

Phylogenetic analysis

In this study, the concatenated nucleotide sequences of 57 PCGs of 58 cp genomes of Malvidae group were used to reconstruct the phylogenetic relationships by the ML and Bayesian method. These 57 genes were present in all the cp genomes so the problem of missing data from the sequence alignment was minimized. The sequence alignment used comprised 36,206 characters. The final alignment was submitted and assigned as 21,047 in the TreeBASE database (https://treebase.org/). ML analysis resulted in a single tree with ln $L = -249{,}032.011$, and bootstrap values were high with values >80% for 4 of 55 nodes, and 48 nodes with 100% bootstrap (Online Resource 4). Although the Bayesian and ML analyses showed similar topologies, the posterior probabilities in the Bayesian analysis were better than the bootstrap values in the ML (Fig. 4). Therefore, only the Bayesian tree was chosen for discussing the phylogenetic results.

There are congruence areas strongly supported by the phylogeny (PP = 1.0) that include the monophyly of Brassicales and their sister relationship to Malvales and Sapindales and monophyly of Geraniales and Myrtales. Our phylogenies placed Myrtales in a sister relationship to Geraniales with solid support and resolution (PP = 0.95),

Table 4 Codon–anticodon recognition pattern and codon usage for the *Eugenia uniflora* chloroplast genome

Codon	Aminoacid	Count	RSCU	*trn*A	Codon	Aminoacid	Count	RSCU	*trn*A
UUU	F	2308	1.19	*trn*F$^{(GAA)}$	UAU	Y	1456	1.34	*trn*Y$^{(GUA)}$
UUC	F	1587	0.81		UAC	Y	715	0.66	
UUA	L	1080	1.19	*trn*L$^{(UAA)}$	UAA	*	1225	1.21	
UUG	L	1160	1.28	*trn*L$^{(CAA)}$	UAG	*	855	0.84	
CUU	L	1110	1.22	*trn*L$^{(UAG)}$	CAU	H	967	1.4	*trn*H$^{(GUG)}$
CUC	L	717	0.79		CAC	H	416	0.6	
CUA	L	848	0.94		CAA	Q	1102	1.39	*trn*Q$^{(UUG)}$
CUG	L	526	0.58		CAG	Q	478	0.61	
AUU	I	1888	1.21	*trn*I$^{(GAU)}$	AAU	N	1819	1.39	*trn*N$^{(GUU)}$
AUC	I	1230	0.79		AAC	N	795	0.61	
AUA	I	1565	1	*trn*I$^{(CAU)}$	AAA	K	2172	1.32	*trn*K$^{(UUU)}$
AUG	M	958	1	*trn*(f)M$^{(CAU)}$	AAG	K	1117	0.68	
GUU	V	839	1.36	*trn*V$^{(GAC)}$	GAU	D	1025	1.41	*trn*D$^{(GUC)}$
GUC	V	437	0.71		GAC	D	429	0.59	
GUG	V	446	1.22		GAA	E	1379	1.38	*trn*E$^{(UUC)}$
GUA	V	754	0.72	*trn*V$^{(UAC)}$	GAG	E	622	0.62	
UCU	S	1117	1.48	*trn*S$^{(GGA)}$	UGU	C	667	1.2	*trn*C$^{(GCA)}$
UCC	S	855	1.13		UGC	C	449	0.8	
UCG	S	602	0.8		UGA	*	956	0.94	
UCA	S	861	1.14	*trn*S$^{(UGA)}$	UGG	W	704	1	*trn*W$^{(CCA)}$
CCU	P	662	1.07	*trn*P$^{(UGG)}$	CGU	R	321	0.6	*trn*R$^{(ACG)}$
CCC	P	564	0.92		CGC	R	252	0.47	*trn*R$^{(UCU)}$
CCA	P	799	1.3		CGA	R	577	1.08	
CCG	P	440	0.71		CGG	R	363	0.68	
ACU	T	647	1.17	*trn*T$^{(GGU)}$	AGA	R	1079	2.01	
ACC	T	549	0.99		AGG	R	626	1.17	
ACG	T	362	0.65		AGU	S	627	0.83	*trn*S$^{(GCU)}$
ACA	T	656	1.19	*trn*T$^{(UGU)}$	AGC	S	471	0.62	
GCU	A	469	1.3	*trn*A$^{(UGC)}$	GGU	G	510	0.95	*trn*G$^{(GCC)}$
GCC	A	329	0.91		GGC	G	330	0.62	
GCA	A	420	1.16		GGG	G	537	1	
GCG	A	225	0.62		GGA	G	764	1.43	*trn*G$^{(UCC)}$

despite the fact that this order still has a controversial position in respect to other members of the Rosids (Fig. 4).

In analyzing the Myrtales clade, we showed a closer relationship between species from Melastomataceae and Myrtaceae family than to Onagraceae family. Our phylogenetic tree clearly supports the monophyly of the three Myrtoideae tribes: Myrteae, Eucalypteae and Syzygieae (PP = 1.0). Additionally, we corroborated the paraphyly of Corymbia in the Eucalypteae tribe and observed that the latter has a closer relationship to Syzygieae than Myrteae (Bayly et al. 2013). *Eugenia uniflora* is placed along with *Acca sellowiana* (O.Berg) Burret as the diverging lineage, and they have a closer relationship with *S. cumini* (Syzygieae tribe) than to the Eucalypteae tribe.

Discussion

The cp genome of *Eugenia uniflora* was assembled *de novo* from the Illumina NGS reads derived from the whole genome. This approach, without prior purification of the cpDNA, provides a new way to obtain the cp genome and has been successful in several studies (Leseberg and Duvall 2009; Tangphatsornruang et al. 2010; Straub et al. 2011). Our work serves as another example of this approach for obtaining high coverage (99%) of the cp genome.

The *E. uniflora* cp genome has the typical quadripartite structure (Fig. 1) and gene content with a size in range with other Myrtaceae family members (Asif et al. 2013; Bayly et al. 2013; Machado et al. 2017). Major differences among angiosperm cp genomes are due to gene loss, inversions,

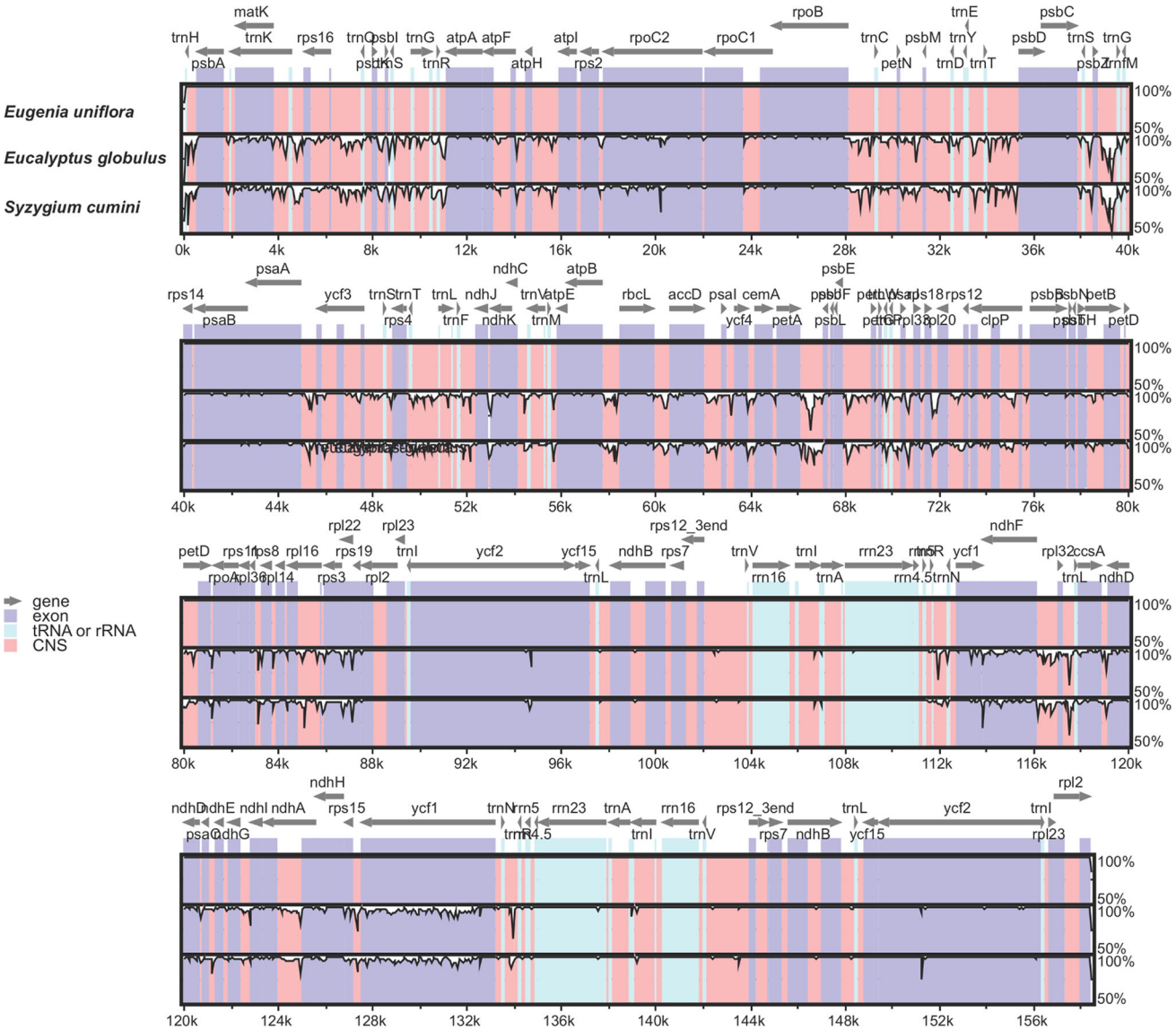

Fig. 2 Sequence identity *plot* comparing the chloroplast genome of *Eugenia uniflora* to other Myrtaceae. Pairwise comparisons between *E. uniflora* and *E. globulus* (*top*) and *Syzygium cumini* (*bottom*) chloroplast genomes using mVISTA. The *y*-axis represents % identity ranging from 50 to 100%. Coding, rRNA, tRNA and conserved noncoding sequences (CNS) are shown as indicated by inner legend

and expansion/contraction of inverted repeat regions. The IR contraction and expansion events, the presence of many stop codons in the coding sequence, or a probable partial duplication are all reasons that could explain the presence of pseudogenes in the cp genome. In our work, this is represented in the *ycf*1, *inf*A and *ycf*15 pseudogenes (Fig. 1; Table 1). Some alternative codons were also identified, ACG was used as an alternative initiation codon in the *psb*L and *ndh*D genes and GUG was only found as a start codon in the *ycf*15 pseudogene and *rps*19 gene. ACG has been shown to convert to an AUG initiation site as reported in *N. tabacum* (Sasaki 2003), and GUG codons have been reported to be more efficient than ACG in translation initiation (Rohde et al. 1994). Most cp genomes are quite AT rich with (above 60%) unevenly distributed AT contents, as well as conserved regions of lower AT contents. The features of the *E. uniflora* cp genome are the same, and of all the cp regions, the IRs have the lowest AT content (57.2%) because of the presence of ribosomal genes (Ravi et al. 2008). These values are congruent with those reported in other Myrtaceae cp genomes (Asif et al. 2013; Bayly et al. 2013; Machado et al. 2017).

Chloroplast SSRs (cpSSRs) are generally short mononucleotide tandem repeats that, when located in the noncoding regions of the cp genome, commonly show intraspecific variation in repeat number. CpSSRs can exhibit high variation within the same species and thus are considered valuable markers for population genetics

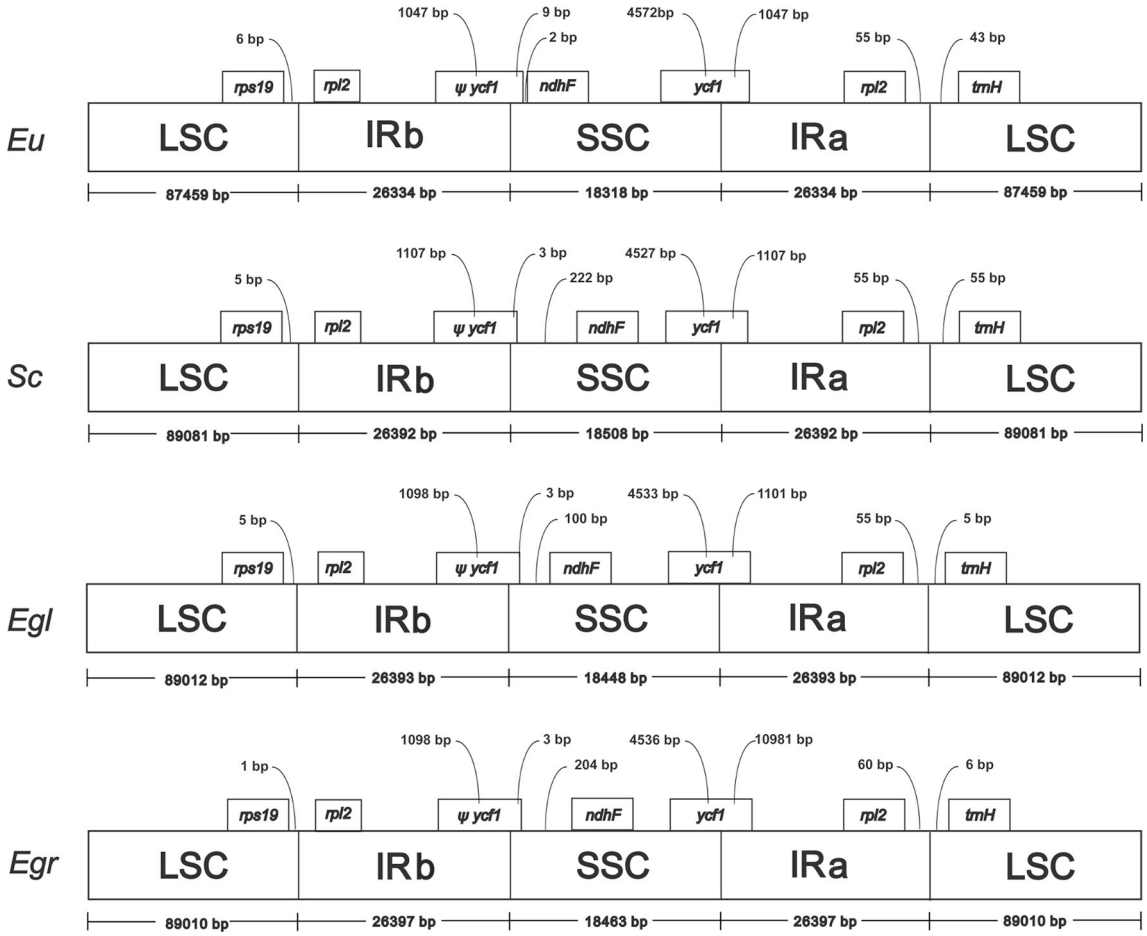

Fig. 3 Comparison of border positions of LSC, SSC and IR among *Eugenia uniflora* and related Myrtaceae family species. *Boxes* above the *main line* indicate the predicted genes, while pseudogenes at the borders are shown by Ψ. Their length is displayed in the corresponding regions. The figure is not scaled and just shows relative changes at or near the IR-SC borders. Sc *Syzygium cumini*, Egl *E. globulus*, Egr *E. grandis*

(Provan et al. 2001). In this work, we identified some SSRs that can be utilized to increase our understanding of the genetic structure of *E. uniflora* populations (Margis et al. 2002; Salgueiro et al. 2004; Ferreira-Ramos et al. 2008). Understanding the effects of spatial isolation on the levels of genetic diversity and gene flow is crucial to providing recommendations for in situ and ex situ conservation of the species. In addition, these SSR markers will also be useful in future studies of other Myrtaceae species from the Neotropics.

Although previous phylogenetic studies improved our understanding of intergeneric relationships within the Myrtales order, the relationship between fleshy-fruited and dry-capsular clades remains unresolved. In this work, some representative cp genomes from Melastomataceae, Onagraceae and Myrtaceae family were selected to build a Malvidae metatree. We used species from the Malvidae group because the order Myrtales belongs to this group and there are several cp genomes available. To do this, 57 protein-coding genes for 13 taxa were analyzed using both the ML and Bayesian methods. Both trees are congruent to that presented in a recent study using 78 cp coding genes from 30 angiosperm taxa (Ruhfel et al. 2014) and to that using 72 complete cp genomes from Rosids (Su et al. 2014). Although our results clearly favor a closer relationship of Myrtales to the Geraniales clade, expanded sampling of complete cp genome sequences of Rosids is needed to resolve this issue, especially since limited taxon sampling can lead to erroneous tree topologies (Leebens-Mack et al. 2005).

Eugenia uniflora formed one monophyletic clade along with *A. sellowiana*, another Myrtaceae from Neotropical region, as previously reported by Machado et al. 2017 using complete cp genomes. These two species were more closely related to *Syzygium cumini* than the Eucalypteae tribe. The Syzygieae tribe has had a long association with the predominantly New World Myrtaceous, mostly because they showed a high similarity between their cp complete

Table 5 Repeated sequences in the *Eugenia uniflora* chloroplast genome

Repeat size (bp)	Start position of first repeat	Type[a]	Start position the repeat found in other region	Copy number	Location[b]	Region
15	55,610	T	55,625	(×2)	IGS (*trn*M$^{(CAU)}$-*atp*E)	LSC
15	130,810	T	130,825, 130,840	(×3)	*ycf*1	SSC, IRB
16	10,031	T	10,047	(×2)	intron (*trn*S$^{(UCC)}$)	LSC
16	87,134	T	87,150	(×2)	IGS (*rpl*22-*rps*19)	LSC
17	102,403	T	102,420	(×2)	IGS (*rps*12-*trn*V$^{(GAC)}$)	IRA
18	66,388	T	66,406	(×2)	IGS (*pet*A-*psb*J)	LSC
18	94,667	T	94,685, 94,703	(×3)	*ycf*2	IRA
20	6373	T	6393	(×2)	IGS (*rps*16-*trn*Q$^{(UUG)}$)	LSC
20	39,036	T	39,056	(×2)	IGS (*psb*Z-*trn*G$^{(GCC)}$)	LSC
20	70,664	T	70,684	(×2)	IGS (*psa*J-*rpl*33)	LSC
21	92,239	T	92,260, 92,281	(×3)	*ycf*2	IRA
31	92,238	F	92,259	(×2)	*ycf*2	IRA
31	45,820	F	102,036	(×2)	intron I (*ycf*3); IGS (*rps*12-*trn*V$^{(GAC)}$)	LSC, IRA
31	153,617	F	153,638	(×2)	*ycf*2	IRB
32	8762	F	38,011	(×2)	IGS (*psb*I-*trn*S$^{(GCU)}$, *trn*S$^{(GCU)}$; IGS (*psb*C-*trn*S$^{(UGA)}$, *trn*S$^{(UGA)}$)	LSC
35	70,665	I	70,665	(×2)	IGS (*psa*J-*rpl*33)	LSC
39	46,761	F	102,014	(×2)	intron II (*ycf*3); IGS (*rps*12-*trn*V$^{(GAC)}$)	LSC, IRA
40	102,014	F	123,971	(×2)	*rps*12; intron (*ndh*A)	IRA, SSC
41	41,804	F	44,028	(×2)	*psa*B; *psa*A	LSC
42	46,758	F	123,968	(×2)	intron II (*ycf*3); intron *ndh*A	LSC, SSC
45	94,666	F	94,684	(×2)	*ycf*2	IRA
45	151,175	F	151,193	(×2)	*ycf*2	IRB
50	38,352	T	38,402	(×2)	IGS (*trn*S$^{(UGA)}$-*psb*Z)	LSC
62	38,351	F	38,401	(×2)	IGS (*trn*S$^{(UGA)}$-*psb*Z)	LSC

[a] *F* Forward; *I* Inverted; *T* Tandem

[b] *IGS* intergenic spacer region

genome sequences. Additionally, they have characteristics in common such as fleshy large-seeded fruits, biotic dispersal, and they are both woody rainforest trees. These results are in agreement with Biffin et al. (2010), who concluded that Syzygieae and Myrteae show highly significant positive variation in diversification rates associated with both of these lineages relative to the overall evolutionary radiation of Myrtaceae. Our phylogenetic tree also confirmed the closer relationship between Melastomataceae and Myrtaceae than to the Onagraceae family as reported by previous analyses based on complete cpDNA (Berger et al. 2016; Reginato et al. 2016). Our phylogenetic analyses based on complete cp genomes further expand the taxon sampling of entire genomes as we included one more Neotropical Myrtaceae genome in a metatree analysis.

Conclusions

The *Eugenia uniflora* cp genome organization and gene content are typical of most angiosperms and are similar to that of Myrtaceae species. It features a relevant number of simple sequence repeats, which could be further explored for population studies within the *Eugenia* genus. Moreover, these data increase the genetic and genomic resources available in Myrtaceae by adding a new strategy of

Table 6 List of simple sequence repeats in *Eugenia uniflora*. The SSR-containing coding regions are indicated in parentheses

Repeat unit	Length (bp)	Number of SSRs	Start position
A	8	31	1992; 4532; 6547; 6772; 8905; 9344; 14,290; 19,917; 23,570; 23,757; 39,529; 45,575; 45,600; 55,722; 62,539; 65,559 (*pet*A); 68,727; 68,795; 72,516; 75,695; 81,223; 113,806 (*ndh*F); 114,538 (*ndh*F); 118,402; 120,673; 120,744; 131,958; 139,500; 143,496; 146,803; 158,425
	9	22	17; 7940; 9048; 12,656; 13,443; 14,472; 31,941; 32,744; 32,817; 38,913; 47,458; 48,054; 57,737; 71,263; 71,766; 92,859 (*ycf*2); 116,745; 117,617; 118,856; 122,698; 126,805; 134,276
	10	10	303; 4705; 4780; 8176; 47,420; 48,211; 48,271; 62,286; 74,562; 131,613
	11	7	5678; 8732; 50,518; 75,032; 117,327; 124,318; 130,258
	12	2	60,317; 84,317
	13	1	8707; 74,079
	15	1	14,765
	19	1	32,343
T	8	33	4296; 5792; 8338; 18,391 (*rpo*C2); 29,269; 31,642; 37,906; 45,860; 69,550; 70,515 (*psa*J); 70,892; 74,164; 76,564 (*psb*B); 78,464; 84,363; 85,593; 85,682; 87,473; 99,095; 102,402; 106,398; 117,953 (*ccs*A); 118,488 (*ccs*A); 119,042; 119,067; 119,739; 127,133; 127,935 (*ycf*1); 128,476 (*ycf*1); 130,326 (*ycf*1); 131,338 (*ycf*1); 131,455 (*ycf*1); 131,573 (*ycf*1)
	9	22	141; 2481; 9565; 14,030; 19,668; 31,404; 34,703; 47,358; 49,898; 54,438; 62,923; 68,293; 74,663; 87,435 (*rps*19); 111,621; 117,239; 122,789; 124,393; 128,707; 130,023 *(ycf*1); 130,846 (*ycf*1); 153,038 (*ycf*2)
	10	17	7863; 9093; 10,950; 15,525; 22,370 (*rpo*C1 - exon II); 27,435 (*rpo*B); 45,944 (*ycf*3 - intron I); 47,321 (*ycf*3 - intron II); 54,299; 54,696; 57,631; 70,066; 72,927; 74,602; 75,767; 85,868; 86,670 (*rpl*22)
	11	8	13,215; 17,551; 19,774 (*rpo*C2); 63,472 (*ycf*4); 66,969; 73,028; 73,341; 124,796
	12	2	66,289; 73,896
	13	1	70,607
	14	2	15,692; 53,692 (*ndh*K)
	15	1	83,775
	20	1	51,433
C	8	2	39,678; 65,485 (*pet*A)
AG	8	2	98,454 (*ndh*B - exon I); 136,156 (*rrn*23)
AT	8	15	1884; 10,527; 45,420; 58,748 (*rbc*L); 60,365; 60,817 (*acc*D); 62,673; 65,199 (*pet*A); 66,816; 70,297; 87,039 (*rpl*22); 124,099; 127,519; 148,203; 157,836
	10	1	33,835
CA	8	1	3100
CT	8	3	31,961; 109,742 (*rrn*23); 147,444 (*ndh*B - exon II)
GA	8	4	38,017 ($trnS^{(UGA)}$; 58,932 (*rbc*L); 90,677 (*ycf*2); 92,880 (*ycf*2)
TA	8	5	7506; 88,061; 96,251 (*ycf*2); 97,694; 149,647 (*ycf*2)
TC	8	3	131,255 (*ycf*1); 153,018 (*ycf*2); 155,221 (*ycf*2)
	10	1	64,285 (*cem*A)
AGA	12	1	139,167
CAG	12	1	1177 (*psb*A)
TTA	12	1	68,856
TTC	12	1	106,726
AATA	12	1	119,348 (*ndh*D)
AGAT	12	1	4894
ATAG	12	1	115,884 (*ndh*F)
ATTA	12	1	33,664
ATTT	12	1	11,090
CTTG	12	1	29,446
TAAG	12	1	46,202
TAAT	12	1	129,206 (*ycf*1)
TCTT	12	1	63,902
TTAT	12	1	78,171
TTTC	12	2	78,202; 85,555

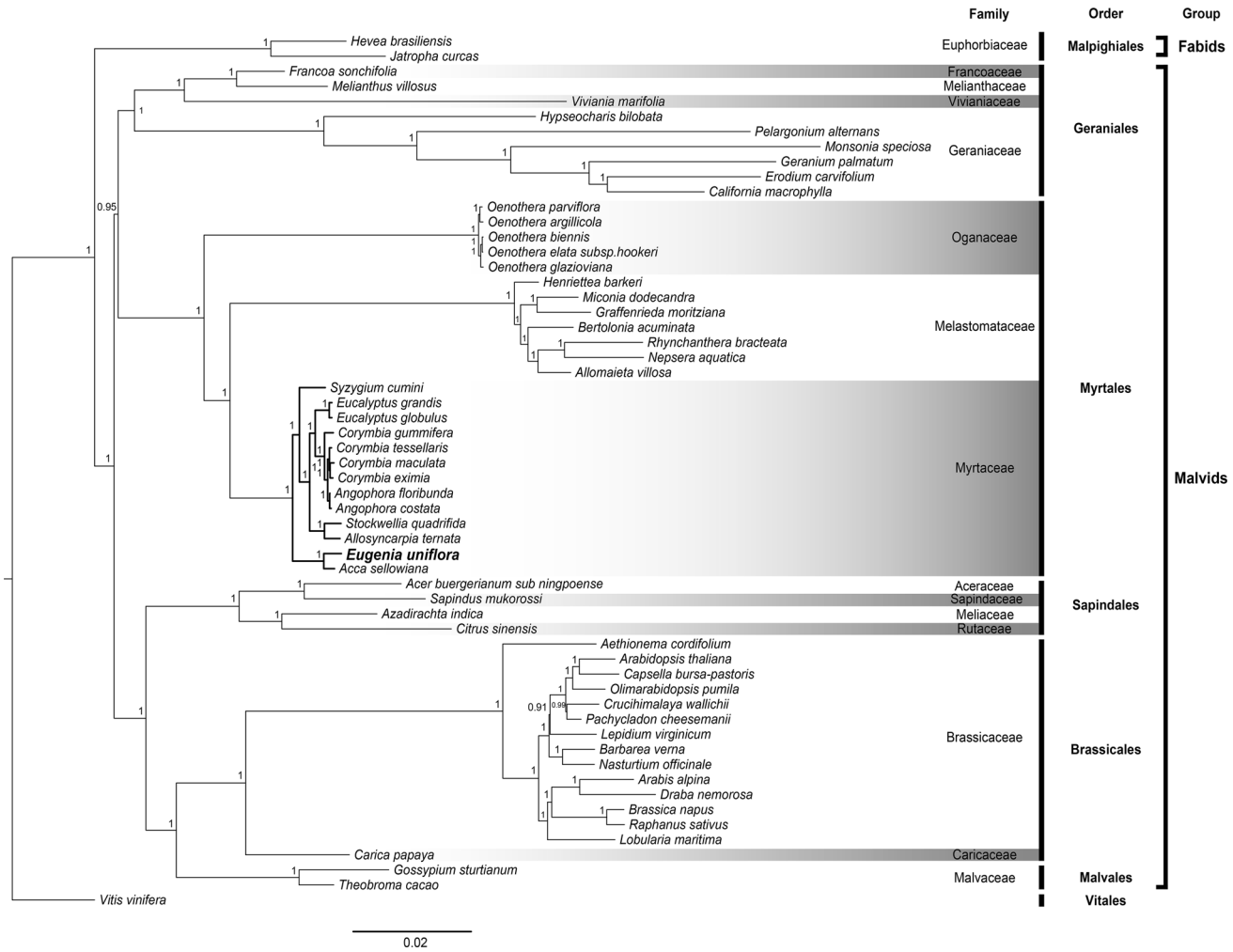

Fig. 4 Phylogenetic tree among 58 eurosids based on 57 protein-coding genes reconstructed by the Bayesian method. The posteriori probabilities are labeled at nodes. Family, order and higher-level group names are also indicated

organelle genome assembly. The cp genome reported here will enrich and help to resolve the phylogeny of the Rosids subclass. In addition, studies of the *Eugenia uniflora* genome will also allow for discovery and interpretation of functional elements encoded within those sequences, providing a basis for understanding key evolutionary changes that correlate with the high diversification rate of Myrteae tribe.

Acknowledgements We would like to thank Prof. Andreia Turchetto Zolet for providing very helpful suggestions to our manuscript and Steven Clipman for correcting the English.

Funding This study was carried out with the support of FAPERGS and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Information on Electronic Supplementary material

Online Resource 1 List of accession numbers of the chloroplast genome sequences used to filter reads from cp genomes.
Online Resource 2 List of the 57 genes used in the ML and Bayesian phylogenetic analysis.
Online Resource 3 List of plastome sequences of Rosids included in the ML and Bayesian phylogenetic analyses.
Online Resource 4 Phylogenetic tree among 58 eurosids based on 57 protein-coding genes reconstructed by the maximum likelihood method.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Molec Biol 215:403–410. doi:10.1016/S0022-2836(05)80360-2

Asif H, Khan A, Iqbal A, Khan IA, Heinze B, Azim MK (2013) The chloroplast genome sequence of *Syzygium cumini* (L.) and its relationship with other angiosperms. Tree Genet Genomes 9:867–877. doi:10.1007/s11295-013-0604-1

Bayly MJ, Rigault P, Spokevicius A, Ladiges PY, Ades PK, Anderson C, Bossinger G, Merchant A, Udovicic F, Woodrow IE, Tibbits J (2013) Chloroplast genome analysis of Australian eucalypts—*Eucalyptus, Corymbia, Angophora, Allosyncarpia* and *Stockwellia* (Myrtaceae). Molec Phylogen Evol 69:704–716. doi:10.1016/j.ympev.2013.07.006

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucl Acids Res 27:573–580. doi:10.1093/nar/27.2.573

Berger BA, Kriebel R, Spalink D, Sytsma KJ (2016) Divergence times, historical biogeography, and shifts in speciation rates of Myrtales. Molec Phylogen Evol 95:116–136. doi:10.1016/j.ympev.2015.10.001

Biffin E, Lucas EJ, Craven LA, Da Costa IR, Harrington MG, Crisp MD (2010) Evolution of exceptional species richness among lineages of fleshy-fruited Myrtaceae. Ann Bot (Oxford) 106:79–93. doi:10.1093/aob/mcq088

De Almeida DJ, Faria MV, Da Silva PR (2012) Biologia experimental em Pitangueira: uma revisão de cinco décadas de publicações científicas/Experimental biology in pitangueira: a review of five decades of scientific publications. Rev Ambiência 8:159–175. doi:10.5777/ambiencia.2012.01.02rb

Dong W, Liu J, Yu J, Wang L, Zhou S (2012) Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLOS One 7:e35071. doi:10.1371/journal.pone.0035071

Doyle J, Doyle J (1990) Isolation of plant DNA from fresh tissue. Focus (Madison) 12:13–15

Ferreira-Ramos R, Laborda PR, De Oliveira Santos M, Mayor MS, Mestriner MA, De Souza AP, Alzate-Marin AL (2008) Genetic analysis of forest species *Eugenia uniflora* L. through of newly developed SSR markers. Conservation Genet 9:1281–1285. doi:10.1007/s10592-007-9458-0

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. Nucl Acids Res 32(Web Server issue): W273–W279. doi:10.1093/nar/gkh458

Govaerts R, Sobral M, Ashton P, Barrie F, Holst BK, Landrum LL, Matsumoto K, Mazine FF, Lughadha EN, Proenca C, Soares-Silva LH, Wilson PG, Lucas E (2015) World checklist of Myrtaceae. Royal Botanic Gardens, Kew

Ibrahim RIH, Azuma J-I, Sakamoto M (2006) Complete nucleotide sequence of the cotton (*Gossypium barbadense* L.) chloroplast genome with a comparative analysis of sequences among 9 dicot plants. Genes Genet Syst 81:311–321

Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee S-B, Peery R, McNeal JR, Kuehl JV, Boore JL (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci USA 104:19369–19374. doi:10.1073/pnas.0709121104

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucl Acids Res 29:4633–4642. doi:10.1093/nar/29.22.4633

Langmead B (2010) Aligning short sequencing reads with Bowtie. Curr Protoc Bioinf 32:11.7.1–11.7.14. doi:10.1002/0471250953.bi1107s32

Leebens-Mack J, Raubeson LA, Cui L, Kuehl JV, Fourcade MH, Chumley TW, Boore JL, Jansen RK, depamphilis CW (2005) Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. Molec Biol Evol 22:1948–1963. doi:10.1093/molbev/msi191

Leseberg CH, Duvall MR (2009) The complete chloroplast genome of coix lacryma-jobi and a comparative molecular evolutionary analysis of plastomes in cereals. J Molec Evol 69:311–318. doi:10.1007/s00239-009-9275-9

Lim TK (2012) *Eugenia uniflora*. In: Lim TK, Edible Medicinal and Non Medicinal Plants, vol. 3, Fruits. Springer, Netherlands, pp 620–630. doi:10.1007/978-94-007-2534-8_85

Lohse M, Drechsel O, Kahlau S, Bock R (2013) OrganellarGenomeDRAW–a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucl Acids Res 41(Web Server issue):W575–W581. doi:10.1093/nar/gkt289

Lucas EJ, Harris SA, Mazine FF, Belsham SR, Nic Lughadha EM, Telford A, Gasson PE, Chase MW (2007) Suprageneric phylogenetics of Myrteae, the generically richest tribe in Myrtaceae (Myrtales). Taxon 56:1105–1128. doi:10.2307/25065906

Machado LO, Vieira LD, Stefenon VM, Pedrosa OF, De Souza EM, Guerra MP, Nodari RO (2017) Phylogenomic relationship of feijoa (*Acca sellowiana* (O.Berg) Burret) with other Myrtaceae based on complete chloroplast genome sequences. Genetica 145:1–12. doi:10.1007/s10709-017-9954-1

Margis R, Felix D, Caldas JF, Salgueiro F, De Araujo DSD, Breyne P, Van Montagu M, De Oliveira D, Margis-Pinheiro M (2002) Genetic differentiation among three neighboring Brazil-cherry (*Eugenia uniflora* L.) populations within the Brazilian Atlantic rain forest. Biodivers & Conservation 11:149–163. doi:10.1023/A:1014028026273

Mazine FF, Souza VC, Sobral M, Forest F, Lucas E (2014) A preliminary phylogenetic analysis of *Eugenia* (Myrtaceae: Myrteae), with a focus on Neotropical species. Kew Bull 69:1–14. doi:10.1007/s12225-014-9497-x

Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE (2010) Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc Natl Acad Sci USA 107:4623–4628. doi:10.1073/pnas.0907801107

Pennington RT, Lavin M, Oliveira-Filho A (2009) Woody Plant Diversity, Evolution, and Ecology in the Tropics: perspectives from Seasonally Dry Tropical Forests. Annual Rev Ecol Evol Syst 40:437–457. doi:10.1146/annurev.ecolsys.110308.120327

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818. doi:10.1093/bioinformatics/14.9.817

Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecol Evol 16:142–147. doi:10.1016/S0169-5347(00)02097-8

Ravi V, Khurana JP, Tyagi AK, Khurana P (2008) An update on chloroplast genomes. Pl Syst Evol 271:101–122. doi:10.1007/s00606-007-0608-0

Reginato M, Neubig KM, Majure LC, Michelangeli FA (2016) The first complete plastid genomes of Melastomataceae are highly structurally conserved. PeerJ 4:e2715. doi:10.7717/peerj.2715

Rohde W, Gramstat A, Schmitz J, Tacke E, Prufer D (1994) Plant viruses as model systems for the study of non-canonical translation mechanisms in higher plants. J Gen Virol 75:2141–2149. doi:10.1099/0022-1317-75-9-2141

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574. doi:10.1093/bioinformatics/btg180

Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG (2014) From algae to angiosperms - inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol Biol 14:23. doi:10.1186/1471-2148-14-23

Salgueiro F, Felix D, Caldas JF, Margis-Pinheiro M, Margis R (2004) Even population differentiation for maternal and biparental gene markers in *Eugenia uniflora*, a widely distributed species from the Brazilian coastal Atlantic rain forest. Diversity & Distrib 10:201–210. doi:10.1111/j.1366-9516.2004.00078.x

Sasaki T (2003) Identification of RNA editing sites in chloroplast transcripts from the maternal and paternal progenitors of tobacco (*Nicotiana tabacum*): comparative analysis shows the involvement of distinct trans-factors for ndhB editing. Molec Biol Evol 20:1028–1035. doi:10.1093/molbev/msg098

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043–2049

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19:1117–1123. doi:10.1101/gr.089532.108

Spada PDS, De Souza GGN, Bortolini GV, Henriques JAP, Salvador M (2008) Antioxidant, mutagenic, and antimutagenic activity of frozen fruits. J Med Food 11:144–151. doi:10.1089/jmf.2007.598

Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. doi:10.1093/bioinformatics/btu033

Steane DA (2005) Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae). DNA Res 12:215–220. doi:10.1093/dnares/dsi006

Straub SCK, Fishbein M, Livshultz T, Foster Z, Parks M, Weitemier K, Cronn RC, Liston A (2011) Building a model: developing genomic resources for common milkweed (*Asclepias syriaca*) with low coverage genome sequencing. BMC Genomics 12:211. doi:10.1186/1471-2164-12-211

Su H-J, Hogenhout SA, Al-Sadi AM, Kuo C-H (2014) Complete Chloroplast Genome Sequence of Omani Lime (*Citrus aurantiifolia*) and Comparative Analysis within the Rosids. PLOS One 9:e113049. doi:10.1371/journal.pone.0113049

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Molec Biol Evol 24:1596–1599. doi:10.1093/molbev/msm092

Tangphatsornruang S, Sangsrakru D, Chanprasert J, Uthaipaisanwong P, Yoocha T, Jomchai N, Tragoonrung S (2010) The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. DNA Res 17:11–22. doi:10.1093/dnares/dsp025

Thornhill AH, Ho SYW, Külheim C, Crisp MD (2015) Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. Molec Phylogen Evol 93:29–43. doi:10.1016/j.ympev.2015.07.007

Wilson PG, O'Brien MM, Heslewood MM, Quinn CJ (2005) Relationships within Myrtaceae sensu lato based on a *mat*K phylogeny. Pl Syst Evol 251:3–19. doi:10.1007/s00606-004-0162-y

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84:9054–9058. doi:10.1073/pnas.84.24.9054

Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–3255. doi:10.1093/bioinformatics/bth352