

UNIVERSIDADE ESTADUAL PAULISTA “Julio de Mesquita Filho”

INSTITUTO DE BIOCÊNCIAS

CAMPUS BOTUCATU



Helena Sanches Marcon

**Caracterização de Retrotransposons com LTR
no gênero *Eucalyptus***

Tese de Doutorado

2014

Helena Sanches Marcon

**Caracterização de Retrotransposons com LTR
no gênero *Eucalyptus***

Orientador: Dr. Celso Luis Marino
Co-orientador: Dr. Douglas Silva Domingues

Tese de Doutorado apresentada ao Instituto de Biociências, Câmpus de Botucatu, UNESP, para obtenção de Título de Doutora no Programa de Pós-Graduação em Ciências Biológicas (Genética).

2014

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO DE AQUIS. E TRAT. DA INFORMAÇÃO
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CAMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: **ROSEMEIRE APARECIDA VICENTE**

Marcon, Helena Sanches.

Caracterização de Retrotransposons com LTR no gênero *Eucalyptus* / Helena Sanches Marcon. – Botucatu : [s.n.], 2014

Tese (doutorado) - Universidade Estadual Paulista, Instituto de Biociências de Botucatu

Orientador: Celso Luis Marino

Coorientador: Douglas Silva Domingues

Capes: 20200005

1. Retroelementos. 2. Genomas. 3. Eucalipto. 4. Genética vegetal.
5. Espécies.

Palavras-chave: *Copia*; Espécie-específico; *Gypsy*; Híbrido; IRAP; Quantificação; REMAP; RBIP.

Dedico aos meus pais, Claudionor e Maria Inês, pela dedicação, amor e trabalho, sempre me proporcionando a oportunidade de realizar sonhos e conquistas.

Agradecimientos



“Nenhum dever é mais importante do que a gratidão”

Cícero

Os agradecimentos de uma tese são talvez a melhor parte de se ler, visto que são neles que se pode ter a oportunidade de conhecer rapidamente o processo que o pesquisador passou.

Os quatros anos em que desenvolvi esta pesquisa foram de uma árdua jornada de desafio, construção e amadurecimento. Mas nenhum empreendimento é realizado de forma fácil e sem esforço. Aprendi que qualquer trabalho se torna uma extensão da vida do autor. Assim, para se produzir algo de valor a pessoa deve criar algo de valor em si. Os resultados são o reflexo da consistência do ser e da obra desenvolvida. Deste modo agradeço do fundo do coração a todas as pessoas que me encorajaram e me ajudaram a produzir este trabalho.

Á princípio gostaria de agradecer as agências de fomento que auxiliaram no desenvolvimento deste trabalho: CAPES e CNPq. Além de agradecer a pós-graduação em Ciências Biológicas – Genética pela oportunidade de desenvolvimento deste trabalho junto ao programa para que pudesse vir a obter o título de Doutora.

Ao meu orientador, Celso, por expandir meus horizontes proporcionando a livre escolha do tema de meu doutorado, que me fascinou desde meu mestrado, e o desenvolvimento do mesmo em seu laboratório. Além é claro da convivência nestes 10 anos e meio.

Ao meu co-orientador, Douglas, que de grande veterano da Bio e de lab aceitou me co-orientar, proporcionando-me desafios que instigaram meu intelecto e o aprimoraram. Também pela amizade e paciência em ouvir-me nos diferentes momentos que permearam meu doutorado e por acreditar em mim incondicionalmente.

No decorrer deste doutorado recebi o apoio de todos os professores do Departamento de Genética, não somente para o seu desenvolvimento, disponibilizando seus laboratórios e equipamentos, mas também pelo apoio moral. Em particular agradeço a Profa. Claudia, que em uma relação mais próxima sempre me apoiou e incentivou e ao Prof. Ivan por estar sempre disponível quando solicitado.

Ao longo desta pesquisa também recebi apoio do Prof. Ney, do departamento de Física e Biofísica, do Prof. João Pessoa do departamento de Microbiologia e Imunologia, do Prof. Claudio e de sua técnica Luciana do Departamento de Morfologia e do Prof. Paulo Ribolla do Departamento de Parasitologia, a todos o meu muito obrigada!

Aos funcionários da Genética presentes e ausentes, Aline, Zé, Betina, Rosana, Val e Deisi, obrigada pela ajuda nos momentos que precisei. E também aos funcionários da Pós-graduação, que sempre se dispuseram a ajudar.

Aos colegas do Instituto Agronômico do Paraná que sempre estiveram disponíveis para ajudar-me quando solicitados, Kenia, Thiago e Juliana.

Agradeço também aos colegas de todo o Departamento de Genética, que já saíram e aos que ainda estão por aqui, em especial a Dani e a Magali. E a todos os colegas da pós-graduação, que pude conviver e dividir momentos tão agradáveis durante a organização do Workshop de Genética e mesmo nas disciplinas cursadas.

Meu especial agradecimento aos colegas do laboratório de Genética de Plantas presentes e ausentes pelos dias passados no laboratório, finais de semana, feriados e tanto outros momentos compartilhados: Dra. Virginia, Dra. Juliana, Dr. Bonsai, Dra. Tânia, Suzana, Cíntia, Dr. Júlio, Fábio, Dra. Vanusa, Alessandra Morena, Alessandra Loira, Rodrigo, Paola, Aletéa, Karine, Rodrigo, Márcio, Eder, Cris, Gicela, Marcela, Pit, Sarah.

E em especial agradeço a Sarah, minha estagiária, pela ajuda, paciência com meu jeito metódico e pelos momentos de cumplicidade, vai morar para sempre no meu coração.

Agradecer a Jú e a Alessandra, pois além de serem ótimas colegas de trabalho também foram ombro amigo nas caminhadas no Lageado, nas noites de ajuda e nos momentos de descontração.

Também deixar meu infinito obrigada a Vir, minha parceira de trabalho no mestrado, que hoje carrego no coração como se fosse minha irmã mais velha, era só lembrá-la que aparecia no laboratório, “transmissão de pensamento Vir”, e podia receber o abraço amigo mais sincero.

E meu carinho especial pela Suzana e Cíntia, amigas que no começo acompanharam de pertinho, na cadeira do lado, mas com a saída de ambas mesmo de longe

me deram o estímulo verdadeiro e a força para não desanimar e puderam partilhar de felizes momentos.

Além de agradecer aos meus alunos do curso de Ciências Biológicas da faculdade de EDUVALE de Avaré pelo grande aprendizado que estão me proporcionando junto as disciplinas de Genética e Biologia Molecular. Também aos colegas da EDUVALE por terem a paciência de ouvir falar tanto do meu doutorado. E a minha coordenadora Lígia, pela oportunidade, compreensão, apoio e incentivo.

Aos grandes amigos, que estive distante devido as exigências do doutorado, meu muito obrigada pela torcida, apoio, paciência, compreensão e pelo abraço amigo: Aline, Bruno, Janaina, Priscila, Michele, Shelly.

Ao meu amor, William, foram momentos de altos e baixos que marcaram meu doutorado e nosso relacionamento sobreviveu pela força que nos une e fez o nosso amor maior e mais puro. Obrigada pela sua força, por me ouvir, por esperar os meus atrasos frente as inúmeras análises diante do computador, por respeitar-me quando cansada dormia e não assistia ao filme, pela sua compreensão e por acreditar em mim não deixando pairar dúvidas de minha capacidade em nenhum minuto frente aos obstáculos que surgiram, enfim obrigada pela sua presença em minha vida.

Aos meus familiares, tios e primos, que me apoiaram em um dos momentos mais difíceis da minha vida durante este doutorado levando força e apoio a mim e a minha família.

Ao meu irmão, Marcelo - como amadureceu nestes quatro anos! Nossas diferenças ficaram mais evidentes mostrando o quanto nos foi dado a oportunidade de traçar caminhos independentes buscando sempre a nossa felicidade. Desculpa não estar ao seu lado nos momentos que mais precisou e ao mesmo tempo obrigada por não me cobrar a ausência e permitir que continuasse vivendo o caminho que escolhi, irmão melhor não poderia ter sido colocado a meu lado nesta vida.

E finalmente posso agradecer os maiores responsáveis por eu poder estar chegando a este momento, meus pais.

Ao meu Pai, você é um grande homem, com uma grandeza de amor sem igual, queria agradecer pelo seu apoio incondicional, por me acalantar durante um dos períodos mais difíceis que já vivemos me dando força para continuar e compreendendo minha ausência

estando ao lado de minha mãe para que ela também tivesse forças para continuar. Você sempre será um exemplo de dignidade e perseverança. Devo meu doutorado a você, a este amor que tem por nossa família. Te amarei eternamente!

A minha Mãe Inês, obrigada pela sabedoria em me educar, pelo carinho de mãe, por me ensinar os limites da vida, por ter investido e acreditado em mim incentivando-me a trilhar caminhos do conhecimento. Obrigada por ensinar-me a não desistir dos meus sonhos, por acreditar em mim e compartilhar das minhas conquistas e angústias vividas. Sua força em acreditar na vida para poder estar conosco será sempre um exemplo. Agradeço todos os dias por você ser minha mãe. Amo-te muito e sempre!

“Ser herói não significa acertar constantemente. É muito mais que isso. O verdadeiro espírito de um herói encontra-se na intensa convicção de enfrentar e vencer as dificuldades em vez de desistir de tudo. Na vida de todos nós poderão surgir situações inesperadas. Poderão se manifestar obstáculos ou problemas que jamais havíamos imaginado. É justamente nesses momentos que revelamos o que verdadeiramente carregamos no nosso coração.”

Daisaku Ikeda

“A vida só é possível através dos desafios, quando você tem tanto o bom tempo quanto o mau tempo, quando tem prazer e dor; quando tem inverno e verão, dia e noite; quando tem tristeza tanto quanto felicidade, desconforto tanto quanto conforto. A vida passa entre essas duas polaridades. Movendo-se entre essas duas polaridades, você aprende a se equilibrar. Entre essas duas asas, você aprende a voar até a estrela mais brilhante.”

Osho

Resumo



Marcon, H. S. Caracterização de Retrotransposons com LTR no gênero *Eucalyptus*. 2014. 171 f. Tese (Doutorado) – Instituto de Biociências de Botucatu, Universidade Estadual Paulista, 2014.

Os retrotransposons com LTR (*long terminal repeats*) (LTR-RTEs) são os componentes mais abundantes nos genomas vegetais. Eles são divididos em duas superfamílias, *Copia* e *Gypsy*, de acordo com a ordem dos domínios internos e com a similaridade das sequências. São conhecidos pela sua natureza ubíqua e plasticidade, o que os leva a serem explorados como ferramentas para o desenvolvimento de marcadores moleculares, utilizando-se assim o padrão de inserção dos LTR-RTEs para geração de polimorfismos. As famílias de LTR-RTEs apresentam número de cópias variável dentro de um mesmo genoma, mas a amplificação de poucas famílias pode contribuir com grandes diferenças de tamanho entre genomas próximos. Em plantas do gênero *Eucalyptus* existem poucos estudos relacionados aos LTR-RTEs, que geralmente limitam-se a análises em bancos de dados privados. Este trabalho tem como objetivo a caracterização de LTR-RTEs transcricionalmente ativos pertencentes as superfamílias *Copia* e *Gypsy*, encontrados no genoma de eucalipto. A partir de análises preliminares de bioinformática, um total de nove famílias de LTR-RTEs foram identificadas; os elementos da superfamília *Copia* foram classificadas em cinco linhagens e elementos da superfamília *Gypsy* foram divididos em três linhagens. As análises *in silico* no genoma *draft* de *Eucalyptus grandis* identificaram que as famílias caracterizadas possuem entre 38 e 290 cópias. Os LTR-RTEs foram utilizados para o desenvolvimento de marcadores moleculares IRAP, REMAP e RBIP, que foram avaliados em cinco espécies de eucalipto. Dezesesseis *primers* IRAP foram avaliados e um total de 1521 fragmentos foram gerados, com 321 fragmentos polimórficos, em um intervalo de 250 a 2500 bp e uma média de 19,01 fragmentos amplificados por *primer*. Vinte e oito combinações de *primers* REMAP foram avaliadas e geraram 1910 fragmentos, dos quais 492 são polimórficos, em um intervalo de 200 a 1250 bp, com uma média de 13,64 fragmentos por combinação de *primer*. O padrão de distância obtido pela análise destes marcadores não foi o mesmo observado em estudos de sistemática. Ainda assim, estes marcadores moleculares podem ser utilizados para análises de diversidade populacional e individual. Os marcadores RBIP foram gerados a partir de três polimorfismos sítio-específicos em três famílias de LTR-RTEs da superfamília *Copia*. O marcador desenvolvido a partir da família RLC_egMax_1 possibilitou identificar locus codominantes entre diferentes espécies de eucalipto. Também foram avaliados o número

relativo de cópias de seis famílias de LTR-RTEs em duas espécies de eucalipto (*E. grandis* e *E. urophylla*) e do híbrido F₁ gerado a partir do cruzamento destas espécies, através de análises quantitativas por qPCR. Foi avaliada a razão entre a LTR/domínio interno, e os dados observados sugerem que parte das famílias apresentam um grande número de cópias completas. O híbrido F₁ apresentou mais cópias completas se comparado aos genitores em três famílias de LTR-RTEs, indicando uma ativação destas famílias devido o evento de hibridação. Foi também possível identificar uma variação na atividade transcricional de cada família de LTR-RTE. A atividade transcricional de algumas famílias foi maior em caule e raiz quando comparadas as folhas. A diferença da atividade transcricional entre os LTR-RTEs entre os genitores e no híbrido sugere que o controle transcricional não possui uma relação com a atividade de transposição em F₁. Com isso, este trabalho tem importante papel na compreensão da dinâmica de LTR-RTEs em espécies de eucalipto.

Palavras-chave: *Copia*, *Gypsy*, IRAP, REMAP, RBIP, híbrido, espécie-específico, quantificação.

Abstract



Marcon, H. S. **Characterization of LTR Retrotransposon in *Eucalyptus* genus**. 2014. 171f. Thesis (PhD) – Biosciences Institute of Botucatu, Universidade Estadual Paulista, 2014.

Long Terminal Repeat Retrotransposons (LTR-RTEs) are the most abundant component of plant genomes. They are divided into two superfamilies, *Copia* and *Gypsy*, based on coding domain order and sequence similarity. As dynamic and ubiquitous entities in plant genomes, marker systems based in LTR-RTEs were developed to exploit LTR-RTEs insertion pattern polymorphism. In most angiosperm genomes, LTR-RTEs families have a wide range of copy number, but the amplification of a few families individually contribute to a large fraction of the genome. LTR-RTEs were scarcely studied in the *Eucalyptus* genus, with most findings derived from global analyses of private genomic databases. The aim of this work is to characterize transcriptionally active *Copia* and *Gypsy* LTR-RTEs in *Eucalyptus* sp genomes. After preliminary bioinformatics analyses, a total of 9 full-length LTR-RTEs families were classified into five *Copia* and three *Gypsy* lineages. BLAT analysis in *Eucalyptus grandis* draft genome identified that these elements have between 38 and 290 copies. These LTR-RTEs were used to develop IRAP, REMAP and RBIP markers, which were evaluated in five *Eucalyptus* species (*Eucalyptus brassiana*, *E. grandis*, *Eucalyptus saligna*, *Eucalyptus tereticornis* and *Eucalyptus urophylla*). Sixteen IRAP primers produced 321 polymorphic fragments from a total of 1521 bands, ranging from 250 to 2500 bp, with an average of 19.01 bands amplified per IRAP primer. Twenty-eight REMAP primer combinations produced 1910 bands. 492 of them were polymorphic, ranging from 200 to 1250 bp. An average of 13.64 REMAP bands was scored per primer. Our results showed that the genetic similarity assessed by both IRAP and REMAP markers did not reflect molecular phylogenetic analysis. Nonetheless, these markers are useful for the assessment of genetic diversity the individual level and population analysis. We developed RBIP markers for three site-specific polymorphisms. Three *Copia*-type families were as candidates for the development of DNA marker systems. A candidate marker from family RLC_egMax_1 was particularly capable of identifying codominant locus among different *Eucalyptus* species. This study also assessed the relative copy number of six LTR-RTEs families in two *Eucalyptus* species (*E. grandis* and *E. urophylla*) and their F₁ hybrids using quantitative real-time PCR. Differences in LTR/internal domain ratio suggest that some elements have a higher number of complete copies. The F₁ hybrid have more copies than parents in three families, indicating the activation of these families during the hybridization event. The transcriptional pattern of each

LTR-RTE family is variable, even within the same class of elements. Transcriptional activity of several LTR-RTEs is higher in stalks and roots, comparing to leaves. The difference of transcriptional activity between LTR-RTEs in the parentals and the hybrid suggests that transcriptional control was not related to their transpositional activity in the F₁. This study makes an important overview of LTR-RTEs dynamism among *Eucalyptus* species.

Key words: *Copia*, *Gypsy*, IRAP, REMAP, RBIP, hybrids, species-specific, quantification.

Lista de Figuras

- Figura 1:** Variação de fenótipos com relação à coloração em grãos de milho devido a ação de elementos transponíveis. A inconstância de fenótipos ocorre pela inserção e excisão do elemento em gene responsável pela produção da enzima que confere coloração ao aleurona da semente do milho. Setores revertentes (com pigmento) ocorrem com a excisão do elemento. O tamanho do setor reflete o momento durante o desenvolvimento onde a excisão aconteceu. Adaptado de Feschotte et al. (2002)..... 32
- Figura 2:** Classificação sugerida para elementos transponíveis. Traduzido de Wicker et al, (2007)..... 34
- Figura 3:** Organização dos LTR-RTEs. Os elementos podem ser classificados dentro de dois maiores grupos, as superfamílias *Copia* e *Gypsy*, nomeados após serem descobertos em *D. melanogaster* e levedura. Os elementos são flanqueados por 5 bp repetitivos (setas tracejadas). Os RTEs consistem em dois LTRs que contêm pequenas repetições invertidas nas extremidades (triângulos pretos) e um conjunto interno de poliproteínas responsáveis pela síntese, que contêm o domínio *gag*, que codifica a proteína formadora do capsídeo viral, similar ao de retrovírus, e uma protease aspártica (AP), no outro domínio estão as regiões responsáveis pela clivagem de poliproteínas dentro das unidades funcionais, a integrase (IN), responsável pela inserção de cópias dentro do genoma, e uma transcriptase reversa (RT) e uma RNaseH (RH), que sintetiza o cDNA a partir do RNA transcrito. Os LTR-RTEs da superfamília *Gypsy* são diferentes de *Copia* em relação a posição no domínio IN. Muitos LTR-RTEs da superfamília *Gypsy* contêm, como no retrovírus, um domínio *env*, geralmente expresso em uma posição diferenciada. Adaptado de Schulman et al. (2012)..... 36
- Figura 4:** Mecanismo de Transposição nos LTR-RTEs: (1) Presença do RNA intermediário transcrito (marrom) de um LTR-RTE com o pareamento de um tRNA (azul), o iniciador está posicionado próximo a extremidade 5"; (2) isto serve como um iniciador para síntese de DNA pela transcriptase reversa; (3) a primeira sequência é copiada uma única vez a partir da porção 5" final do RNA (U5 – vermelho) e por uma pequena sequência repetitiva (R – verde) presente em ambas extremidades do RNA. A transcrição reversa pode continuar porque a cópia de R em um novo DNA sintetizado foi pareado com a sequência complementar a de outra porção do DNA; (4) isto permite que a síntese de DNA prossiga; (5) começando pela sequência única U3 (lilás) adjacente a R, e continua até U5 e R serem copiados novamente. A fita molde de RNA é degradada pela RNaseH; (6) um fragmento de polipurina é deixado para formação da fita de DNA complementar; (7) transcriptase reversa dá início a transcrição reversa a partir da complementariedade entre R e U5; (8) síntese procede em ambas as direções para formar uma dupla fita de DNA e LTR, feita a partir da U5, R e U3, em cada extremidade; (9) integrase age inserindo no DNA cromossomal. Modificado a partir de Finnegan (2012)..... 37

Figura 5:	Tamanho do genoma e conteúdo de TE em diferentes espécies de angiospermas, que apresentam-se correlacionados. Dados retirados de Tenaillon et al. (2010).....	40
Figura 6:	Métodos para construção de marcadores baseado em LTR-RTEs adaptado de Kalendar e Schulman (2006). (a) IRAP: Amplificação entre as LTRs de dois RTEs (setas vermelhas). DNA genômico (gDNA), identificado pela linha azul escuro, os <i>primers</i> estão indicados pelas setas vermelhas, o LTR-RTEs é representado pelas LTRs e pelo domínio central (<i>core</i>). Outras características também podem ser observadas no genoma, como: microsátélites (SSRs) e os sítios de restrição (R), porém IRAP não as utilizam. (b) REMAP: Amplificação dos <i>primers</i> entre um LTR (seta vermelha) e um SSRs (seta azul). (c) RBIP: Todas as posições (contendo um LTR-RTEs) apresentam geralmente um produto de amplificação com um <i>primer</i> LTR (seta vermelha) e um <i>primer</i> que flanqueia uma região adjacente a este (seta azul).....	42
Figura 7:	Etapas realizadas durante a identificação <i>in silico</i> de LTR-RTEs.....	50
Figura 8:	Representação do posicionamento dos <i>primers</i> para: (A) IRAP, cujas setas em rosa setas indicam a posição que os mesmos ocupam flanqueando dois LTR-RTEs adjacentes; (B) REMAP, cuja seta aponta loca de pareamento dos <i>primers</i> (laranja: microsátélite; rosa: LTR), que flanqueiam a região entre o LTR-RTE e o microsátélite (SSRs) adjacente. gDNA: DNA genômico.....	56
Figura 9:	Representação do posicionamento dos <i>marcadores</i> RBIP obtidos a partir das técnicas: (A) IRAP, setas em azul indicam região entre LTRs (LTR-RTEs adjacentes), que os marcadores flanqueiam; e (B) seta em azul e rosa indicam região flanqueada por marcadores obtidos por REMAP.....	58
Figura 10:	Representação do posicionamento dos <i>primers</i> para qPCR na LTR (setas em verde); regiões internas, <i>gag</i> (seta amarela) ou transcriptase reversa (Rvt) (setas em azul). gDNA: DNA genômico; GAG e Pol: regiões internas a ORF.....	66
Figura 11:	Etapas realizadas durante a identificação <i>in silico</i> , com o número de sequência identificadas em cada passo.....	78
Figura 12:	Estrutura hipotética de LTR-RTEs da superfamília <i>Copia</i> identificados no presente trabalho.....	81
Figura 13:	Estrutura hipotética de LTR-RTEs da superfamília <i>Gypsy</i> identificados no presente trabalho.....	82
Figura 14:	Comparação entre as sequências PBS das linhagens identificadas no trabalho com as respectivas sequências consenso das linhagens correspondentes (extraído de Wicker e Keller 2007).....	83
Figura 15:	Distribuição genômica dos LTR-RTEs pertencentes a superfamília <i>Copia</i> ao longo dos 11 pseudocromossomos de <i>E. grandis</i>	85
Figura 16:	Distribuição genômica dos LTR-RTEs pertencentes a superfamília <i>Gypsy</i> ao longo dos 11 pseudocromossomos de <i>E. grandis</i>	86
Figura 17:	Relações filogenéticas a partir de sequências conservadas do domínio transcriptase reversa de LTR-RTEs pertencentes a superfamília <i>Copia</i> . A árvore foi enraizada utilizando-se o elemento <i>DM</i> de <i>Drosophila melanogaster</i> depositada no banco de dados RepBase como grupo externo. As diferentes cores nos ramos indicam as linhagens	

	evolutivas, verde escuro - <i>Maximus</i> , rosa - <i>Ale</i> , verde claro - <i>Tar</i> , roxo - <i>Angela</i> , laranja - <i>GMR</i> , azul - <i>Bianca</i> e vermelho - <i>Ivana</i>	88
Figura 18:	Relações filogenéticas a partir de sequências conservadas do domínio transcriptase reversa de LTR-RTEs pertencentes a superfamília <i>Gypsy</i> de diferentes linhagens evolutivas descritos por Du et al. (2010) e apresentadas por Neumann et al (2011), com adição dos acessos XM_004497754.1 e ERP54440.1. A árvore foi enraizada utilizando-se o elemento <i>Invader</i> de <i>Drosophila melanogaster</i> , depositada no banco de dados RepBase, como grupo externo. As diferentes cores dos ramos indicam as linhagens evolutivas: azul - <i>Tekay</i> , vermelho - <i>Galadriel</i> , verde claro - <i>Reina</i> , roxo - <i>CRM</i> , verde escuro - <i>Caulivírus</i> , laranja <i>Athila</i> e rosa - <i>Tat</i>	90
Figura 19:	Análise IRAP em gel de agarose 1,8%: (A) utilizando o <i>primer</i> EgAleLTR-3; (B) utilizando <i>primer</i> EgBiancaLTR-3. L: GeneRuler 1kb DNA Ladder (Fermentas – código SM0311).....	94
Figura 20:	Análise IRAP em Gel de agarose 1,8%: (A) utilizando o <i>primer</i> EgAle2LTR-3; (B) utilizando <i>primer</i> EgAle2LTR-5. L: GeneRuler 1kb DNA Ladder (Fermentas – código SM0311). As setas (amarelas) indicam locais em que foram observados polimorfismos.....	95
Figura 21:	Análise IRAP em gel de agarose 1,8%: (A) utilizando o <i>primer</i> EgGMRLTR-5; (B) utilizando <i>primer</i> EgMaxLTR-3. L: GeneRuler 1kb DNA Ladder (Fermentas – código SM0311).....	96
Figura 22:	Análise IRAP em gel de agarose 1,8%: (A) utilizando o <i>primer</i> EgTatLTR-3; (B) utilizando <i>primer</i> EgTekayLTR-3. L: GeneRuler 1kb DNA Ladder (Fermentas – código SM0311).....	97
Figura 23:	Análises REMAP em gel de agarose 1,8% corado com brometo de etídio. A: Combinação dos <i>primers</i> Micro1 e EgAle2LTR-3. B: Combinação dos <i>primers</i> Micro1 e EgTatLTR-5. L: GeneRuler 1kb DNA Ladder (Fermentas – código SM0311).....	101
Figura 24:	Análises REMAP em gel de agarose 1,8% corado com brometo de etídio. A: Combinação dos <i>primers</i> Micro1 e EgCauliLTR-5, com indivíduos <i>E. brassiana</i> , <i>E. grandis</i> e <i>E. saligna</i> ; B: Combinação dos <i>primers</i> Micro1 e EgCauliLTR-3, com indivíduos <i>E. brassiana</i> e <i>E. grandis</i> . L: GeneRuler 1kb DNA Ladder (Fermentas – código SM0311).....	101
Figura 25:	Análises REMAP em gel de agarose 1,8% corado com brometo de etídio. A: Combinação dos <i>primers</i> Micro1 e EgAleLTR-3. B: Combinação dos <i>primers</i> Micro1 e EgBiancaLTR-3. L: GeneRuler 1kb DNA Ladder (Fermentas – código SM0311).....	102
Figura 26:	Análises REMAP em gel de agarose 1,8% corado com brometo de etídio. A: Combinação dos <i>primers</i> Micro12 e EgMaxLTR-5. B: Combinação dos <i>primers</i> Micro12 e EgTatLTR-5. L: GeneRuler 1kb DNA Ladder (Fermentas – código SM0311).....	103
Figura 27:	Dendrograma UPGMA de distância genética com os dados obtidos por meio das técnicas IRAP e REMAP.....	107
Figura 28:	Região polimórfica identificada a partir do <i>primer</i> CoEu4-3 da amostra de <i>E. grandis</i> no genoma <i>draft</i> de <i>E. grandis</i>	108
Figura 29:	Detecção da família RLC_egGMR_1. A: Ilustração do <i>primer</i> RBIP desenhado. Setas vermelhas e azul indicam a posição e direção dos	

	<i>primers</i> usados. TGMR1_F X TGMR1_R: conjunto de <i>primers</i> que flanqueiam a região genômica a esquerda (F - <i>forward</i>) e a direita (R - <i>reverse</i>). B: Presença ou ausência de RLC_egGMR_1 nas cinco espécies de eucalipto. L: GeneRuler 1kb DNA Ladder (Fermentas - código SM0311).....	110
Figura 30:	Detecção da família RLC_egAle2_1. A: Ilustração do <i>primer</i> RBIP desenhado. Setas rosas indicam a posição e direção dos <i>primers</i> usados. EAle2.1_F X EAle2.1_R: conjunto de <i>primers</i> que flanqueiam a região genômica a esquerda (F) e a direita (R). B: Presença ou ausência de RLC_egAle2_1 nas cinco espécies de eucalipto. L: GeneRuler 1kb DNA Ladder (Fermentas - código SM0311).....	111
Figura 31:	Detecção da família RLC_egMax_1 em cinco espécies de eucalipto com tamanhos de fragmentos diferentes indicados (setas laranja). L: GeneRuler 1kb DNA Ladder (Fermentas - código SM0311).....	111
Figura 32:	Eficiência de amplificação do qPCR feito para os pares de <i>primers</i> utilizados na quantificação de LTR-RTEs. Um conjunto de quatro a seis diluições foram usadas, cujos valores de C_T no eixo Y indicam o valor do <i>threshold</i> , enquanto que o eixo X indica a quantidade de DNA. Em todos os casos a regressão (R_2) foi acima de 0,99.....	113
Figura 33:	Análises de qPCR do RLC_egAle_2 nos parentais <i>E. grandis</i> e <i>E. urophylla</i> e no híbrido (F_1). Os resultados de <i>E. grandis</i> foram utilizados como calibrador. (A) RQ da região interna (Gag) do RLC_egAle_2. (B) RQ da LTR do RLC_egAle_2. (C) Razão LTR:Gag do RLC_egAle_2. Nas barras estão indicados o erro padrão (n=9).....	115
Figura 34:	Análises de qPCR do RTE RLC_egMax_1 nos parentais <i>E. grandis</i> e <i>E. urophylla</i> e no híbrido F_1 . Os resultados de <i>E. grandis</i> foram utilizados como calibrador. (A) RQ da região interna (Gag) do RTE RLC_egMax_1. (B) RQ da LTR do RTE RLC_egMax_1. (C) LTR:Gag do RTE RLC_egMax_1. Nas barras estão indicados o erro padrão (n=9).....	117
Figura 35:	Análises de qPCR do RLC_egGMR_1 nos parentais <i>E. grandis</i> e <i>E. urophylla</i> e no híbrido (F_1). Os resultados de <i>E. grandis</i> foram utilizados como calibrador. (A) RQ da região interna (Rvt) do RLC_egGMR_1. (B) RQ da LTR do RLC_egGMR_1. (C) LTR:Rvt do RLC_egGMR_1. Nas barras estão indicados o erro padrão (n=9).....	118
Figura 36:	Análises de qPCR do RLC_egIvana_1 nos parentais <i>E. grandis</i> e <i>E. urophylla</i> e no híbrido (F_1). Os resultados de <i>E. grandis</i> foram utilizados como calibrador. (A) RQ da região interna (Gag) do RLC_egIvana_1. (B) RQ da LTR do RLC_egIvana_1. (C) LTR:Gag do RLC_egIvana_1. Nas barras estão indicados o erro padrão (n=9).....	119
Figura 37:	Análises de qPCR do RLG_egTat_1 nos parentais <i>E. grandis</i> e <i>E. urophylla</i> e no híbrido (F_1). Os resultados de <i>E. grandis</i> foram utilizados como calibrador. (A) RQ da região interna (Gag) do RLG_egTat_1. (B) RQ da LTR do RLG_egTat_1. (C) LTR:Gag do RLG_egTat_1. Nas barras estão indicados o erro padrão (n=9).....	121
Figura 38:	Análises de qPCR do RLG_egTekay_1 nos parentais <i>E. grandis</i> e <i>E. urophylla</i> e no híbrido (F_1). Os resultados de <i>E. grandis</i> foram utilizados como calibrador. (A) RQ da região interna (Gag) do RLG_egTekay_1. (B) RQ da LTR do RLG_egTekay_1. (C) LTR:Gag do RLG_egTekay_1. Nas barras estão indicados o erro padrão (n=9)....	122

Figura 39:	Expressão relativa da família RLC_egAle_1 em diferentes tecidos de cinco espécies de eucalipto e o híbrido F ₁ . Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, ANOVA seguido de teste LSD).....	123
Figura 40:	Expressão relativa da família RLC_egAle_2 em diferentes tecidos de cinco espécies de eucalipto e o híbrido F ₁ . Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, ANOVA seguido de teste LSD).....	124
Figura 41:	Expressão relativa da família RLC_egMax_1 em diferentes tecidos de cinco espécies de eucalipto e o híbrido F ₁ . Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, ANOVA seguido de teste LSD).....	125
Figura 42:	Expressão relativa da família RLC_egBianca_1 em diferentes tecidos de cinco espécies de eucalipto e o híbrido F ₁ . Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, ANOVA seguido de teste LSD).....	126
Figura 43:	Expressão relativa da família RLC_egGMR_1 em diferentes tecidos de cinco espécies de eucalipto e o híbrido F ₁ . Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, ANOVA seguido de teste LSD).....	127
Figura 44:	Expressão relativa da família RLC_egIvana_1 em diferentes tecidos de cinco espécies de eucalipto e o híbrido F ₁ . Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, ANOVA seguido de teste LSD).....	128
Figura 45:	Expressão relativa da família RLC_egTat_1 em diferentes tecidos de cinco espécies de eucalipto e o híbrido F ₁ . Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, ANOVA seguido de teste LSD).....	129
Figura 46:	Expressão relativa da família RLC_egTekay_1 por meio de RT-qPCR em diferentes tecidos de cinco espécies de eucalipto e o híbrido F ₁ . Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, ANOVA seguido do teste LSD).....	129
Figura 47:	Expressão relativa da família RLC_egCauli_1 em diferentes tecidos de cinco espécies de eucalipto e o híbrido F ₁ . Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, ANOVA seguido de teste LSD).....	130
Figura 48:	Expressão relativa das famílias de LTR-RTEs por meio de RT-qPCR em plantas de <i>E. grandis</i> estressadas osmoticamente com PEG após 6 e 24h. Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, teste ANOVA seguido de teste LSD).....	131

Lista de Tabelas

Tabela 1:	Material vegetal utilizado para os diferentes experimentos do trabalho.....	49
Tabela 2:	<i>Primers</i> microssatélites utilizados conjuntamente com <i>primers</i> IRAP, na técnica REMAP (Kalendar et al. 1999).....	57
Tabela 3:	<i>Primers</i> utilizados para a qPCR em que se obteve os valores de massa de uma copia do plasmídeo e seus respectivos valores de eficiência e R ²	69
Tabela 4:	Concentração de cDNA dos diferente tecidos usados para o desenvolvimento do RT-qPCR.....	74
Tabela 5:	LTR-RTEs pertencentes a superfamília <i>Copia</i> identificados a partir de LTR-RTEs de famílias já descritas na literatura e armazenadas no banco de dados RepBase, comparados ao genoma draft de <i>E. grandis</i>	80
Tabela 6:	LTRs-RTEs pertencente a superfamília <i>Gypsy</i> identificados a partir de LTRs-RTEs de famílias já descritas e armazenados no banco de dados RepBase, comparados ao genoma draft de <i>E. grandis</i>	81
Tabela 7:	Caracterização das sequências de referência de LTR-RTEs da superfamília <i>Copia</i> utilizadas para análises de bancada.....	85
Tabela 8:	Caracterização das sequências de LTR-RTEs da superfamília <i>Gypsy</i> utilizadas para análises de bancada.....	85
Tabela 9:	Proporção a cada 100.000 ESTs Sanger de cada LTR-RTE nas diferentes espécies de eucalipto disponíveis no banco de dados EUCANEXT (Nascimento et al. 2011; Salazar et al. 2013). Foram selecionados ESTs que atendiam as regras de Wicker et al. (2007).....	93
Tabela 10:	Total de fragmentos e polimorfismos obtidos com <i>primers</i> IRAP.....	99
Tabela 11:	Fragmentos polimórficos selecionados entre os <i>primers</i> desenhados nos LTR-RTEs das superfamílias <i>Copia</i> e <i>Gypsy</i>	101
Tabela 12:	Total de fragmentos e polimorfismos obtidos com os <i>primer</i> IRAP em conjunto com os <i>primers</i> Micro 1 e 12.....	106
Tabela 13:	Fragmentos polimórficos selecionados a partir da combinação dos <i>primers</i> Micro1 e 12 com os <i>primers</i> IRAP.....	107
Tabela 14:	Marcadores obtidos a partir dos polimorfismos encontrados através das técnicas IRAP e REMAP.....	110

Sumário

Resumo	
Abstract	
Lista de Figuras	
Lista de Tabelas	
1. Introdução	26
1.1. O Gênero <i>Eucalyptus</i>	27
1.1.1. A cultura do eucalipto no Brasil	28
1.1.2. Estudos moleculares no gênero <i>Eucalyptus</i>	29
1.2. Elementos Transponíveis: aspectos históricos	31
1.2.1. Classificação e estrutura dos Elementos Transponíveis	33
1.2.2. Retrotransposons com LTR e sua distribuição	35
1.2.3. O número de cópias de retrotransposons com LTR no genoma vegetal	39
1.2.4. Os retrotransposons com LTR como marcadores moleculares	41
1.2.5. Atividade transcricional de Retrotransposons com LTR em plantas	44
2. Objetivos	46
2.1. Objetivo Geral	47
2.2. Objetivos Específicos	47
3. Material e Métodos	48
3.1. Material Vegetal	49
3.2. Identificação <i>in silico</i> de LTR-RTEs transcionalmente ativos	50
3.2.1. Caracterização inicial de LTR-RTEs transcionalmente ativos	51
3.3. Identificação de ORFs nos LTR-RTEs candidatos	51
3.3.1. Identificação de Domínios conservados nos LTR-RTEs selecionados	52
3.4. Quantificação e distribuição genômica <i>in silico</i> dos LTR-RTEs selecionados no genoma <i>draft</i> de <i>E. grandis</i>	52
3.5. Análises filogenéticas	53
3.6. Identificação <i>in silico</i> de sequências expressas de LTR-RTEs em espécies do gênero <i>Eucalyptus</i>	53
3.7. Extração de DNA genômico	54
3.7.1. Extração de DNA genômico para análises IRAP, REMAP e RBIP	54
3.7.2. Extração do DNA genômico para quantificação do número de cópias de LTR-RTEs	55

3.7.3. Quantificação de DNA	55
3.8. Manipulação do DNA genômico	55
3.8.1. Iniciadores	55
3.8.1.1. Desenhos de <i>primers</i> IRAP e REMAP	55
3.8.1.2. Desenhos dos marcadores RBIP	57
3.8.2. Amplificação IRAP	58
3.8.3. Amplificação REMAP	59
3.8.4. Análises de fragmentos IRAP e REMAP	59
3.8.4.1. Análise de similaridade genética	59
3.8.5. Purificação de fragmentos polimórficos em IRAP e REMAP	60
3.9. Plasmídeos e Bactérias utilizados	60
3.9.1. Plasmídeos	60
3.9.2. Bactéria Competente	60
3.9.2.1. Obtenção das Bactérias Competentes	61
3.9.2.2. Transformação de bactérias competentes	61
3.9.2.2.1. Ligação de fragmento de DNA	61
3.9.2.2.2. Transformação das bactérias competentes por choque térmico	62
3.9.3. Extração de DNA plasmidial em pequena escala (Miniprep)	62
3.9.4. Análise das Sequências	63
3.10. Amplificação RBIP	64
3.10.1. Seleção de fragmentos para a técnica RBIP	64
3.10.2. Padronização da técnica RBIP	65
3.10.3. Análises de fragmentos RBIP	65
3.11. Quantificação de LTR-RTEs por PCR Quantitativo (qPCR)	65
3.11.1. Escolha do normalizador	65
3.11.2. Desenho dos <i>primers</i> para qPCR	66
3.11.3. PCR quantitativo	66
3.11.3.1. Obtenção de DNA plasmidial de fragmentos utilizados para construção da curva padrão	67
3.11.3.1.1. Análise das Sequências de DNA plasmidial para construção da curva padrão	68
3.11.4. Obtenção da curva padrão para quantificação e cálculo de eficiência	68
3.11.5. Análises quantitativas de LTR-RTEs selecionados	69
3.12. Análise transcricional	70
3.12.1. Condições experimentais	70

3.12.2. Extração de RNA total	71
3.12.2.1. Quantificação do RNA total	72
3.12.3. Desenho dos <i>primers</i> para o RT-qPCR	72
3.12.4. Síntese da primeira fita de cDNA	72
3.12.5. PCR em Tempo Real (RT-qPCR)	73
3.12.6. Cálculo de expressão relativa	75
3.12.6.1. Análises Estatísticas	75
4. Resultados	76
4.1. Identificação <i>in silico</i> de LTR- RTEs transcricionalmente ativos	77
4.1.1. Identificação da ORF completa, domínios conservados e características gerais das cópias completas de referência	81
4.1.2. Número de cópias completas e distribuição genômica dos LTR-RTEs no genoma <i>draft</i> de <i>E. grandis</i>	82
4.1.3. Nomenclatura e Classificação dos LTR-RTEs identificados	86
4.1.4. Identificação de sequências expressas dos LTR-RTEs	91
4.2. Análises IRAP	93
4.2.1. Análises de polimorfismos	93
4.3. Análises REMAP	101
4.3.1. Amplificação de marcadores REMAP	101
4.4. Similaridade Genética – IRAP e REMAP	107
4.5. Análises dos fragmentos polimórficos IRAP e REMAP	108
4.6. Desenvolvimento e análise de marcadores RBIP	112
4.7. Quantificação de famílias de LTR-RTEs por qPCR	123
4.8. Análise transcricional de LTR-RTEs	132
5. Discussão	132
5.1. LTR-RTEs caracterizados em eucalipto	133
5.2. Número de cópias dos LTR-RTEs	136
5.3. A diversidade e similaridade genética de espécies do gênero <i>Eucalyptus</i> por meio das técnicas IRAP e REMAP	138
5.4. Marcadores Moleculares gerados a partir dos LTR-RTEs	140
5.5. Análise transcricional dos LTR-RTEs	141
6. Conclusões	144

1. Introdução



1. 1. O Gênero *Eucalyptus*

O eucalipto é uma espécie vegetal pertencente a uma das maiores famílias do reino vegetal – a família Myrtaceae. Ela possui duas subfamílias, que compreendem 17 tribos, e estão divididas em 132 gêneros e 5671 espécies (Govaerts et al. 2008).

As plantas da família Myrtaceae distribuem-se por toda a América do Sul, Austrália e Sul da Ásia, muitas são economicamente importantes, atingindo diferentes setores industriais, como de especiarias, frutas, madeira e farmacológico (Govaerts et al. 2008). Entre os gêneros que apresentam destaque, devido sua importância econômica, é o está o *Eucalyptus*.

O gênero *Eucalyptus* compreende mais de 700 espécies que ocorrem predominantemente no Hemisfério Sul, e cuja origem é tida como no continente australiano (Ladiges et al. 2003). Seu primeiro relato data de 1770 realizado por exploradores europeus no norte da Austrália, que encaminharam as plantas para o Museu Britânico, onde o botânico inglês L'Héritier cunhou o nome “Eucalyptus”, vindo do grego raízes *eu* e *calyptos*, significando “*well*” e “*covered*”, respectivamente, o que faz referência ao opérculo do botão floral (Brooker e Kleinig 1994) e hoje, popularmente, é conhecido como eucalipto.

As plantas que compõem o gênero *Eucalyptus* apresentam diferentes formas, desde arbustos até árvores gigantes. Este gênero distribui-se em 15 subgêneros, dentre estes: *Corymbia* (~70 spp.), *Eucalyptus* (~110 spp.) e *Symphyomyrtus* (~500 spp.) (Merchant et al. 2006). Em relação ao seu cariótipo, o gênero *Eucalyptus* apresenta número cromossômico $2n=22$, com raros casos de aneuploidias (Atchison, 1947), e quebra dos cromossomos durante a metáfase (Oudjehih e Bentouati 2006).

Dentre as espécies do subgênero *Eucalyptus*, com interesse agrônômico, destaca-se o *Eucalyptus globulus*, que está entre as mais distribuídas por todo o mundo, por ser a principal fonte de óleo antisséptico (Khan et al. 2012), utilizado tanto em tratamentos medicinais, como na produção de madeiras para o setor de papel e celulose (Rodrigues et al. 1999); no entanto, esta espécie não se adaptou as regiões tropicais do globo (revisado por Salazar et al. 2013). Outra de grande cultivo é *Eucalyptus tereticornis*, reconhecida por ser uma das maiores fontes para madeira, lenha e matéria-prima para celulose e papel (Aggarwal et al. 2010), e também *Eucalyptus saligna*, que é encontrada em plantações no Brasil e Estados Unidos (Doi et al. 2010).

No Brasil, as espécies mais plantadas são *Eucalyptus urophylla* e *Eucalyptus grandis*, ambas amplamente utilizadas em programas de melhoramento pelas empresas do setor florestal devido à resistência a doenças e tolerância a condições ambientais adversas (revisado por Salazar et al. 2013).

Além de áreas verdes cultivadas com uma única espécie, também são encontrados híbridos, que podem ocorrer a partir de um sincronismo no tempo de florescimento entre plantas de espécies diferentes, bem como por cruzamentos direcionados por ação antrópica. A planta híbrida formada pode ser fértil ou não. Na Austrália, existem populações naturais de híbridos formados a partir do cruzamento entre *Eucalyptus crebra* e *Eucalyptus melanophloia*, além de outras espécies (Brooker e Kleinig 1994). No Brasil, o cruzamento entre as espécies de *E. grandis* e *E. urophylla*, gera um híbrido muito difundido devido sua qualidade e rendimento na produção de produtos em empresas do setor de papel e celulose (Carvalho e Nahuzs 2001).

1.1.1. A cultura do eucalipto no Brasil

No Brasil, dois terços das áreas de reflorestamento são compostos por árvores do gênero *Eucalyptus* e o restante são áreas ocupadas por árvores de *Pinus* (Sociedade Brasileira de Silvicultura – SBS - 2010).

Sua introdução no país deu-se no início do século XX, quando se planejava estimular o plantio de árvores ao longo das linhas férreas, pela Companhia Paulista de Estradas de Ferro, com o intuito de produzir madeira para combustível, estacas, toras, postes e dormentes.

O responsável pela escolha da espécie arbórea foi Edmundo Navarro de Andrade, que avaliou tempo de crescimento, adaptação e durabilidade da madeira de diferentes espécies florestais locais e exóticas, destacando-se o *Eucalyptus* sp (Martini 2004). Com a introdução do eucalipto por Navarro entre as árvores utilizadas para o reflorestamento, a ocupação das áreas passou por três fases distintas ao longo do último centenário: a primeira fase, até 1980, foi marcada pela formação da base florestal; a segunda fase, de 1980 a 2000, ocorreu à profissionalização do setor florestal; e a terceira fase, de 2000 até os dias atuais onde houve o estabelecimento no Brasil do setor de áreas verdes plantadas, com a ampliação de áreas de livo e a consolidação do processo de desenvolvimento tecnológico do setor (ABRAF 2012).

Atualmente, o plantio de eucalipto ocupa 5,1 milhões hectares no Brasil, e os principais estados produtores são Minas Gerais e São Paulo, levando o setor florestal a arrecadar mais de 7,6 bilhões de reais em tributos no ano de 2012, o que corresponde a 0,51% da arrecadação nacional (ABRAF 2013).

O eucalipto plantado é destinado para diferentes finalidades, como no uso medicinal para produção de xaropes, loções e pastilhas peitorais (Rocha e Santos 2008), e nas grandes plantações para produção de madeira para energia, celulose e papel, produção de móveis, construção civil, laminação, compensados, chapas de fibras, placas de OBS, moirões, postes, dormentes, óleos essenciais, processamento mecânico.

O desempenho do setor florestal brasileiro é favorecido pelas condições climáticas, mas pode também ser atribuído a tecnologia desenvolvida por instituições de pesquisa e empresas do país, o que leva o setor a ser reconhecido mundialmente por suas inovações na área de genética, melhoramento e propagação florestal (Soares 2010). O melhoramento florestal clássico é muito explorado para a avaliação de progênies, entretanto dados moleculares também são de grande importância para o melhoramento florestal, dado os longos ciclos rotacionais em estratégias de genética clássica (Harfouche et al. 2012).

1.1.2. Estudos moleculares no gênero *Eucalyptus*

Em plantas do gênero *Eucalyptus*, os estudos moleculares tiveram início na década de 90 (Steane et al. 1991), e ainda estão em fase de expansão, visto que a sua domesticação encontra-se nos estádios iniciais se comparada a outras culturas (Grattapaglia e Kirst 2008).

O melhoramento vegetal, nesse sentido, utiliza desde técnicas clássicas como a seleção de fenótipos superiores e análise de progênies com marcadores moleculares, transformação genética, entre outros, vindo a obter, em alguns casos, os genótipos com as características desejadas (Sartoretto et al. 2008).

Entre as diferentes técnicas mencionadas, destaca-se o uso de marcadores moleculares, que tem como principal objetivo a identificação de genes candidatos ao melhoramento, obtenção de dados sobre a variabilidade genética, garantia de pureza genética, construção de mapas genéticos e mapeamento de loci de herança quantitativa (QTLs) (Grattapaglia e Kirst 2008).

Entre as técnicas mais utilizadas dentro do gênero *Eucalyptus* e também em diferentes espécies da família Myrtaceae, os marcadores moleculares microsátélites são os que tem o maior número de trabalhos (Grattapaglia et al. 2012), substituindo grande parte dos estudos em que se utilizava as técnicas RAPD (*Random Amplification Polymorphisms*) e AFLP (*Amplified Fragment Length Polymorphism*), que eram empregadas para avaliar a pureza de espécies (Gaiotto et al. 1997; Leite et al. 2002), construção de mapas (Gan et al. 2003; Myburg et al. 2003), além de auxiliar na identificação de marcas relacionadas a características de interesse (Junghans et al. 2003; Domingues et al. 2006).

Recentemente, foram desenvolvidos marcadores microsátélites baseados em repetições com tetra, penta e hexanucleotídeos, o que unido a marcadores baseados em mutações pontuais ou *snips* (SNP – *Single Nucleotide Polymorphisms*), proporcionaram um avanço nas análises em distância genética, caracterizações espécie específicas, e também nas análises parentais em híbridos (conforme revisado por Grattapaglia et al. 2012).

A maioria dos trabalhos realizados com eucalipto tem por objetivo a melhoria de características primárias de produção, como a qualidade da madeira (Grattapaglia et al. 2009). Com uso das novas tecnologias de pirosequenciamento iniciaram-se estudos em larga escala para a descoberta e caracterização de novos marcadores e genes-chave para processos de interesse no melhoramento.

O primeiro trabalho nesta linha produziu sequências expressas pela técnica 454 (Novaes et al. 2008) para a descoberta de SNPs e caracterização geral de genes em eucalipto, a partir de um *pool* de tecidos de diversos genótipos. Posteriormente, foi feito um catálogo de genes a partir do sequenciamento de cDNA de xilema, proveniente de um híbrido de interesse para a indústria florestal (cruzamento *E. grandis* X *E. urophylla*), que resultou no primeiro banco de dados de sequências expressas de eucalipto baseadas em RNA-Seq, o *Eucspresso* (<http://eucspresso.bi.up.ac.za/>) (Mizrachi et al. 2010). Foram também desenvolvidos estudos de RNA-seq avaliando-se o impacto do déficit hídrico na expressão gênica (Villar et al. 2011), bem como o transcriptoma comparativo do desenvolvimento do xilema (que dá origem a madeira) das três principais espécies de eucalipto usadas na silvicultura: *E. globulus*, *E. grandis* e *E. urophylla* (Salazar et al. 2013). Os principais dados de RNA-seq obtidos nos últimos anos para o gênero *Eucalyptus* foram recentemente compilados em um banco de dados público, o *EUCANEXT* (<http://bioinfo03.ibi.unicamp.br/eucalyptusdb/>) (Nascimento et al. 2011; Salazar et al. 2013).

O genoma *draft* de *Eucalyptus grandis* foi liberado para pesquisas públicas em Janeiro de 2011, tendo como principal financiador o Departamento de Energia dos Estados

Unidos (Myburg et al. 2011; http://www.phytozome.net/search.php?method=Org_Egrandis). O genótipo de referência, BRASUZ1, teve seu DNA doado pela Suzano Papel e Celulose. Os dados são de uso público para pesquisas em pequena escala, e está em fase de preparação um manuscrito descrevendo as características gerais do genoma (<http://web.up.ac.za/eucagen/Default.aspx?a1=141&home=1>).

Mesmo com todos estes esforços em genômica e transcriptômica, pouco se conhece sobre a composição do seu genoma além de genes clássicos, codificantes de proteínas. Em plantas, é largamente conhecido que a porção mais representativa do genoma, corresponde a retrotransposons com LTR (*Long Terminal Repeats*) (Wicker e Keller 2007). Entretanto, os estudos até o momento sobre retrotransposons em eucalipto, limitam-se a descrever sua presença em sequências expressas ou ESTs (*Target Expressed Sequence*) (Bacci Júnior et al. 2005; Marcon et al. 2011), e em sequenciamento genômico aleatório (Lourenço 2004). Com isso, é de grande importância estudos que compreendam com maiores detalhes a dinâmica dos retrotransposons com LTR no genoma de eucalipto. Deste modo, com informações de diferentes níveis moleculares do eucalipto e com o desenvolvimento de modelos que definem o fenótipo com base na genética, epigenética, e propriedades do ambiente que o eucalipto ocupa (Grattapaglia e Kirst 2008), poderemos ter diversas respostas a questões evolutivas e de melhoramento genético neste gênero.

1.2. Elementos Transponíveis: aspectos históricos

Barbara McClintock, entre as décadas de 1940 e 1950, contemplou alterações no cromossomo nove de milho devido a existência de dois loci, que foram denominados como dissociador (Ds) e ativador (Ac), e propôs que estavam ligados a determinação da cor do aleurona da semente. As proposições de McClintock foram por observar a quebra do braço curto do cromossomo nove, e relacionou a excisão pela transposição de um elemento transponível, o que poderia ser constatado através da mudança de cor no aleurona. McClintock propôs que Ac controla a transposição de Ds, e este ao “saltar” para outra região liberava a síntese do pigmento na semente do milho (McClintock et al. 1950). A figura a seguir indica como ocorre a distribuição da pigmentação no aleurona na semente do milho.

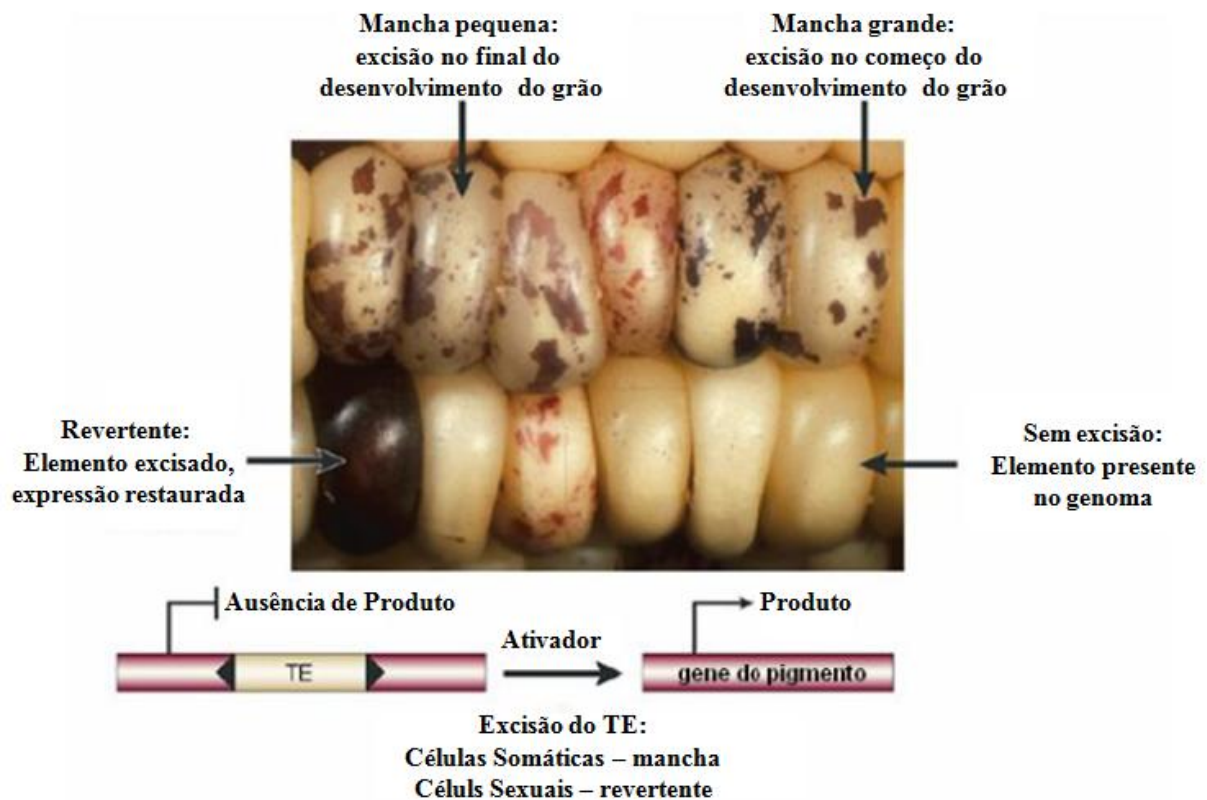


Figura 1: Variação de fenótipos com relação à coloração em grãos de milho devido à ação de elementos transponíveis. A inconstância de fenótipos ocorre pela inserção e excisão do elemento em gene responsável pela produção da enzima que confere coloração ao aleurona da semente do milho. Setores revertentes (com pigmento) ocorrem com a excisão do elemento. O tamanho do setor reflete o momento durante o desenvolvimento onde a excisão aconteceu. Adaptado de Feschotte et al. (2002).

Mais de uma década depois da descoberta de McClintock foi realizada a primeira caracterização molecular de elementos transponíveis (TEs) em *Escherichia coli*, por Jordan e colaboradores (1967), no entanto, muito se relutou para aceitar que os TEs compusessem os genomas dos seres vivos, e inicialmente passou a ser chamado de “DNA lixo” (*Junk DNA*), termo denominado por Sozumu Ohno (1972), na tentativa de explicar o excesso de material existente, porém sem desempenhar um papel evolutivo aparente (Ohno 1972). Já entre 1970 e 80, algumas hipóteses foram lançadas, como a de que os TEs seriam “parasitas genômicos”, pois apresentavam seu próprio modo de sobrevivência e mudança de local no genoma (Doolittle e Sapienza 1980). Outro termo também lançado foi o de “DNA egoísta”, em função dos elementos utilizarem a maquinaria celular somente para a sua própria replicação (Orgel e Crick 1980).

A partir dos anos de 1990, com o maior número de estudos em larga escala de sequenciamento, pôde-se observar que os TEs podem ser co-optados pelo genoma para a realização de funções celulares. Feschotte (2008) e Lisch (2013) revisaram casos nos quais os

TEs foram “domesticados”, visto que perderam a capacidade de transposição, tornando-se integrante de funções imprescindíveis para o ser vivo, inclusive como responsáveis pela adaptação dos genomas diferentes ambientes (Casacuberta e Gonzalez 2013).

1.2.1. Classificação e estrutura dos Elementos Transponíveis

A classificação molecular de TEs apresentada por Finnegan (1989) propôs a divisão dos TEs de acordo com a presença ou ausência de um RNA intermediário, dividindo-os em duas classes: os elementos da classe I referem-se ao retrotransposons, que utilizam um RNA intermediário para a sua retrotransposição, estes apresentam a estrutura genômica similar a de um retrovírus, e sua propagação é realizada por meio de um mecanismo conhecido como “copia e cola” (*copy and paste*), no final tem-se duas cópias a partir de uma inicial; e a classe II de elementos, ou transposons de DNA, que é capaz de se inserir em uma região do genoma diferente, realizando sua própria excisão e inserção. Este mecanismo ocorre com o auxílio da enzima transposase, que é codificada por uma sequência presente nesta região.

Além da divisão proposta por Finnegan, inúmeras outras classificações conflitantes foram propostas para os TEs, no entanto, Wicker e colaboradores (2007) propuseram um sistema universal de classificação de TEs eucariotos, que é de entendimento mais simples para não especialistas (figura 2).

Da mesma forma que Finnegan (1989), Wicker e colaboradores (2007) propuseram uma divisão inicial em duas classes de elementos, cuja divisão também é feita de acordo com a presença ou não de um RNA intermediário, mas inserindo critérios enzimáticos para classificá-los. A divisão dos elementos continua sendo entre Classe I, que ainda faz referência aos retrotransposons (RTEs) e a classe II, composta pelos transposons de DNA. Porém, além desta divisão, as classes são subdivididas em: 1) ordens, baseado no mecanismo de inserção, estrutura e na codificação da proteína, 2) superfamílias, mecanismos bioquímicos de transposição, e 3) famílias e subfamílias, com base na identidade de sequências nucleotídicas.

Classificação	Estrutura	TSD	Código	Ocorrência
Ordem Superfamília				
Classe I (retrotransposons)				
LTR	<i>Copia</i> → GAG AP INT RT RH →	4-6	RLC	P, M, F, O
	<i>Gypsy</i> → GAG AP RT RH INT →	4-6	RLG	P, M, F, O
	<i>Bel-Pao</i> → GAG AP RT RH INT →	4-6	RLB	M
	<i>Retrovirus</i> → GAG AP RT RH INT ENV →	4-6	RLR	M
	<i>ERV</i> → GAG AP RT RH INT ENV →	4-6	RLE	M
DIRS	<i>DIRS</i> → GAG AP RT RH YR ←	0	RYD	P, M, F, O
	<i>Ngaro</i> → GAG AP RT RH YR → →	0	RYN	M, F
	<i>VIPER</i> → GAG AP RT RH YR → →	0	RYV	O
PLE	<i>Penelope</i> ← RT EN →	Variável	RPP	P, M, F, O
LINE	<i>R2</i> RT EN -	Variável	RIR	M
	<i>RTE</i> APE RT -	Variável	RIT	M
	<i>Jockey</i> ORF1 APE RT -	Variável	RIJ	M
	<i>L1</i> ORF1 APE RT -	Variável	RIL	P, M, F, O
	<i>I</i> ORF1 APE RT RH -	Variável	RII	P, M, F
SINE	<i>tRNA</i> - -	Variável	RST	P, M, F
	<i>7SL</i> - -	Variável	RSL	P, M, F
	<i>5S</i> - -	Variável	RSS	M, O
Classe II (transposons de DNA) - Subclasse 1				
TIR	<i>Tc1-Mariner</i> Tase* ←	TA	DTT	P, M, F, O
	<i>hAT</i> Tase* ←	8	DTA	P, M, F, O
	<i>Mutator</i> Tase* ←	9-11	DTM	P, M, F, O
	<i>Mertin</i> Tase* ←	8-9	DTE	M, O
	<i>Transib</i> Tase* ←	5	DTR	M, F
	<i>P</i> Tase ←	8	DTP	P, M
	<i>PiggyBac</i> Tase ←	TTAA	DTB	M, O
	<i>PIF-Harbinger</i> Tase* ORF2 ←	3	DTH	P, M, F, O
	<i>CACTA</i> Tase ORF2 ←	2-3	DTC	P, M, F
Crypton	<i>Crypton</i> YR -	0	DYC	F
Classe II (transposons de DNA) - Subclasse 2				
Helitron	<i>Helitron</i> RPA Y2 HEL -	0	DHH	P, M, F
Maverick	<i>Maverick</i> C-INT ATP CYP POL B ←	6	DMM	M, F, O

Aspectos Estruturais

→ Longas Terminções repetidas ← Repetições terminais invertidas [] Região Codificante - - - Região não Codificante

- - - Motivo diagnóstico em região não codificante - / - Região que pode conter uma ou mais ORFs adicionais

Domínios das Proteínas codificadas

AP, Protease aspártica	APE, Endonuclease apuninica	ATP, ATPase de empacotamento	C-INT, C-integrase	CYP, Protease de cisteína	EN, Endonuclease
ENV, Proteína do envelope	GAG, Proteína do capsídeo	HEL, Helicase	INT, integrase	ORF, Quadro aberto de leitura de função desconhecida	
POL B, DNA polimerase B	RH, RNase H	RPA, Proteína de replicação A (somente encontrada em plantas)	Y2, YR com domínio deYY		
Tase, Transposas (*com domínio DDE)		YR, Recombinase de tirosina			

Grupos de espécies

P, Plantas M, Metazoários F, Fungos O, outros

Figura 2: Classificação sugerida para elementos transponíveis. Traduzido de Wicker et al. (2007).

1.2.2. Retrotransposons com LTR e sua distribuição

Os RTEs são elementos transponíveis pertencentes à classe I, cuja movimentação para novas localizações no cromossomo ocorre através de um RNA intermediário, que é convertido em DNA extracromossomal pela enzima transcriptase reversa, antes de ser inserido no genoma (Kumar e Bennetzen 1999; Sabot e Schulman 2006). A partir destas características de transposição e de replicação, pode ocorrer um aumento exponencial no número de cópias dos RTEs e no tamanho do genoma de seus hospedeiros (Kumar e Rzhetsky 1996; SanMiguel e Bennetzen 1998, Wicker e Keller 2007).

De acordo com a divisão de Wicker e colaboradores (2007), os RTEs são subdivididos em duas subclasses diferenciadas pela presença de longas repetições terminais nas porções 5' e 3' (LTRs) ou ausência (estes se dividem em grandes e pequenos elementos (LINE e SINE)). O tamanho das LTRs é variável de centenas a milhares de nucleotídeos, o que não interfere na classificação.

Os RTEs com LTRs (LTR-RTEs) possuem entre suas características a presença dos dinucleotídeos de ponto de partida 5'-TG-3' e de finalização 5'-CA-3' que os delimitam. Estes são flanqueados por *target site duplications* (TSDs) com quatro a seis pares de base (bp). Nas regiões 5' e 3' LTRs estão presentes os *primer binding site* (PBS) e *polypurine tract* (PPT), respectivamente, que apresentam de 20 a 25 bp e são conservados entre as espécies (Figura 3). Estas regiões indicam o local para a transcrição reversa do LTR-RTE a partir do DNA complementar (cDNA) (Wicker et al. 2007).

Entre as LTRs estão contidas uma ou mais matrizes de leitura aberta (ORFs), que são responsáveis pela produção das enzimas relacionadas à replicação do LTR-RTE. A ORF corresponde a dois domínios, a *gag* (que codifica uma proteína semelhante a um capsídeo viral), e *pol* (que codifica uma protease aspártica, integrase, transcriptase reversa e uma RNaseH) (Boeke e Corces 1989; Kumar e Bennetzen 1999) (Figura 3). Os LTR-RTEs autônomos são aqueles que apresentam a maquinaria citada acima, que garante a transposição do elemento de uma região para outra.

Outras ORFs são também encontradas em muitos LTRs-RTEs (Jern et al. 2005; Novikova et al. 2009; Steinbauerová et al. 2011), no entanto, não se sabe ao certo qual a sua função. Em muitos LTRs-RTEs de plantas são encontrados o gene *env*, (Du et al. 2010; Wright e Voytas 2002), que é responsável por codificar uma proteína transmembrana em retrovírus.

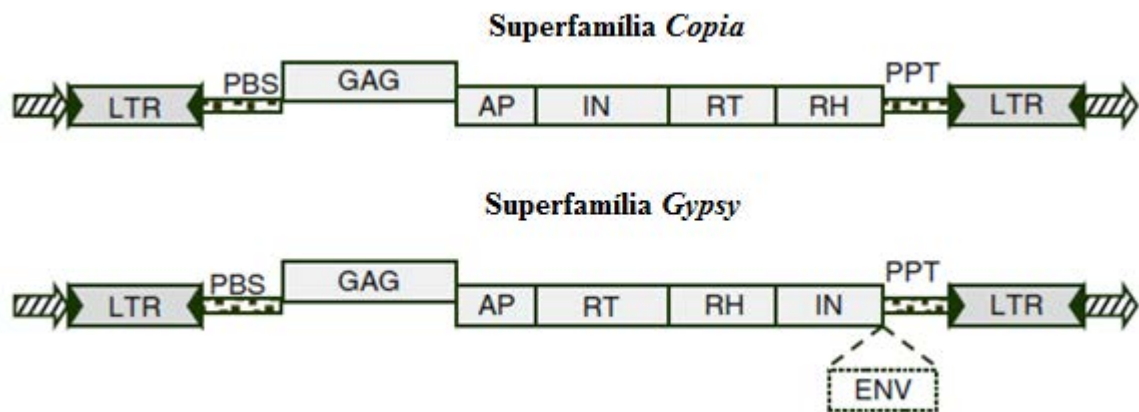


Figura 3: Organização dos LTR-RTEs. Os elementos podem ser classificados dentro de dois maiores grupos, as superfamílias *Copia* e *Gypsy*, nomeados após serem descobertos em *D. melanogaster* e levedura. Os elementos são flanqueados por cinco bp repetitivos (setas tracejadas). Os RTEs consistem em dois LTRs que contêm pequenas repetições invertidas nas extremidades (triângulos pretos) e um conjunto interno de poliproteínas responsáveis pela síntese, que contêm o domínio *gag*, que codifica a proteína formadora do capsídeo viral, similar ao de retrovírus, e uma protease aspártica (AP), no outro domínio estão as regiões responsáveis pela clivagem de poliproteínas dentro das unidades funcionais, a integrase (IN), responsável pela inserção de cópias dentro do genoma, e uma transcriptase reversa (RT) e uma RNaseH (RH), que sintetiza o cDNA a partir do RNA transcrito. Os LTR-RTEs da superfamília *Gypsy* são diferentes de *Copia* em relação a posição no domínio IN. Muitos LTR-RTEs da superfamília *Gypsy* contêm, como no retrovírus, um domínio *env*, geralmente expresso em uma posição diferenciada. Adaptado de Schulman et al. (2012).

A transcrição dos LTR-RTEs é iniciada na região U3 da 5'LTR, onde comumente encontra-se o TATA *box*, visto que entre as LTRs estão as regiões promotoras que regulam sua transcrição (Sugimoto et al. 2000). A partir da sequência de RNA transcrita terá início a síntese do DNA por meio da fita de RNA. A princípio, um *primer* tRNA emparelha-se próximo a porção 5' do RNA, o que também ocorre na região terminal 3', deste modo, a síntese ocorre primeiramente nas regiões ricas em repetições. O processo continua a partir da região 3', completando a sequência. O tRNA é degradado pela RNaseH, e por fim, uma fita complementar a esta é sintetizada, resultando em uma dupla fita de DNA, que será inserida ao DNA genômico através da ação da integrase, que permite a inserção nas porções 3' com OH livres, como pode ser observado na figura 4 (Kalendar et al. 2011; Finnegan 2012).

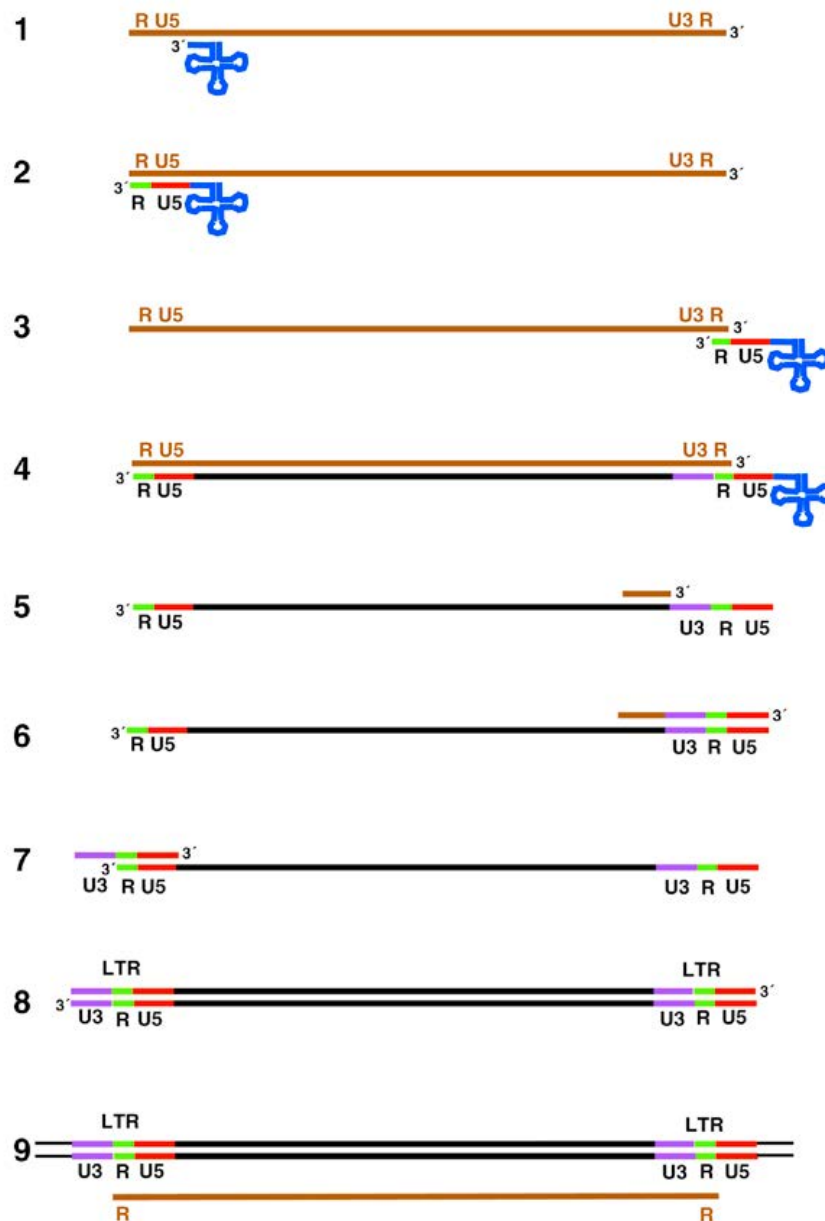


Figura 4: Mecanismo de Transposição nos LTR-RTEs: (1) Presença do RNA intermediário transcrito (marrom) de um LTR-RTE com o pareamento de um tRNA (azul), o iniciador está posicionado próximo a extremidade 5'; (2) isto serve como um iniciador para síntese de DNA pela transcriptase reversa; (3) a primeira sequência é copiada uma única vez a partir da porção 5' final do RNA (U5 – vermelho) e por uma pequena sequência repetitiva (R – verde) presente em ambas extremidades do RNA. A transcrição reversa pode continuar porque a cópia de R em um novo DNA sintetizado foi pareado com a sequência complementar a de outra porção do DNA; (4) isto permite que a síntese de DNA prossiga; (5) começando pela sequência única U3 (lilás) adjacente a R, e continua até U5 e R serem copiados novamente. A fita molde de RNA é degradada pela RNaseH; (6) um fragmento de polipurina é deixado para formação da fita de DNA complementar; (7) transcriptase reversa dá início a transcrição reversa a partir da complementariedade entre R e U5; (8) síntese prossegue em ambas as direções para formar uma dupla fita de DNA e LTR, feita a partir da U5, R e U3, em cada extremidade; (9) integrase age inserindo no DNA cromossomal. Modificado a partir de Finnegan (2012).

O modo como os LTR-RTEs realizam sua replicação é similar ao de um retrovírus, o que também ocorre quanto a organização da sequência, e como descrito anteriormente, a presença de vestígios de regiões codificantes do envelope nuclear. A partir destas observações e dos estudos filogenéticos, notou-se uma estreita relação entre os LTR-RTEs e os retrovírus, conforme revisado por Laten e Gaston (2012).

A diferença entre *Copia* e *Gypsy* está na ordem dos domínios em sua matriz de leitura aberta, INT-RT-Rvt e RT-Rvt-INT, respectivamente. Alguns elementos *Gypsy*, de linhagens mais semelhantes a um retrovírus, podem conter um domínio a mais (*env*, que facilita a infecção das células) (figura 3). Ambos os grupos são amplamente encontrados no reino vegetal, desde briófitas a angiospermas (Voytas 1990; Suoniemi et al. 1998; Novikova et al. 2008; Kolano et al. 2013).

Atualmente, sabe-se que em plantas, dentro das superfamílias *Copia* e *Gypsy*, há subdivisões internas (linhagens evolutivas), que muitas vezes trazem indícios particulares da distribuição e características estruturais de uma família de retrotransposons (Gorinsek et al. 2004; Wicker e Keller 2007; Llorens et al. 2009; Du et al. 2010; Domingues et al. 2012; El Baidouri e Panaud 2013), relacionados ao tamanho dos LTR-RTEs e das LTRs, número de cópias e organização de suas ORFs (Wicker e Keller 2007; Du et al. 2010), e estas linhagens se distribuem entre as monocotiledôneas e dicotiledôneas.

As famílias pertencentes à superfamília *Copia* são compostas comumente por elementos menores que seis quilobases (Kb) e com LTRs menores que 500 bp (Wang e Liu 2008). Geralmente são divididas dentro de seis linhagens evolutivas (*Bianca*, *TAR*, *Angela*, *Ale*, *Ivana* e *Maximus*). A linhagem evolutiva *Ale* é conhecida como a menor em tamanho, enquanto *Maximus* apresenta longos LTRs e regiões internas não codificantes, além de possuir uma segunda ORF, cuja função é desconhecida. Na linhagem evolutiva *Bianca* também ocorre uma segunda ORF anterior a poliproteína. As linhagens *Angela* e *TAR* apresentam grande similaridade entre si, sendo compactas e com LTRs pequenas (Wicker e Keller 2007).

Já os membros da superfamília *Gypsy* são classificados dentro de cinco linhagens evolutivas (*Tekay*, *Reina*, *CRM*, *Athila* e *Tat*). Comumente os membros desta superfamília apresentam mais que 6 Kb de comprimento com LTRs maiores que 500 bp (Wang e Liu 2008). A linhagem evolutiva *Athila* é caracterizada por possuir famílias de tamanho pequeno e conseqüentemente LTR pequenos e são registradas a presença de ORFs adicionais, assim como em *Tat*. Em *Reina* e *CRM* são comumente encontrados ORFs adicionais (Wang e Liu 2008).

Todas as linhagens listadas pertencentes as superfamília *Copia* e *Gypsy* tiveram sua origem antes mesmo da divergência entre monocotiledôneas e dicotiledôneas (Wang e Liu 2008; Du et al. 2010).

Vários trabalhos demonstraram que os LTR-RTEs estão relacionados à domesticação dos vegetais (revisado por Lisch 2013), como na diferenciação da cor em frutos de uvas (Kobayashi et al. 2004; Cadle-Davidson e Owens 2008; Shimazaki et al. 2011).

Deste modo, toda esta natureza ubíqua dos LTR-RTEs, leva a impactos nos genomas das plantas (Zhou et al. 2009; Kalendar et al. 2011; Agren e Wright 2011; Lisch 2013), e os LTR-RTEs somente são passados ao longo das gerações devido à transposição ter ocorrido nas células germinativas (Finnegan 2012), no entanto, a transposição que ocorre nas células somáticas, resulta em uma heterogeneidade entre indivíduos de uma mesma espécie.

1.2.3. O número de cópias de retrotransposons com LTR no genoma vegetal

Os TEs são responsáveis por grande parte da variação do tamanho do genoma, seu comportamento auxilia na explicação do paradoxo do Valor C, em que o tamanho do genoma de um organismo não tem relação direta com sua complexidade (Griffiths et al. 2009).

Genomas pequenos, como de *A. thaliana*, são compostos de 20 a 30% de TEs, já grandes genomas, como de milho e cevada, são compostos por mais de 85% de TEs (conforme revisado por Agren et al. 2011, El Baidouri e Panaud 2013). Com isto, os TEs são considerados os responsáveis pela evolução do tamanho do genoma de populações naturais (Whitney e Garland 2010), além de apresentarem um papel relevante na estrutura e função dos genomas (Feschotte 2008; Hollister e Gaut 2009), o que está diretamente relacionado à espécie vegetal (Bennetzen et al. 2005), conforme observado na figura 5.

No entanto, o aumento no tamanho do genoma das plantas não ocorre devido a proliferação de um grande número de famílias, e sim por algumas famílias específicas, que levam a um “inchaço genômico” (revisado por Estep et al. 2013). Entre os membros deste grupo estão algumas famílias de retrotransposons das superfamílias *Copia* e *Gypsy*.

No pinhão-mansão, a caracterização de LTR-RTEs da superfamília *Copia* demonstraram que três famílias estão distribuídas em um grande número no seu genoma (Alipour et al. 2013). Já em algodão, uma única família de LTR-RTE, *Gorge3*, da

superfamília *Gypsy* é a principal responsável por diferenças de tamanho de genoma no gênero *Gossypium* (Hawkins et al. 2006). De maneira similar, El Baidouri e Panaud (2013) indicaram que há uma relação de especificidade entre as superfamílias e o tamanho do genoma da espécie em que se encontram, ao investigarem oito genomas de plantas (*Arabidopsis thaliana*, *Arabidopsis lyrata*, uva, soja, arroz, *Brachypodium distachyon*, sorgo e milho).

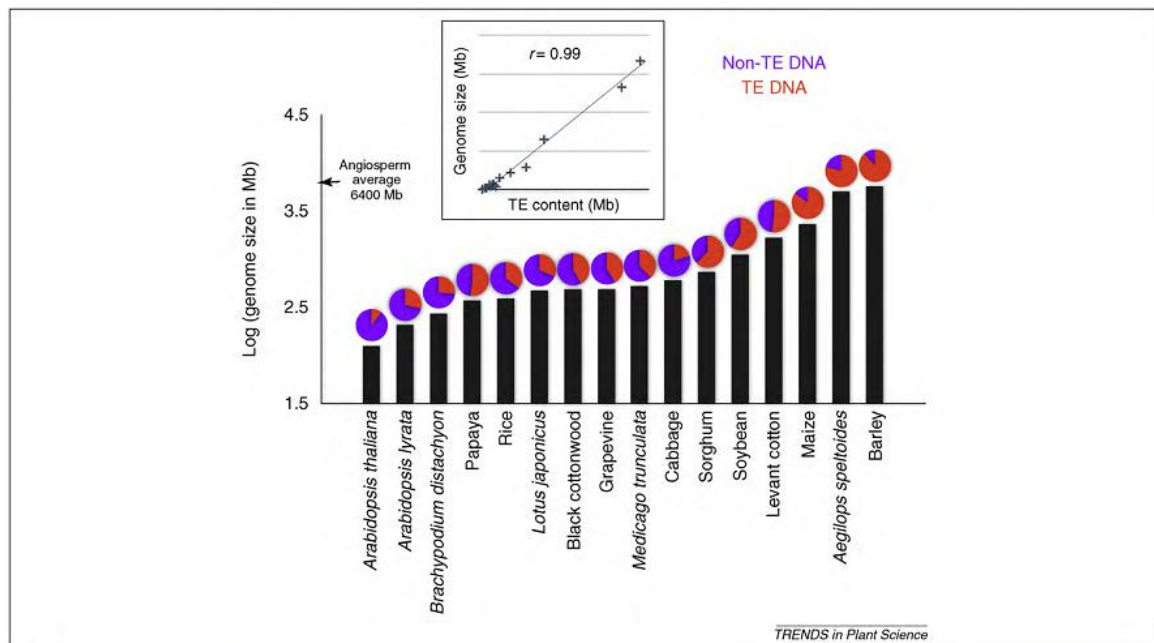


Figure 1. Genome size and TE content in angiosperm species are strongly correlated. All species are diploids. Data are taken from [1–4,7,12,90–101], <http://www.genome.ou.edu/medicago.html> for the estimate of *Medicago truncatula* genome size and from <http://data.kew.org/cvalues/> for *Aegilops speltoides* genome size.

Figura 5: Tamanho do genoma e conteúdo de TE em diferentes espécies de angiospermas, que se apresentam correlacionados. Dados retirados de Tenailon et al. (2010).

A busca pelo número de cópias dos LTR-RTEs nos genomas das plantas, podem ser realizadas através de diferentes métodos, baseados em citogenética molecular, hibridações de DNA genômico e avaliações diretas por qPCR, cujos resultados trazem maior confiabilidade devido às análises estatísticas empregadas (Lockey et al. 1998; Bustin 2000; Staton et al. 2009). Entre as técnicas usadas, a quantificação absoluta e a relativa por PCR quantitativo em tempo real (qPCR), possuem grande potencial por apresentar alta sensibilidade e reprodutibilidade.

As análises quantitativas por qPCR possibilitam inclusive inferir o número de LTR-RTEs completos e aqueles em que são defectivos. Em cevada, a família BARE1 ocupa 9,6% do seu genoma, embora parte deste valor corresponda somente a LTR. Deste modo,

somente parte dos elementos distribuídos no genoma encontram-se completos. Os autores propõem que o alto número de regiões LTRs é devido ao grande número de recombinações de BARE1 (Soleimani et al. 2006).

Em espécies de trigo, inferências realizadas pela quantificação relativa de LTR-RTEs por qPCR, possibilitaram dimensionar o quanto pode ser variável o número de um mesmo LTR-RTE em plantas, de um mesmo gênero, sugerindo ao menos 3 processos: 1) variação da atividade de LTR-RTEs ao longo a escala evolutiva de cada espécie, 2) a contribuição dos genitores para formação de um genoma híbrido são distintas, 3) eles não vão necessariamente levar um aumento no número de cópias de uma família de LTR-RTEs na formação de genomas híbridos ou poliplóides (Yaakov et al. 2013).

Deste modo, a abundância, dispersão e atividade dos LTR-RTEs possibilitam sua aplicação como marcadores moleculares (Kalendar e Schulman 2006; Kalendar et al. 2011), o que tem sido reforçado com o grande número de estudos já desenvolvidos em diversas espécies vegetais (Nair et al. 2005; Branco et al. 2007; D'Onofrino et al. 2009; Campbell et al. 2011; Pandotra et al. 2013; Fan et al. 2013).

1.2.4. Os retrotransposons com LTR como marcadores moleculares

A natureza ubíqua dos LTR-RTEs autônomos faz com que estes sejam referidos como boa fonte para desenvolvimento de marcadores moleculares (Smýkal et al. 2006; Kalendar et al. 2011; Schulman 2012). O grande número de marcadores moleculares já desenvolvidos em diferentes espécies vegetais confirma esta proposição (Smýkal et al. 2006; Branco et al. 2007; D'Onofrio et al. 2010; Hosid et al. 2012; Newaskar et al. 2013).

Os marcadores moleculares comumente são utilizados em análises de diversidade e variabilidade, e na construção de mapas genéticos ou mesmos em análises inter ou intraespecíficas. São também aplicados na identificação de genótipos de interesse ou na identificação dos parentais, além das análises fenotípicas, sistemáticas, biologia da conservação, ecologia molecular e no desenvolvimento biológico, conforme revisado por Poczai e colaboradores (2013).

Os LTR-RTEs possuem vantagens sobre outros tipos de marcadores, como os SNPs e os SSRs (*simple sequence repeats*), por apresentarem uma inserção no genoma irreversível, pela sua natureza ubíqua, estando presente em grande número de cópias e por ser

altamente heterogêneo em populações. Apresentam-se também amplamente distribuídos em regiões de eucromatina dos cromossomos e mostram polimorfismos tanto dentro, quanto entre espécies de plantas (Kumar e Hirochika 2001; Schulman 2012).

Desde a década de 1990, algumas técnicas foram implementadas para a utilização dos LTR-RTEs como marcadores moleculares (revisadas por Schulman et al. 2012), entre estas se destacam o IRAP (*Inter-Retrotransposon Amplified Polymorphism*) e o REMAP (*Retrotransposon-Microsatellite Amplified Polymorphism*), que são amplamente utilizados em plantas e se baseiam na amplificação de fragmentos entre dois LTR-RTEs com LTRs muito próximos no genoma e na detecção de LTR-RTEs próximos a microsatélites, conforme as figuras 5A e 5B, respectivamente. Ambas as técnicas já foram aplicadas para diferentes propósitos como mapeamento genético, diversidade genética e populacional, modificações da cromatina (Kalendar 2011). Os marcadores IRAP e REMAP foram desenvolvidos por Kalendar e colaboradores (1999), com o objetivo de que não houvesse a digestão do DNA a ser utilizado, como ocorre em marcadores derivados de AFLP.

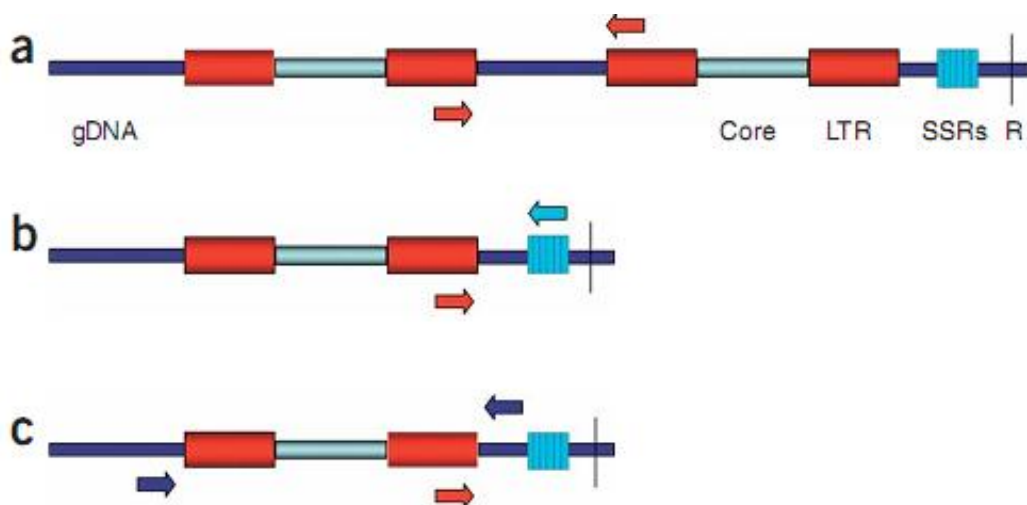


Figura 6: Métodos para construção de marcadores baseado em LTR-RTEs adaptado de Kalendar e Schulman (2006). (a) IRAP: Amplificação entre as LTRs de dois RTEs (setas vermelhas). DNA genômico (gDNA), identificado pela linha azul escuro, os *primers* estão indicados pelas setas vermelhas, o LTR-RTEs é representado pelas LTRs e pelo domínio central (*core*). Outras características também podem ser observadas no genoma, como: microsatélites (SSRs) e os sítios de restrição (R), porém IRAP não as utilizam. (b) REMAP: Amplificação dos *primers* entre um LTR (seta vermelha) e um SSRs (seta azul). (c) RBIP: Todas as posições (contendo um LTR-RTEs) apresentam geralmente um produto de amplificação com um *primer* LTR (seta vermelha) e um *primer* que flanqueia uma região adjacente a este (seta azul).

Os marcadores IRAP e REMAP já foram utilizados conjuntamente em diferentes espécies vegetais, como gengibre (Pandotra et al. 2013), alfafa (Mandoulakani et al. 2012), cacau (Santana et al. 2012), trigo (Carvalho et al. 2012), laranja (Biswas et al. 2010), caqui (Du et al. 2010), uva (Carcamo et al. 2010), trigo (Carvalho et al. 2010), arroz (Branco et al. 2007). Há um grande número de trabalhos que utilizaram somente o marcador IRAP, como em algodão (Newaskar et al. 2013), cevada (Campbell et al. 2011), língua de tucano (Jawdat et al. 2010), girassol (Vukich et al. 2009; Buti et al. 2013), batata (Lightbourn et al. 2007). Já o uso do marcador REMAP é utilizado sobretudo em espécies de interesse agrônômico, como em aveia (Tanhuanpaa et al. 2007) e lúpulo (Peredo et al. 2009).

A partir de observações de Smýkal (2006), as técnicas de IRAP e REMAP passaram a ser demonstradas como ferramentas de grande valia para o melhoramento vegetal, pois poderiam investigar como os LTR-RTEs estão agindo sob alguns genótipos de interesse, além de serem potenciais métodos para a obtenção de marcadores moleculares específicos para determinada espécie ou mesmo clones específicos de uma espécie. Em uva, D'Onofrio e colaboradores (2010), utilizaram 11 marcadores para diferenciação de variedades nesta planta. Em algodão, foi possível calcular na F_1 (produto do cruzamento entre anfidiplóides de diferentes espécies) a porcentagem de contribuição do genoma parental puro (Newaskar et al. 2013), ressaltando que estes marcadores podem ser utilizados conjuntamente a outras técnicas, como realizado por Palhares e colaboradores (2012), que empregaram os marcadores desenvolvidos a partir de LTR-RTEs junto a técnica de AFLP (*Amplified fragment length polymorphism*), confirmando assim a plasticidade destes marcadores.

Os marcadores IRAP e REMAP apresentam variações de eficiência na geração de dados. Um número maior de estudos utiliza a técnica IRAP em relação à REMAP, embora a maioria dos trabalhos empreguem os dois marcadores ao mesmo tempo. Conforme proposto por Castro e colaboradores (2012), a eficácia das técnicas está diretamente relacionada ao modo como os LTR-RTEs estão dispersos pelo genoma, que podem estar clusterizados e inseridos em tandem ou distantes entre si, ou a uma maior distância dos microsatélites adjacentes.

Outra técnica derivada das mencionados acima é o RBIP (*Retrotransposon-Based Insertion Polymorphism*), que necessita do conhecimento prévio das regiões 5' e 3', que flanqueiam o local onde o LTR-RTE foi inserido (Figura 5C). Neste caso, são utilizados dois marcadores conjuntamente, um deles já conhecido do elemento e outro na região que flanqueia o mesmo (Flavell et al. 1998). O marcador RBIP, trata-se de uma inserção locus específica, que produz um número menor de resultados que outras técnicas, mas é mais

preciso para estudos filogenéticos (Jing et al. 2005, 2010; Martin-Sanz et al. 2011) e também para a estudos populacionais (Vitte et al. 2004; Jing et al. 2005; Huang et al. 2008; Kokaeva et al. 2012) e de variabilidade (Baruch e Kashkush 2012).

Em análises de diversidade em ervilhas, pode-se avaliar o quanto estavam diversos os grupos em estudo com um pequeno número de marcadores, o que ocorre pelo fato dos LTR-RTEs constituírem grande parte do genoma mutável dos organismos (Jing et al. 2010).

Além do impacto pela sua prevalência nos genomas, os LTR-RTEs podem influenciar os genomas vegetais de acordo com sua localização cromossômica. Sabe-se que um dos componentes prevalentes da manutenção dos centrômeros em plantas são um subgrupo de LTR-RTEs da superfamília *Gypsy* (Yang et al. 2005; Luo et al. 2012), enquanto outros tem prevalência em regiões ricas em genes (Gao et al. 2012). Em termos transcricionais, os retrotransposons podem também ativar-se em situações ou tecidos bastante específicos (Vicient, 2010; Ochogavía et al. 2011) e influenciar inclusive a atividade transcricional de genes adjacentes (Kashkush et al. 2003; Kashkush e Khasdan, 2007).

1.2.5. Atividade transcricional de Retrotransposons com LTR em plantas

Os retrotransposons apresentam sua atividade transcricional altamente controlada pela metilação, com a participação da maquinaria de silenciamento gênico neste processo (Feschotte et al. 2002; Ito et al. 2011; Cui et al. 2012). Muitas vezes, o estado silenciado é herdado ao longo das gerações como uma estratégia em que o genoma hospedeiro se protege contra alterações.

Em relação a esta alta “afinação” da maquinaria de silenciamento gênico é comum a identificação de LTR-RTEs transcricionalmente ativos em plantas (Vicient et al. 2001; Araújo et al. 2005; Vicient 2010, Lopes et al. 2013), ressaltando que um LTR-RTE transcricionalmente ativo não é, em muitos casos, transposicionalmente ativo, ou seja, a identificação de um transcrito relacionado a um LTR-RTE não significa necessariamente que ele será inserido em uma outra região do genoma, conforme revisado por Feschotte e colaboradores (2002). A expressão “LTR-RTE ativo” é adequada a retrotransposons que tem simultaneamente sua transcrição e transposição ativada, como já descrito em arroz (Picault et al. 2009).

A ativação transcricional dos LTR-RTEs nos vegetais pode estar relacionada a diversos fatores, de modo que a modulação da sua transcrição em vegetais é realizada de forma *cis*, pelo reconhecimento do TATA Box e diferentes motivos responsivos a estímulos presentes na LTR. A ativação transcricional de famílias selecionadas de LTR-RTEs em cultura de tecidos já foi relatada para a família *Tos17* em arroz (Hirochika et al. 1996), *Tto1* em tabaco (Takeda et al. 1999), em diversas famílias em cana-de-açúcar (Araújo et al. 2005) e cevada (Alzohairy et al. 2012).

Hormônios vegetais também são desencadeadores de atividade transcricional de LTR-RTEs. Em morango, observou-se maior atividade transcricional do LTR-RTE *FaRE1*, em resposta a 2,4-D e ácido abscísico (He et al. 2012), e em Solanaceae, há diversos relatos de elementos regulados por hormônios, estresses bióticos e abióticos, conforme revisado por Grandbastien et al. (2005).

Análises em larga escala de ESTs em diversas espécies demonstraram que os LTR-RTEs não são apenas transcricionalmente ativos em resposta a estresses, mas também podem ser modulados entre diferentes órgãos e tecidos de uma planta (Rossi et al. 2001; Echenique et al. 2003; Vicient 2010; Rocheta et al. 2012). Análises individuais de famílias de LTR-RTEs evidenciaram inclusive que existem flutuações transcricionais entre elementos de uma mesma linhagem evolutiva (Domingues et al. 2012).

A família *Corky*, pertencente à superfamília *Gypsy*, em *Quercus súber*, apresentou atividade transcricional diretamente relacionada a resposta ao estresse ambiental e ao desenvolvimento da planta, permanecendo com a mesma expressão nas diferentes fases, mas com expressão diferencial nos tecidos (Rocheta et al. 2012). Por outro lado, já se conhece LTR-RTEs que controlam a expressão gênica como a coloração avermelhada da laranja-de-sangue, que é desencadeado pelo calor (Butelli et al. 2012).

A hibridação é um evento que ocorre em muitos casos de forma natural e tem sido demonstrado que levam a ativação dos LTR-RTEs (Kashkush et al. 2002), inclusive contribuindo para o rearranjo cromossomal e a modificação de genes adjacentes como observado em trigo (Kashkush et al. 2003). Além disso, também leva ao aumento no tamanho do genoma (Ungerer et al. 2009; Ungerer e Kawakami 2013).

No Brasil, o cultivo de híbridos de eucalipto é largamente utilizado, conhecidos como híbridos “urograndis” (*E. grandis* x *E. urophylla*) e “uroglobulus” (*E. grandis* x *E. globulus*). Situações como esta, provocam um “choque genômico” da hibridação de espécies, que podem eventualmente ter impacto na regulação transcricional de LTR-RTEs em eucalipto.

2. Objetivos



2.1. Objetivo Geral

O objetivo geral do presente trabalho foi caracterizar LTR-RTEs transcricionalmente ativos no genoma de cinco espécies do gênero *Eucalyptus*.

2.2. Objetivos Específicos

1. Identificar em bancos de dados públicos LTR-RTEs transcricionalmente ativos.
2. Caracterizar a sequência completa de LTR-RTEs selecionados no genoma sequenciado de *Eucalyptus grandis*.
3. Desenvolver marcadores moleculares baseados nos LTR-RTEs caracterizados.
4. Determinação do número de cópias de LTR-RTEs transcricionalmente ativos selecionados em duas espécies do gênero *Eucalyptus* e em seu híbrido F₁.
5. Compreender a atividade transcricional dos elementos caracterizados em cinco espécies do gênero *Eucalyptus* e um híbrido F₁.

3. Material e Métodos



3.1. Material Vegetal

O material vegetal utilizado para o desenvolvimento do trabalho pertence ao programa de melhoramento da Suzano Papel e Celulose mantidos em Itapetininga-SP e foi cultivado em viveiros da empresa (Tabela 1).

As técnicas de IRAP, REMAP e RBIP foram realizadas em cinco espécies de eucalipto: *Eucalyptus brassiana*, *Eucalyptus grandis*, *Eucalyptus saligna*, *Eucalyptus tereticornis* e *Eucalyptus urophylla* (Tabela 1).

Já para a quantificação de LTR-RTEs o material foi composto por mudas clonais de indivíduos de duas espécies: *E. urophylla* e *E. grandis* e o híbrido F₁ proveniente do cruzamento destas plantas (Tabela 1).

As análises transcricionais foram feitas em folhas, caule e raízes laterais de *Eucalyptus brassiana*, *Eucalyptus grandis*, *Eucalyptus saligna*, *Eucalyptus tereticornis*, *Eucalyptus urophylla* e do híbrido F₁ proveniente do cruzamento das plantas *E. urophylla* e *E. grandis*. Além disso, foram utilizadas as raízes de mudas de *E. grandis* submetidas a estresse osmótico por polietilenoglicol (PEG) (Tabela 1).

Tabela 1: Material vegetal utilizado para os diferentes experimentos do trabalho.

Espécies	Número de indivíduos	Tecido	Experimento utilizado
<i>E. brassiana</i>	20	Folha	Análises IRAP, REMAP e RBIP
<i>E. grandis</i>	20	Folha	Análises IRAP, REMAP e RBIP
<i>E. saligna</i>	20	Folha	Análises IRAP, REMAP e RBIP
<i>E. tereticornis</i>	20	Folha	Análises IRAP, REMAP e RBIP
<i>E. urophylla</i>	20	Folha	Análises IRAP, REMAP e RBIP
<i>E. grandis</i>	2	Folha	Quantificação por qPCR
<i>E. urophylla</i> **	2	Folha	Quantificação por qPCR
Híbrido F ₁ ***	2	Folha	Quantificação por qPCR
<i>E. brassiana</i>	10	Caule, folha e raiz	Atividade transcricional – RT-qPCR
<i>E. grandis</i>	10	Caule, folha e raiz	Atividade transcricional – RT-qPCR
<i>E. saligna</i>	10	Caule, folha e raiz	Atividade transcricional – RT-qPCR
<i>E. tereticornis</i>	10	Caule, folha e raiz	Atividade transcricional – RT-qPCR
<i>E. urophylla</i> **	10	Caule, folha e raiz	Atividade transcricional – RT-qPCR
Híbrido F ₁ ***	10	Caule, folha e raiz	Atividade transcricional – RT-qPCR
<i>E. grandis</i>	2	Raiz	Atividade transcricional – RT-qPCR

***Híbrido F₁ obtido do cruzamento de *E. grandis* e *E. urophylla*.

3.2. Identificação *in silico* de LTR-RTEs transcricionalmente ativos

A figura 7 sintetiza os passos realizados nesta fase do trabalho.

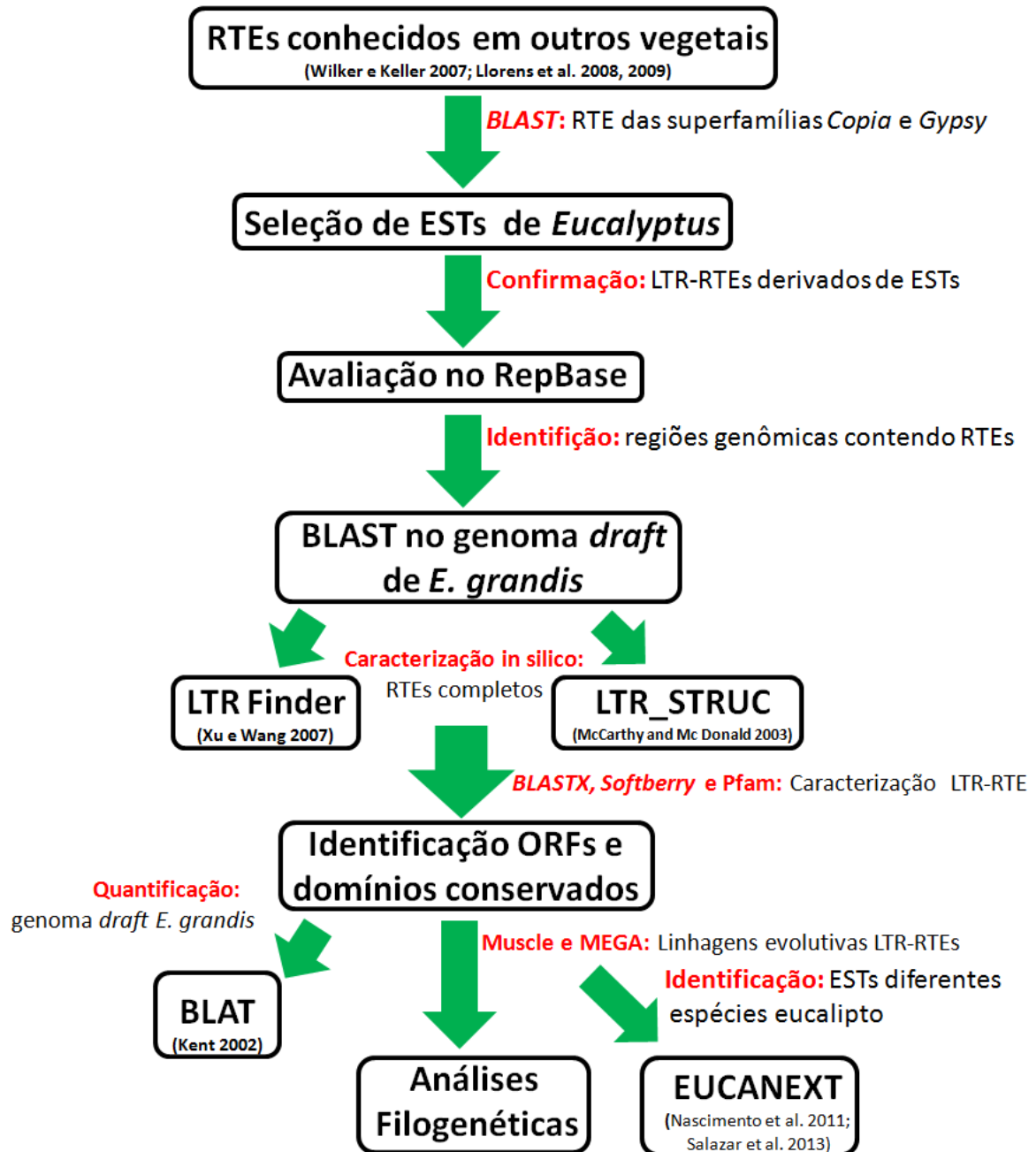


Figura 7: Etapas realizadas durante a identificação *in silico* de LTR-RTEs.

3.2.1. Caracterização inicial de LTR-RTEs transcricionalmente ativos

A identificação preliminar de LTR-RTEs transcricionalmente ativos candidatos foi realizada através da ferramenta *BLAST* utilizando-se como sondas eletrônicas (*queries*) LTR-RTEs das subdivisões *Copia* e *Gypsy* já descritos em outras espécies vegetais e/ou que estavam armazenados em bancos de dados. No caso de elementos da superfamília *Copia*, tomou-se como ponto de partida o trabalho de Wicker e Keller (2007) e para a superfamília *Gypsy*, trabalhos de Lloréns e colaboradores (2008, 2009). Para ambas superfamílias, também foram utilizados como referência LTR-RTEs armazenados no banco de dados RepBase (Jurka et al. 2005).

Estas sequências foram comparadas por *TBLASTX* a sequências de EST de espécies do gênero *Eucalyptus* disponíveis no banco de dados dbEST presente no GenBank. Os ESTs de *Eucalyptus sp.* com mais de 200 bp e identidade maior que 80% foram selecionados para avaliação no banco de dados RepBase (Jurka et al. 2005), para confirmar a presença de LTR-RTEs no transcrito identificado. Estas sequências então foram comparadas ao genoma *draft* (versão 1.63) de *E. grandis* (http://www.phytozome.com/search.php?show=blast&method=Org_Egrandis) por *BLASTN*, para a seleção de regiões genômicas com potenciais cópias completas dos LTR-RTEs analisados. Foram selecionados 20 kb à 5' e a 3' da região de identidade entre o EST e o genoma, para identificação de uma cópia completa de referência.

A identificação das cópias completas nas regiões selecionadas do genoma *draft* de *E. grandis* foram realizadas com os programas LTR_Finder (Xu e Wang 2007) e LTR_STRUC (McCarthy e McDonald 2003).

3.3. Identificação de ORFs nos LTR-RTEs candidatos

Para determinação do quadro aberto de leitura (ORFs) dos LTR-RTEs completos, foram retiradas as regiões LTR do elemento, seguida por avaliações *BLASTX* e utilização da ferramenta de predição de genes FGENESH, disponível na plataforma Softberry (<http://linux1.softberry.com/berry.phtml>), utilizando-se como referência ORFs completas de

retrotransposons depositados no GenBank. Todo este processo foi corrigido manualmente no caso de erros no quadro de leitura e códons de parada prematuros. Posteriormente, para avaliar se as proteínas obtidas correspondiam de fato a ORFs existentes em eucalipto, foram comparadas a versão 1.63 do genoma *draft* de *E. grandis*, que encontra-se armazenado no banco público *Phytozome* v. 7.0 (<http://www.phytozome.net/>).

3.3.1. Identificação de Domínios conservados nos LTR-RTEs selecionados

As ORFs encontradas foram analisadas utilizando o banco de dados Pfam (<http://pfam.sanger.ac.uk/>; Punta et al., 2011), o banco de dados CDD v 3.10 do NCBI (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>; Marchler-Bauer et al. 2010), e predição de domínios na ferramenta HHpred (<http://toolkit.tuebingen.mpg.de/hhpred/>; Kalev e Habec 2011). Os LTR-RTEs selecionados deveriam apresentar no mínimo duas regiões de domínios conservados característicos de retrotransposons em sua ORF, como: transcriptase reversa (Rvt), integrase (Int), gag, dedo de zinco (Zn) e *Chromodomain*.

3.4. Quantificação e distribuição genômica *in silico* dos LTR-RTEs selecionados no genoma *draft* de *E. grandis*

O levantamento do número de cópias de cada LTR-RTE das superfamílias *Copia* e *Gypsy* foi feito utilizando o algoritmo BLAT (Kent 2002), de modo análogo ao realizado por Moisy et al. (2008) para determinação do número de cópias de famílias de LTR-RTEs em uva. Foi utilizado o genoma *draft* de *E. grandis* versão 1.63, disponível na plataforma *Phytozome* 7.0. Os parâmetros utilizados foram: identidade mínima de 75% em uma região com de tamanho mínimo de 2000 bp, com o limite de criação de três *gaps*.

Os resultados obtidos foram curados manualmente, eliminando-se regiões de LTR-RTEs que apresentassem *gaps* maiores que 100 bp e também tamanho menor que 80% do LTR-RTE completo.

3.5. Análises filogenéticas

A região da transcriptase reversa identificada nos LTR-RTEs de eucalipto foram classificadas filogeneticamente com base em sequências utilizadas por Du e colaboradores (2010) e Neumann e colaboradores (2011) assim definindo-se linhagens evolutivas dos LTR-RTEs identificados. O alinhamento das sequências nucleotídicas foi feito com o programa MUSCLE (Edgar 2004), e foi construída uma árvore filogenética pelo método de máxima verossimilhança (Saitou e Nei 1987), com *bootstrap* de 1000 réplicas utilizando o modelo de substituição Kimura-2-parâmetros (Kimura 1980) e eliminação parcial de dados com gaps de 80%. Todas as análises filogenéticas foram feitas em ferramentas implementadas no software MEGA 6 (Tamura et al. 2013).

3.6. Identificação *in silico* de sequências expressas de LTR-RTEs em espécies do gênero *Eucalyptus*

As sequências completas de LTR-RTEs foram utilizadas como sondas eletrônicas contra ESTs Sanger de cinco espécies de eucalipto (*Eucalyptus camaldulensis*, *E. globulus*, *E. grandis*, *E. pellita* e *E. urophylla*) disponíveis no GenBank. As comparações foram realizadas no banco de dados EUCANEXT (Nascimento et al. 2011; Salazar et al. 2013; <http://bioinfo03.ibi.unicamp.br/eucalyptusdb/>), Foram somente considerados similares aqueles que apresentaram identidade maior de 80% e tamanho maior que 200 bp.

Foi também investigada a atividade transcricional de LTR-RTEs em dados de RNA-seq de *E. grandis* e *E. urophylla* (Salazar et al. 2013).

3.7. Extração de DNA genômico

3.7.1. Extração de DNA genômico para análises IRAP, REMAP e RBIP

Para as técnicas IRAP e REMAP o DNA foi extraído de folhas de 100 indivíduos não relacionados de cinco espécies do gênero *Eucalyptus* (item 4.1., Tabela 1), utilizando o protocolo de Ferreira e Grattapaglia (1998) com modificações.

A extração de DNA foi realizada utilizando-se uma massa de 110 mg de folhas, maceradas em nitrogênio líquido. A seguir, foi adicionado 700 µl de solução de extração CTAB (2%) com PVP e 2-mercaptoetanol (2 µl/ml), os quais foram aplicados em capela de exaustão e homogeneizados para que todo o material macerado entrasse em contato com a solução. Em seguida, a solução resultante foi mantida em banho-maria a 65 °C em um intervalo de 40 a 60 minutos. A primeira extração foi realizada com 600 µl de clorofórmio-álcool isoamílico (CIA 24:1) misturando-se a solução em vórtex com posterior centrifugação a 12000 rpm por 5 min.

Com a centrifugação ocorreu à separação do material em três fases. A fase superior foi transferida para outro tubo para ser utilizada nas etapas seguintes e as demais foram descartadas. A ela foi adicionado 400 µl de álcool isopropílico gelado e centrifugada nas mesmas condições anteriores. Após era esperada a formação de um precipitado; caso ele não se formasse, a etapa anterior era repetida. O sobrenadante do tubo foi descartado e o precipitado lavado com 800 µl de etanol 70% por 10 min, com posterior lavagem com 800 µl de etanol 95% por 3 min.

O precipitado foi seco em estufa a aproximadamente 50°C e ressuspendido em 100 µl de tampão TE (10 mM Tris, 1 mM EDTA pH 7,5) contendo 10 mg/µl de RNase. Para que ocorresse a digestão do RNA pela RNase, a solução foi incubada em banho-maria a 37°C até a diluição do precipitado por um período de até 120 min.

3.7.2. Extração do DNA genômico para quantificação do número de cópias de LTR-RTEs

Para se obter um DNA com maior integridade a ser utilizado em ensaios de quantificação de famílias de LTR-RTEs foi utilizado o kit Plant DNeasy[®] Plant Mini Kit da Qiagen[®], a partir do protocolo sugerido pelo fabricante.

3.7.3. Quantificação do DNA

A verificação da integridade do DNA obtido foi realizada após por eletroforese nas condições de 3 V/cm por 1 h a 25°C em tampão TBE 1X em gel de agarose 1% (Invitrogen) corado com brometo de etídeo. As amostras também foram quantificadas e avaliadas quanto a pureza em espectrofotômetro (NanoDrop[®] ND-1000). O DNA de cada amostra foi diluído para a concentração trabalho de 50 ng/μl.

3.8. Manipulação do DNA genômico

3.8.1. Iniciadores

Os *primers* foram desenhados utilizando o software PerlPrimer v1.1.17 (<http://perlprimer.sourceforge.net>), sendo excluídas as porcentagens GC (%GC) e os *clamp* GC.

3.8.1.1. Desenhos de *primers* IRAP e REMAP

Após o levantamento do número de cópias foram confeccionados 18 *primers* de LTR-RTEs das subdivisões *Copia* e *Gypsy*, para as análises IRAP e REMAP. Os *primers*

foram desenhados baseados na região da LTR, conforme proposto por Kalendar e Schulman (2006), utilizando-se o software PerlPrimer v1.1.17 (<http://perlprimer.sourceforge.net>) (figura 8).

Os marcadores foram denominados conforme o nome da família de LTR-RTE da qual foram gerados, com as iniciais „Eg“, que indicam a espécie vegetal em que foram obtidas as sequências, seguido pelo nome da família e LTR. Os números 3 e 5 ao final de cada nome indicam em qual das LTRs o *primer* foi posicionado (Figura 8A).

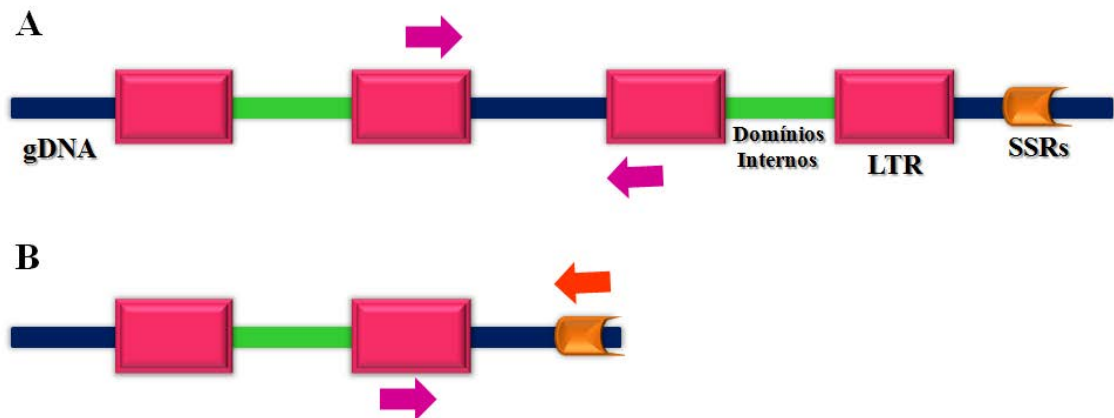


Figura 8: Representação do posicionamento dos *primers* para: (A) IRAP, cujas setas em rosa indicam a posição que os mesmos ocupam flanqueando dois LTR-RTEs adjacentes; (B) REMAP, cuja seta aponta local de pareamento dos *primers* (laranja: microsatélite; rosa: LTR), que flanqueiam a região entre o LTR-RTE e o microsatélite (SSRs) adjacente. gDNA: DNA genômico.

Para a técnica REMAP seguiu-se o proposto por Kalendar et al. (1999), cujos *primers* microsatélites foram selecionados (Tabela 2) e usados em conjuntamente com os *primers* confeccionados para a técnica IRAP (Figura 8B).

Tabela 2: *Primers* microssatélites utilizados conjuntamente com *primers* IRAP, na técnica REMAP (Kalendar et al. 1999).

<i>Primers</i>	<i>Sequências</i>
Micro 1	CTCTCTCTCTCTCTCTCTG
Micro 5	CACACACACACACACACAT
Micro 8	TGTGTGTGTGTGTGTGTGTC
Micro 12	AGAGAGAGAGAGAGAGAGAGT
Micro 13	ACACACACACACACACACG
Micro 18	CTCCTCCTCCTCCTCCTCCTCCTCA
Micro 21	GTGGTGGTGGTGGTGGTGGTGGTGA
Micro 23	CACCACCACCACCACCACCACCACCT
Micro 25	ACCACCACCACCACCACCACCACCACCG
Micro 30	TCGTCGTCGTCGTCGTCGTCGTCGTCGA

3.8.1.2. Desenhos dos marcadores RBIP

Para o desenvolvimento dos marcadores RBIP foi feita uma adequação da metodologia empregada por Flavell e colaboradores (1998). Por meio das técnicas IRAP e REMAP foram selecionados fragmentos polimórficos que identificassem polimorfismos espécie ou indivíduo específicos, que após clonagem e sequenciamento, tiveram as sequências consenso comparadas com o genoma *draft* de *E. grandis* e do marcador correspondente, para que se pudesse agrupá-los de acordo com o intervalo de identidade.

Os marcadores foram confeccionados utilizando as sequências consenso, obtidas pela técnica IRAP (Figura 9A), com o intuito de se obter marcadores espécie e/ou indivíduos específicos.

Do mesmo modo procurou-se identificar polimorfismos que estivessem relacionados à especificidade e ao genótipo pela técnica REMAP, foi utilizado o *primer* correspondente ao LTR conjuntamente com o marcador desenhado a partir da sequência consenso (Figura 9B).

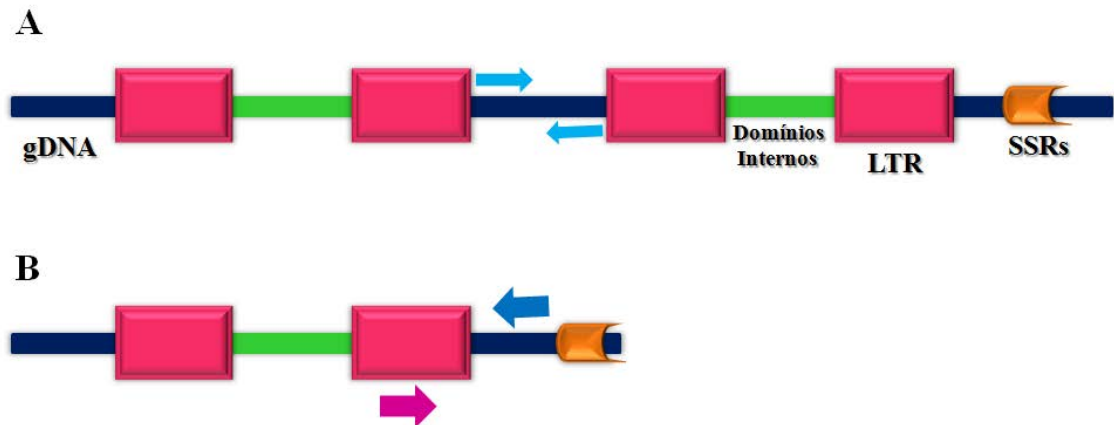


Figura 9: Representação do posicionamento dos *marcadores* RBIP obtidos a partir das técnicas: (A) IRAP, setas em azul indicam região entre LTRs (LTR-RTEs adjacentes), que os marcadores flanqueiam; e (B) seta em azul e rosa indicam região flanqueada por marcadores obtidos por REMAP.

3.8.2. Amplificações IRAP

Os testes iniciais para a técnica IRAP consistiram na padronização da reação em cadeira de polimerase (PCR), visando à adequação do que foi proposto por Kalendar e Schulman (2006). No entanto, em razão de problemas amplificação dos *primers* IRAP, foi utilizado o protocolo de Smýkal (2006), o qual realizou modificações a partir do proposto por Kalendar et al. (1999). A reação de PCR proposta por Smýkal (2006) consistiu no uso de 20 a 50 ng de DNA, 1 X PCR Buffer (75 mM Tris-HCl, pH 9,0; 50 mM KCl; 20 mM $(\text{NH}_4)_2\text{SO}_4$), 2,5 μM de MgCl_2 , 4 pmol de *primer*, 10 μM de dNTP e 1 U de *Taq* polimerase para um volume final de 20 μl . A amplificação proposta foi de 94 °C por 4 min, 35 ciclos de 94 °C por 30 s, 60 °C por 30 s e 72 °C por 2,5 min e uma extensão fina de 72 °C por 10 min.

A reação de PCR utilizada consistiu em 0,7 X de tampão (750 mM de Tris-HCl, 200 mM de $(\text{NH}_4)_2\text{SO}_4$), 25 mM de MgCl_2 , 40 ng de *primer*, 0,4 mM de dNTP, 25 ng de DNA e 0,3 U de *Taq* polimerase para um volume final de 10 μl . A amplificação utilizada foi 94°C por 4 min, 35 ciclos de 94°C por 40 s, 50°C por 2 min e 72°C por 3,50 min, com uma extensão final de 72°C por 5 min.

Para a visualização dos fragmentos foi utilizado gel de agarose corado com brometo de etídio seguindo o proposto por Kalendar e Schulman (2006). Foram testados diferentes concentrações de agarose, das quais a mais adequada foi 1,8 %. Além disso, para se obter um gel com melhor qualidade para visualização dos fragmentos, o mesmo

obrigatoriamente deve ser polimerizando por um período mínimo de 30 min. A corrida do gel foi realizada a 70 Volts (V) por um período de 5 a 7 h. Para a fotodocumentação utilizou-se luz ultra-violeta e máquina digital. Os géis foram fotodocumentados por 3 períodos de corrida (5, 6 e 7hS), para se confirmar o padrão de bandas para análises posteriores.

4.8.3. Amplificações REMAP

Do mesmo modo para a técnica IRAP, a técnica REMAP foi baseada no protocolo proposto por Smýkal (2006), com algumas adequações. A PCR consistiu em 1,0 X de tampão (750 mM de Tris-HCl, 200 mM de $(\text{NH}_4)_2\text{SO}_4$), 25 mM de MgCl_2 , 10 ng de cada *primer*, 0,6 mM de dNTP, 40 ng de DNA e 0,3 U de *Taq* polimerase para um volume final de 10 μl . A amplificação utilizada foi a 94°C por 4 min, 35 ciclos de 94°C por 1 min, 56°C por 1 min e 72°C por 1 min 50 s com uma extensão final de 72°C por 5 min.

A visualização dos fragmentos do gel de agarose foi realizada assim como o protocolo estabelecido para a técnica a IRAP (item 4.11.1.1).

3.8.4. Análises IRAP e REMAP

A análise dos fragmentos foi realizada visualmente, a partir dos arquivos de imagem digital abertos no Adobe Photoshop 12.0x0.32.

Para ambas as técnicas ao se identificar polimorfismos com um *primer*, no caso do IRAP, ou nas diferentes combinações de *primers*, em REMAP, os mesmos foram avaliados em mais 10 indivíduos da mesma espécie, para validação do perfil observado.

3.8.4.1. Análise de similaridade genética

Nas técnicas IRAP e REMAP foram detectados fragmentos polimórficos e monomórficos, cuja classificação foi realizada quanto a presença (1) e ausência em cada

indivíduo. A similaridade genética entre as cinco espécie de estudo foi realizada no software NTSYS-pc 2.10 (Rohlf 2000).

A média da similaridade para todos os genótipos foi usada como “*cutoff*” para definir os grupos por meio do coeficiente de Jukes & Cantor (1969) e o dendrograma foi obtido utilizando a média das distâncias entre todos os pares de genótipos para formação de cada grupo, utilizando-se o método de distância média (UPGMA).

3.8.5. Purificação de fragmentos polimórficos em IRAP e REMAP

Os fragmentos polimórficos observados a partir das combinações de *primer* foram excisados em três amostras diferentes e transferidas para tubos de 1,5 ml. A seguir, foram purificados com o kit *GFXTM PCR DNA and Gel Band Purification* da Amersham *Pharmacia Biotech* e o kit *DNA Gel Extraction Kit* da Norgen *Biotek Corp.*.

3.9. Plasmídeos e Bactérias utilizados

3.9.1. Plasmídeos

Para o clonagem de fragmentos purificados de gel foi utilizado o vetor plasmidial PGEM (PGEM®-T Easy Vector System – Promega), que possui 3016 pb e apresenta resistência a ampicilina.

3.9.2. Bactéria Competente

A cepa de *Escherichia coli* utilizada em todo o experimento foi DH10β.

3.9.2.1. Obtenção das Bactérias Competentes

Uma alíquota da cepa armazenada em freezer (-80°C) foi selecionada e inoculada em meio *Circle Grow* (CG, Bio 101, Inc.) líquido, previamente autoclavado, sendo colocada em agitação por 16 h, a 37°C e 220 rpm.

Após este período, foi adicionado 2 ml de solução MgCl₂ (1M), 2 ml de MgSO₄ (1 M) e 2 ml de meio contendo a cepa, a qual foi transferida para 200 ml de meio CG em erlenmeyer autoclavado, e colocado sob agitação de 220 rpm, a 37°C por 1 h e 30 min.

Transcorrido o tempo, uma alíquota de 1 ml foi levada a espectrofotômetro para se realizar a leitura de densidade até atingir-se o valor de OD₆₆₀ entre 0,5 e 0,6.

Ao alcançar este valor, a cultura foi transferida para tubo plástico de 50 ml e deixadas em gelo por 10 min. Em seguida foi centrifugada a 3600 rpm, por 10 min, a 4°C. O sobrenadante foi descartado e adicionou-se ao precipitado 15 ml de solução tampão RF1 pH 5,8 (30 mM C₂H₃O₂K – KAc-, 100 mM KCl, 50 mM MnCl, 10 mM CaCl, 15% de glicerol), com agitação leve por vortex e mantida em gelo por 30 min.

Posteriormente foi centrifugado a 37°C por 10 min a uma rotação de 3000 rpm. Transcorrido este período o sobrenadante foi descartado e se adicionou 8 ml de solução RF2 pH 6,8 (10 mM MOPS, 10 mM KCl, 75 mM CaCl, 15% de glicerol), a qual foi deixada em gelo por 30 min. A seguir, as células competentes foram alíquotadas em eppendorfs 1,5 ml, previamente autoclavados, e resfriados imediatamente em nitrogênio líquido.

3.9.2.2. Transformação de bactérias competentes

3.9.2.2.1. Ligação de fragmento de DNA

Os fragmentos foram ligados em 25 ng do vetor pGEM-T (Promega) utilizando-se diferentes volumes de DNA conforme a sua concentração, a partir da razão de 1:3, respectivamente, sendo mantida a 16°C por 16 h.

3.9.2.2.2. Transformação das bactérias competentes por choque térmico

A metodologia empregada segue o descrito por Hanahan (1983) com algumas modificações. Após a ligação do fragmento ao vetor pGEM-T (Promega), o produto da ligação foi utilizado para transformação de células competentes DH10- β de *Escherichia coli*, que foram retirados do freezer (-80°C) e mantidas em gelo por 15 min para sua posterior adição. Os tubos foram incubados em gelo por 30 min e em seguida levados a banho-maria por um período de 1 min e 30 s a 42°C, e voltaram novamente para o gelo por 5 min. Logo depois, adicionou-se 700 μ l de meio CG, seguindo-se a uma incubação sob agitação de 190 rpm por um período de 1 a 2 h. A cultura foi inoculada em meio sólido CG com ágar contendo 40 μ g/ml de X-Gal (5-bromo-4chloro-3-indolyl- β -D-galactoside), 250 μ g/ml de IPTG (isopropil thiogalactoside) e 100 mg/ml de ampicilina, posteriormente incubou-se por um período de 16 h a 37°C.

De seis a 10 colônias que poderiam conter o fragmento de interesse foram selecionadas com ponteiros estéreis para obtenção do DNA plasmidial.

3.9.3. Extração de DNA plasmidial em pequena escala (Miniprep)

Após a seleção das bactérias transformadas, prosseguiu-se com a extração do DNA plasmidial. Em fluxo laminar, de seis a dez colônias de bactérias foram coletadas individualmente com ponteira esterilizada em autoclave e colocadas separadamente em tubos para cultura de bactérias contendo 3 ml de meio CG com o antibiótico de seleção e levada ao agitador de tubos a 37°C (200 rpm) por um período de 18 h. Após este período 1 ml de cultura da bactéria foi transferido para microtubo de 2 ml e centrifugado por 1 min a 12000 rpm. O sobrenadante foi descartado e 1,5 ml de cultura da bactéria foi novamente adicionado ao tubo nas mesmas condições anteriores.

O material sobrenadante foi descartado e o pellet eluído em 75 μ l de Tampão TE (pH 8,0). Em seguida adicionou-se 300 μ l da solução Teens (TE pH 8,0, 0,5% SDS, 0,1 N NaOH) e 150 μ l de solução de acetato de potássio 3M (pH 5,3), vortexando por 5 s a cada nova solução adicionada. Posteriormente, o material foi centrifugado por 15 min a 12000 rpm.

A fase aquosa foi transferida para um novo tubo, na qual foi adicionado 1 µl de RNase (10 mg/ml) e mantido a 37°C por 20 min. A seguir, o material foi centrifugado por 5 min a 12000 rpm e a fase superior foi transferida para um novo tubo.

Um volume de 800 µl de etanol absoluto gelado (-20°C) foi adicionado ao sobrenadante, o qual foi agitado por vórtex por alguns segundos e centrifugado por 5 min a 12000 rpm. O pellet obtido foi lavado com 500 µl de etanol 70% (-20°C) e centrifugado por 2 min a 12000 rpm. Em seguida o sobrenadante foi descartado e o material foi seco a 37°C por 10 min. O pellet formado foi ressuscitado em 40 µl de água MilliQ autoclavada e armazenado a -20°C.

O DNA plasmidial obtido foi utilizado na reação de PCR com o *primer* M13 utilizando-se 1 X tampão (750 mM de Tris-HCl, 200 mM de (NH₄)₂SO₄), 25 mM de MgCl₂, 10 ng de cada *primer*, 0,5 mM de dNTP, 0,20 U de DNA *Taq* polimerase (5 U/µl) e água destilada ou deionizada autoclavada para o volume final de 25 µl. O programa de amplificação consistiu em uma desnaturação inicial de 3 min a 95°C, 40 ciclos de 45 s a 95°C, 45 s a 55°C e 1 min a 72°C, com extensão de 5 min a 72°C.

Para a confirmação da presença do fragmento de interesse foi feita a observação em gel de agarose a 1,2 %, a partir do tamanho do inserto obtido. Posteriormente três reações de PCR positivas, para cada fragmento polimórfico de cada indivíduo, foram purificadas utilizando a enzima Exosap (GE), conforme protocolo do fabricante e quantificadas em espectrofotômetro (NanoDrop ND-1000). Os produtos de PCR purificados foram enviados para sequenciamento Sanger pela empresa Macrogen, Inc, utilizando os *primers* universais SP6 e T7.

3.9.4. Análise das Sequências

A sequência consenso para cada clone foi montada através do programa DNA Baser 3.5.3 (Heracle Software, Germany), e a sequência de vetor foi excluída com base em análises na ferramenta *Vecscreen* do NCBI (*National Center for Biotechnology Information*, <http://www.ncbi.nlm.nih.gov/tools/vecscreen/>).

Em todas as sequências, procurou-se localizar os respectivos *primers*, e sua região de identidade no genoma *draft* de *E. grandis*. As sequências que não possuíam qualquer

identidade com o genoma foram comparadas com o banco de dados do NCBI utilizando a ferramenta *TBLASTN*. A partir desta análise, foram selecionadas as sequências para o desenvolvimento de marcadores RBIP.

3.10. Amplificação RBIP

A técnica de RBIP é fundamentada na técnica de PCR, e utiliza um conjunto de iniciadores baseado na sequência do LTR-RTE em estudo e na região genômica flanqueante de uma inserção individual (Flavell et al. 1998). A partir disso, pode-se identificar, para uma determinada inserção de LTR-RTE, se há alelo(s) com a presença do elemento bem como o(s) alelo(s) no(s) qual(is) não estão inseridos. Portanto, a inserção individual de um LTR-RTE poderá resultar na conversão para um marcador molecular codominante (Schulman et al. 2004). Para a amplificação de RBIPs foi utilizado o protocolo descrito por Schulman e colaboradores (2004), onde foram empregados os iniciadores desenhados conforme descrito no item 4.10.1.2.

3.10.1. Seleção de fragmentos para a técnica RBIP

A seleção dos fragmentos a partir das técnicas IRAP e REMAP, baseou-se na ausência e presença dos mesmos, os quais deveriam estar presente ou ausente em 90% das amostras analisadas (10 amostras de cada espécie), o que foi confirmado ampliando-se as análises para outras 10 amostras da espécie em questão.

As etapas posteriores de excisão, purificação, inserção, transformação, sequenciamento e análise das sequências foram realizadas conforme descrito anteriormente.

3.10.2. Padronização da técnica RBIP

Os testes iniciais para a técnica RBIP consistiram na padronização da reação de PCR a partir das reações já utilizadas para as técnicas IRAP e REMAP. A reação de PCR consistiu no uso de 75 a 100 ng de DNA, 1 X PCR Buffer (75 mM Tris-HCl, pH 9,0; 50 mM KCl; 20 mM $(\text{NH}_4)_2\text{SO}_4$), 25 μM de MgCl_2 , 20 pmol de *primer*, 10 μM de dNTP e 1,5 U de *Taq* polimerase para um volume final de 20 μl . A amplificação proposta foi de 94°C por 1 min, 38 ciclos de 94°C por 1 min, diferentes temperaturas de pareamento conforme o *primer* por 1 min e 15 seg e 72°C por 2 min e uma extensão fina de 72°C por 10 min.

Os fragmentos foram visualizados em gel de agarose 1,2 %, corado com brometo de etídio. A corrida do gel foi realizada a 90 V por um período de 2 h. A fotodocumentação foi feita em luz ultra-violeta com máquina digital.

3.10.3. Análises de fragmentos RBIP

As análises RBIP foram realizadas visualmente, a partir dos arquivos de imagem digital abertos no Adobe Photoshop 12.0 x 0.32.

3.11. Quantificação de LTR-RTEs por PCR Quantitativo (qPCR)

3.11.1. Escolha do normalizador

Foi utilizado como normalizador o gene *DUR3*, um transportador de membrana com alta afinidade a uréia e que se apresenta em cópia simples em diversos genomas investigados até o momento (de Michele et al. 2012). Utilizando-se o gene de *A. thaliana* (acesso NCBI NM_123906.4), foi identificado seu ortólogo no genoma *draft* de *E. grandis*. confirmando o observado em outras espécies, foi identificada apenas uma cópia do gene no genoma *draft* de *E. grandis*, que se encontra no scaffold 5, no intervalo de coordenadas 13.076.040 a 13.079.479.

3.11.2. Desenho dos *primers* para qPCR

Para a quantificação absoluta foi desenhado um par de *primers* para o gene normalizador de cópia única (DUR3), e desenhados *primers* para seis famílias de LTR-RTEs. Os *primers* foram baseados nas regiões LTR, *gag* e Rvt, de modo que resultassem em amplicons na faixa de 79 a 120 bp (figura 10).

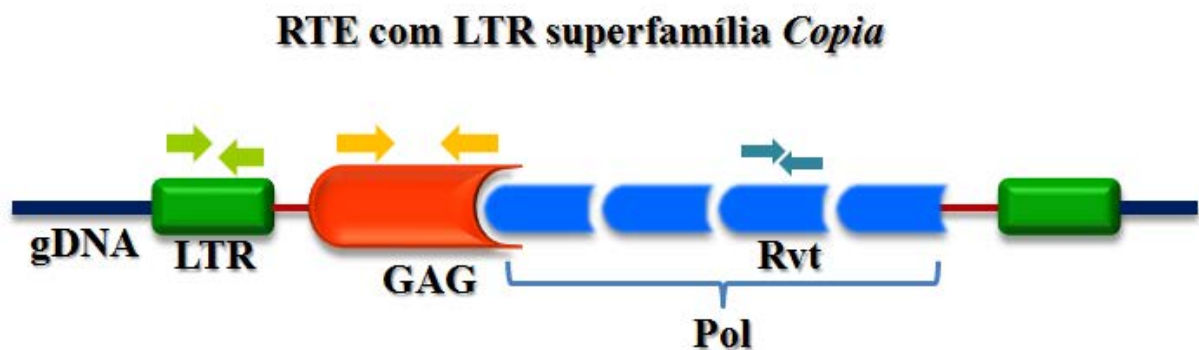


Figura 10: Representação do posicionamento dos *primers* para qPCR na LTR (setas em verde); regiões internas, *gag* (seta amarela) ou transcriptase reversa (Rvt) (setas em azul). gDNA: DNA genômico; GAG e Pol: regiões internas a ORF.

3.11.3. PCR quantitativo

A análise quantitativa dos LTR-RTEs foi realizada seguindo o proposto por Kraitshtein et al. (2010) e Yaakov et al. (2013), que é uma quantificação relativa baseada na amplificação de um gene de cópia única e utilizando-se as regiões de LTR-RTEs como “genes-alvo”. Foram selecionados seis LTR-RTEs para quantificação, com base nos dados de polimorfismo obtidos nos marcadores IRAP e REMAP: RLC_egAle_2, RLC_egMax_1, RLC_egGMR_1, RLC_egIvana_1, RLG_egTat_1 e RLG_egTekay_1.

Dado o impacto genômico de hibridizações interespecíficas na dispersão de elementos transponíveis, os LTR-RTEs foram quantificados em duas espécies de *Eucalyptus* (*E. grandis* e *E. urophylla*) e em uma geração filial híbrida (F₁) destas duas espécies.

Antes de iniciarem-se os experimentos por qPCR, os *primers* foram validados por PCR convencional. As reações de PCR foram baseadas no que foi descrito por Manetti et al. (2009) e Pagnotta et al. (2009). A reação de PCR consistiu em 1 X de tampão (750 mM de

Tris-HCl, 200 mM de $(\text{NH}_4)_2\text{SO}_4$, 25 mM de MgCl_2 , de 10 a 30 ng de cada *primer*, 0,6 mM de dNTP, 100 ng de DNA e 0,3 U de *Taq* polimerase para um volume final de 20 μl . A amplificação utilizada foi a 94°C por 4 min, 38 ciclos de 94°C por 1 min, diferentes temperaturas de pareamento específicas para cada *primer*, por 30 s e 72°C por 2 min, com uma extensão final de 72°C por 10 min.

Os fragmentos observados foram excisados e purificados com o kit DNA *Gel Extraction Kit* da Norgen *Biotek* Corp., com algumas modificações para se obter uma maior quantidade de material. Foi realizada a ligação e transformação (itens 4.11.1.6.1 e 4.11.1.6.2).

3.11.3.1. Obtenção de DNA plasmidial de fragmentos utilizados para construção da curva padrão

Com a transformação das bactérias foram selecionadas de duas a seis colônias de bactérias individualmente com ponteira estéril e colocadas em tubos contendo meio de cultura CG com o antibiótico de seleção e incubadas a 37°C por 18 h a 190 rpm.

Para garantir parâmetros de pureza e qualidade do DNA plasmidial, a extração de DNA utilizou o kit *Concert Rapid Plasmid Purification Systems* (Life Technologies™), conforme as recomendações do fabricante.

Para confirmar a presença do fragmento de interesse foram feitas duas reações de PCR, com dois pares de *primers* diferentes: uma com o conjunto de *primers* universais M13 (ver item 4.11.1.7) e outra com os *primers* específicos do LTR-RTE selecionado. O programa de amplificação consistiu em uma desnaturação inicial de 1 min a 96°C, 35 ciclos de 10 s a 96°C, 5 s a 50°C e 4 min a 60°C, permanecendo a 10°C. A eletroforese para visualização dos fragmentos foi feita em gel de agarose 1,2%, e a presença do fragmento com o tamanho de interesse classificou a amostra como positiva. Os fragmentos amplificados para qPCR foram clonados e sequenciados para confirmação da especificidade dos *primers* desenhados. O sequenciamento do DNA plasmidial foi feito para as duas fitas, com base em *primers* dos LTR-RTEs selecionados. O sequenciamento utilizou 120 ng de DNA. A reação de sequenciamento Sanger utilizou 1,0 μl de Tampão de Sequenciamento (Tris-HCl 1 M pH 9.0 e MgCl_2 50 mM), 2 μl *Big Dye*, 1,0 μl de *primer* (5 pmol/ μl) e água destilada ou deionizada

autoclavada para o volume final de 10 µl. O sequenciamento Sanger foi realizado no instrumento 3130 *Genetic Analyzer* (Applied Biosystems).

3.11.3.1.1. Análise das Sequências de DNA plasmidial para construção da curva padrão

O programa DNA Baser 3.5.3 (Heracle Software, Germany) foi utilizado para montar a sequência consenso a partir da leitura das duas fitas. Para confirmação dos resultados de sequenciamento, os fragmentos foram comparados ao RTE de referência utilizando-se *blast2seq*, e também ao genoma draft de *E. grandis* por BLASTN.

3.11.4. Obtenção da curva padrão para quantificação e cálculo de eficiência

O DNA plasmidial foi diluído inicialmente para uma concentração de 20 ng/µl. A partir desta diluição foram realizadas diluições seriadas de: 5×10^{-1} , 5×10^{-2} , $5 \cdot 10^{-3}$, 5×10^{-4} , 5×10^{-5} e 5×10^{-6} ng/µl.

As diferentes diluições foram utilizadas em reações de qPCR, que continham 5 µl de GoTaq® 2-Step-RT-qPCR System (Promega), 0,15 ng de cada iniciador, 3,7 µl de água ultrapura autoclavada. Os parâmetros de amplificação foram 95°C por 10 min, 40 ciclos de 95°C por 45 s e 60°C por 1 min, 95°C por 15 s, 60°C por 30 s e 95°C por 15 s.

A curva padrão foi produzida a partir dos resultados avaliados utilizando o software Step One 2.1 da Applied Biosystems, cuja montagem é realizada a partir de uma linha de tendência dos valores de C_T (valor da fluorescência acima do *threshold*) e o logaritmo do número de cópias. Foi utilizada uma faixa de C_T entre 18 e 32.

O número de cópias do DNA plasmidial para a primeira diluição ($5 \cdot 10^{-1}$ ng/µl) foi calculado através do número de massa da molécula de um plasmídeo, que foi baseado no número de bases do vetor plasmidial (3016 bp – PGEM-T Easy vector) mais o tamanho do fragmento inserido e na concentração do DNA plasmidial (Tabela 3), cujos valores foram calculados em www.scienceprimer.com/copy-number-calculator-for-realttime-pcr.

A tabela 3 mostra o índice de eficiência obtido para os iniciadores a serem utilizados em ensaios de quantificação de LTR-RTEs.

Tabela 3: *Primers* utilizados para a qPCR em que se obteve os valores de massa de uma cópia do plasmídeo e seus respectivos valores de eficiência e R².

LTR-RTEs	Região ORF	Primers	Massa*	Eficiência	R ²
RLC_egAle2_1	Gag	RTAle2Gag	6,17 x 10 ¹⁰	1,09	0,994
RLC_egAle2_1	LTR	RTAle2LTR	2,98 x 10 ¹⁰	1,01	0,999
RLC_egMax_1	Gag	RTMaxGag	2,17 x 10 ¹⁰	0,96	0,997
RLC_egMax_1	LTR	RTMaxLTR	5,96 x 10 ⁹	1,05	0,991
RLC_egGMR_1	Rvt	RTGMRRvt	3,55 x 10 ¹⁰	1,10	0,995
RLC_egGMR_1	LTR	RTGMRLTR	3,94 x 10 ¹⁰	1,04	0,998
RLC_egIvana_1	Gag	RTIvanaGag	1,47 x 10 ¹⁰	1,10	0,991
RLC_egIvana_1	LTR	RTIvanaLTR	2,65 x 10 ¹⁰	1,07	0,989
RLG_egTat_1	Gag	RTTatGag	3,73 x 10 ¹⁰	0,91	0,992
RLG_egTat_1	LTR	RTTatLTR	2,97 x 10 ¹⁰	1,03	0,999
RLG_egTekay_1	Gag	RTTekayGag	5,24 x 10 ¹⁰	1,10	0,994
RLG_egTekay_1	LTR	RTTekayLTR	4,03 x 10 ¹⁰	1,10	0,996
Gene normalizador	-	DUR3	2,47 x 10 ¹⁰	1,09	0,994

*Massa de uma cópia do plasmídeo

3.11.5. Análises quantitativas de LTR-RTEs selecionados

As análises aqui apresentadas foram desenvolvidas conforme Kraitshtein e colaboradores (2010) e Yaakov e colaboradores (2013).

Foi realizado um cálculo comparativo de C_T entre o LTR-RTE e o gene de cópia única DUR3 ($\Delta C_T = C_{Tamostra} - C_{TDUR3}$). Os resultados obtidos para *E. grandis* (Gra) foram utilizados como calibrador, para se calcular a quantidade relativa (RQ), em *E. urophylla* e F₁ ($\Delta \Delta C_T = \Delta C_{Tamostra} - \Delta C_{TGra}$). Os valores de RQ foram obtidos a partir da equação descrita por Livak e Schmittgen (2001), em que P é diretamente relacionado a ploidia da amostra e E a eficiência obtida a partir da curva de diluição:

$$RQ = P \times (2 \times E)^{-\Delta \Delta C_T}$$

Também foram realizadas comparações entre a LTR e região interna (*gag* ou *Rvt*) ($\Delta C_T = C_{TLTR} - C_{T(gag \text{ ou } Rvt)}$), cujo LTR foi utilizado como alvo e a região interna como referência. Os valores de RQ foram obtidos pela equação:

$$RQ = (1 + E)^{-\Delta C_T}$$

3.12. Análise transcricional

3.12.1. Condições experimentais

Sementes de *E. brassiana*, *E. grandis*, *E. saligna*, *E. tereticornis*, *E. urophylla* e do híbrido F₁ (*E. grandis* X *E. urophylla*) foram semeados em tubetes e cultivadas no viveiro da empresa Suzano Papel e Celulose mantidos em Itapetininga-SP, em condições de casa de vegetação. O caule principal, folhas maduras e raízes laterais de plantas com 90 dias de idade foram coletadas para as análises transcricionais. O material biológico foi congelado imediatamente em nitrogênio líquido e armazenados em freezer -80°C.

Foram também utilizadas raízes laterais de plantas de *E. grandis* (clone SP0791) com 60 dias de idade, cultivados pela Suzano Papel e Celulose, em que foram submetidos a estresse osmótico. Após o estabelecimento do estresse osmótico, as plantas foram transferidas para cultivo hidropônico e mantidos em solução contendo nutriente Hoagland's (75%) pH 6,0.

Após um mês as plantas foram submetidas a estresse osmótico com a adição de 1 L de polietileno glicol (PEG) 8000 (Sigma, USA), 25% das plantas foram mantidas na condição anterior para serem utilizadas como controle. O potencial osmótico de plantas tratadas foi de -0,6 MPa (WP4-T, Decagon Devices, Inc., England). Um sistema de compressão de ar foi usado para homogeneizar a solução e evitar a falta de oxigênio nas raízes.

As plantas controle e as estressadas compreenderam em 16 plântulas por bandeja, e 2 bandejas por tratamento, representando a replicatas biológicas, que foram usadas. A coleta das raízes foi feita a partir de 2 mudas por bandeja, que foi realizada no controle (tempo 0 h), após 6 h e 24 h após o tratamento com PEG.

3.12.2. Extração de RNA total

Para a extração de RNA total de caule, folha e raiz foi utilizado o protocolo de Chang e colaboradores (1993) com adaptações feitas por Korimbocus e colaboradores (2002). Os materiais utilizados nas extrações foram devidamente tratados com água deionizada contendo dietil pirocarbonato (DEPC) (0,01% v/v) e autoclavados a 130 por 45 min.

A maceração do material biológico foi realizada em duplicatas biológicas contendo cinco indivíduos em cada “*pool*”. Para as amostras tratadas com PEG também foram utilizadas duas duplicatas biológicas contendo um único indivíduo por duplicata. Com isso foram realizadas extrações de RNA total de 10 amostras de caule, 10 amostras de folhas, 10 amostras de raiz e seis amostras de raízes tratadas com PEG.

Foram utilizados 100 mg de tecido vegetal por extração, que foi pulverizado e posteriormente homogeneizado em 500 µl de tampão de extração autoclavado (Tris-HCl 1M pH 8,0; EDTA 0,5 M, NaCl 5 M, PVP e CTAB). Após um período de incubação de 15 minutos 65°C, adicionou-se às amostras clorofórmio e álcool isoamílico na proporção de 24:1. Em seguida as amostras foram centrifugadas a 10000 rpm durante 10 min à temperatura de 4°C. Após a centrifugação, a fase superior foi recuperada e transferida para tubos estéreis de 1,5 ml, repetindo-se a etapa anterior. Posteriormente foi adicionado um volume de LiCl 5 M e levados a -4°C por 16 h, para precipitação. Para cada 900 µl de amostra foram acrescentados 540 µl de álcool isopropílico.

Posteriormente foi realizada uma nova centrifugação por 30 min a 12.000 rpm à 4°C, e após o descarte do sobrenadante as amostras foram precipitadas com 200 µl de TE-SDS 1 %, 100 µl de NaCl 5 M e 300 µl de isopropanol sendo deixadas a -20°C por 30 min. Em seguida foi centrifugado a 12000 rpm por 20 min a 4°C.

O sobrenadante foi descartado e lavados com etanol 70 % e centrifugado a 10000 rpm por 5 minutos à 4°C. O material foi ressuspensionado em água DEPC (livre de RNase) e armazenado a -80°C.

3.12.2.1. Quantificação do RNA total

O material quantificado teve sua pureza avaliada em espectrofotômetro Nanodrop (NanoDrop® ND-1000). As medidas de absorvância foram realizadas nos comprimentos de onda de 230, 260 e 280 nm.

Suplementarmente, para avaliação da integridade do RNA total, uma alíquota foi submetida a eletroforese em gel de agarose 2 %, mantido a 90 V por 2 h.

3.12.3. Desenho dos *primers* para o RT-qPCR

Os *primers* desenhados nas LTR-RTEs foram baseados nas regiões *gag* e/ou Rvt, de modo que resultassem em amplicons na faixa de 79 a 120 bp, conforme apresentado na figura 10. Foi também utilizado um par de *primers* para o gene normalizador gliceraldeído-3-fosfato desidrogenase (GAPDH) (Juliana Pereira Bravo e Ivan de Godoy Maia, comunicação pessoal), e *primers* para as famílias de LTR-RTEs.

3.12.4. Síntese da primeira fita de cDNA

As amostras foram tratadas com DNase (MBI Fermentas) para eliminação de resíduos de DNA genômico proveniente da extração de ácidos nucleicos. O tratamento foi realizado com 2 µg de RNA, 2 µl de tampão de reação de DNase I (10x), 1 µl de DNase 1U/µl e água tratada com DEPC para volume final de 20 µl. A mistura foi incubada por a 37°C por 30 min e posteriormente inativou-se a DNase através da adição de 2 µl de EDTA 0,5 M e aquecimento por 10 minutos a 65°C.

Foi feita uma PCR convencional utilizando o gene GAPDH para confirmar a eliminação de DNA genômico residual. A reação de PCR consistiu em 1 X de tampão (750mM de Tris-HCl, 200mM de (NH₄)₂SO₄), 25 mM de MgCl₂, 20 ng de cada *primer*, 0,6 mM de dNTP, 100 ng de cDNA e 0,3 U de *Taq* polimerase para um volume final de 20 µl. A

amplificação utilizada foi a 94°C por 3 min, 38 ciclos de 94°C por 30 s, pareamento dos *primers* a 60°C por 45 s, extensão por 1 min a 72°C, com uma extensão final de 72°C por 10 min. A ausência de fragmentos indicava que o material estava livre de DNA genômico.

A partir deste material foi feita a síntese de cDNA com o kit GoScript™ Reverse Transcription System (Promega). A transcrição reversa foi feita a partir da adição de 1 µl de oligo(dT)₁₅ *primer* (1 µg/reação), 1 µl de *Random primer* (1 µg/reação) e 8 µl de RNA livre de DNA genômico. A mistura foi aquecida a 70°C por 5 min e resfriada a 4°C por 5 min. Posteriormente adicionou-se a mistura 4 µl de tampão GoScript™ (5X); 2 µl de MgCl₂ (25 mM); 1 µl PCR *Nucleotide Mix* (0,5 mM de cada nucleotídeo), 1 µl do inibidor de ribonuclease RNAsin, 1 µl de Transcriptase Reversa GoScript™ e completou-se com água DEPC para o volume final de 20 µl. A mistura foi incubada por 5 min a 25°C para o pareamento, seguida da extensão por 1 h a 42°C e inativação por 15 min a 70°C. As amostras de cDNA prontas foram armazenadas a -20°C.

3.12.5. PCR em Tempo Real (RT-qPCR)

As reações de RT-qPCR foram feitas em termociclador StepOnePlus Real Time PCR System (Applied Biosystems). De 150 a 250 ng de cDNA (tabela 4) foram utilizados com 5 µl GoTaq® 2-Step-RT-qPCR System (Promega), 0,15 ng de cada iniciador e água ultrapura autoclavada para completar o volume final de 10 µl. Os parâmetros da corrida foram 95°C por 10 min, 40 ciclos de 95°C por 45 s e 60°C por 1 min, 95°C por 15 s, 60°C por 30 s e 95°C por 15 s.

Para amostras nas quais o Ct foi superior a 32, uma reação de RT-PCR foi realizada para que se pudesse comprovar a expressão do elemento avaliado. A reação de PCR semi-quantitativa consistiu em 1 X de tampão (750mM de Tris-HCl, 200mM de (NH₄)₂SO₄), 25 mM de MgCl₂, 20 ng de cada *primer*, 0,6 mM de dNTP, 250 ng de cDNA e 0,3 U de *Taq* polimerase para um volume final de 20 µl. A amplificação utilizada foi a 94°C por 4 min, 38 ciclos de 94°C por 1 min, diferentes temperaturas de pareamento específicas para cada *primer* por 30 s e 72°C por 2 min, com uma extensão final de 72°C por 10 min. Os fragmentos foram observados em gel de agarose 2 % submetidos a 90 V por 2 h e fotodocumentados utilizando máquina digital. O gene GAPDH foi também utilizado como normalizador nas reações de RT-qPCR.

Tabela 4: Concentração de cDNA dos diferente tecidos usados para o desenvolvimento do RT-qPCR.

Espécies	Tecidos	Concentração	Primers
<i>E. brassiana</i>	Folha, caule e raiz	200 ng	RTAle1Rvt
<i>E. grandis</i>	Folha, caule e raiz	200 ng	RTAle1Rvt
Híbrido F ₁	Folha, caule e raiz	200 ng	RTAle1Rvt
<i>E. saligna</i>	Folha, caule e raiz	200 ng	RTAle1Rvt
<i>E. tereticornis</i>	Folha, caule e raiz	200 ng	RTAle1Rvt
<i>E. urophylla</i>	Folha, caule e raiz	200 ng	RTAle1Rvt
<i>E. grandis</i> (PEG)	Raiz	200 ng	RTAle1Rvt
<i>E. brassiana</i>	Folha, caule e raiz	200 ng	RTAle2Gag
<i>E. grandis</i>	Folha, caule e raiz	200 ng	RTAle2Gag
Híbrido F ₁	Folha, caule e raiz	200 ng	RTAle2Gag
<i>E. saligna</i>	Folha, caule e raiz	200 ng	RTAle2Gag
<i>E. tereticornis</i>	Folha, caule e raiz	200 ng	RTAle2Gag
<i>E. urophylla</i>	Folha, caule e raiz	200 ng	RTAle2Gag
<i>E. grandis</i> (PEG)	Raiz	200 ng	RTAle2Gag
<i>E. brassiana</i>	Folha, caule e raiz	200 ng	RTBiancaRvt
<i>E. grandis</i>	Folha, caule e raiz	200 ng	RTBiancaRvt
Híbrido F ₁	Folha, caule e raiz	200 ng	RTBiancaRvt
<i>E. saligna</i>	Folha, caule e raiz	200 ng	RTBiancaRvt
<i>E. tereticornis</i>	Folha, caule e raiz	200 ng	RTBiancaRvt
<i>E. urophylla</i>	Folha, caule e raiz	200 ng	RTBiancaRvt
<i>E. grandis</i> (PEG)	Raiz	200 ng	RTBiancaRvt
<i>E. brassiana</i>	Folha, caule e raiz	150 ng	RTMaxGag
<i>E. grandis</i>	Folha, caule e raiz	200 ng	RTMaxGag
Híbrido F ₁	Folha, caule e raiz	150 ng	RTMaxGag
<i>E. saligna</i>	Folha, caule e raiz	150 ng	RTMaxGag
<i>E. tereticornis</i>	Folha, caule e raiz	150 ng	RTMaxGag
<i>E. urophylla</i>	Folha, caule e raiz	150 ng	RTMaxGag
<i>E. grandis</i> (PEG)	Folha, caule e raiz	150 ng	RTMaxGag
<i>E. brassiana</i>	Folha, caule e raiz	200 ng	RTGMRRvt
<i>E. grandis</i>	Folha, caule e raiz	200 ng	RTGMRRvt
Híbrido F ₁	Folha, caule e raiz	200 ng	RTGMRRvt
<i>E. saligna</i>	Folha, caule e raiz	200 ng	RTGMRRvt
<i>E. tereticornis</i>	Folha, caule e raiz	200 ng	RTGMRRvt
<i>E. urophylla</i>	Folha, caule e raiz	200 ng	RTGMRRvt
<i>E. grandis</i> (PEG)	Raiz	200 ng	RTGMRRvt
<i>E. brassiana</i>	Folha, caule e raiz	200 ng	RTIvanaRvt
<i>E. grandis</i>	Folha, caule e raiz	200 ng	RTIvanaRvt
Híbrido F ₁	Folha, caule e raiz	200 ng	RTIvanaRvt
<i>E. saligna</i>	Folha, caule e raiz	200 ng	RTIvanaRvt
<i>E. tereticornis</i>	Folha, caule e raiz	200 ng	RTIvanaRvt
<i>E. urophylla</i>	Folha, caule e raiz	200 ng	RTIvanaRvt
<i>E. grandis</i> (PEG)	Raiz	200 ng	RTIvanaRvt
<i>E. brassiana</i>	Folha, caule e raiz	200 ng	RTTatGag
<i>E. grandis</i>	Folha, caule e raiz	200 ng	RTTatGag
Híbrido F ₁	Folha, caule e raiz	200 ng	RTTatGag
<i>E. saligna</i>	Folha, caule e raiz	200 ng	RTTatGag
<i>E. tereticornis</i>	Folha, caule e raiz	200 ng	RTTatGag
<i>E. urophylla</i>	Folha, caule e raiz	200 ng	RTTatGag
<i>E. grandis</i> (PEG)	Raiz	200 ng	RTTatGag
<i>E. brassiana</i>	Folha, caule e raiz	200 ng	RTTekayGag
<i>E. grandis</i>	Folha, caule e raiz	200 ng	RTTekayGag
Híbrido F ₁	Folha, caule e raiz	200 ng	RTTekayGag
<i>E. saligna</i>	Folha, caule e raiz	200 ng	RTTekayGag
<i>E. tereticornis</i>	Folha, caule e raiz	200 ng	RTTekayGag
<i>E. urophylla</i>	Folha, caule e raiz	200 ng	RTTekayGag
<i>E. grandis</i> (PEG)	Raiz	200 ng	RTTekayGag
<i>E. brassiana</i>	Folha, caule e raiz	200 ng	RTCauliGag
<i>E. grandis</i>	Folha, caule e raiz	200 ng	RTCauliGag
Híbrido F ₁	Folha, caule e raiz	200 ng	RTCauliGag
<i>E. saligna</i>	Folha, caule e raiz	200 ng	RTCauliGag
<i>E. tereticornis</i>	Folha, caule e raiz	200 ng	RTCauliGag
<i>E. urophylla</i>	Folha, caule e raiz	200 ng	RTCauliGag
<i>E. grandis</i> (PEG)	Raiz	200 ng	RTCauliGag

3.12.6. Cálculo de expressão relativa

Os resultados de Ct foram obtidos para cada um das triplicatas por indivíduo por meio do software Step One 2.1 da Applied Biosystems. A eficiência de cada *primer* foi determinada para cada genótipo por meio do software LinReg (versão 2013.0) (Ruijter et al. 2009), utilizando os valores de Ct do ciclo um ao 40. Foram avaliadas reações de RT-PCR que apresentaram eficiência igual ou superior a 90%.

Para todos os casos, foi utilizado como calibrador (expressão = 1) o tecido ou o ponto de estresse osmótico com maior valor de Ct. A expressão relativa foi calculada utilizando o método de $\Delta\Delta C_t$ descrito por Livak e Schmittgen (2001), na expressão $(1+E)^{\Delta\Delta C_t}$.

3.12.6.1. Análises Estatísticas

As diferenças entre os RQ dos tecidos e entre os tempos após a aplicação do PEG em raiz foram analisadas por análise de variância (ANOVA), seguida pela comparação entre os grupos por meio do teste de Fisher LSD (*Least significant difference*), no qual resultados com $p < 0,05$ foram considerados estatisticamente significativos. Os cálculos foram realizados utilizando-se o programa Assistat 7.7 beta (Silva e Azevedo 2009).

4. Resultados



4.1. Identificação *in silico* de LTR-RTEs transcricionalmente ativos

A partir de 146 LTR-RTEs completos caracterizados na literatura, foram identificadas 3727 sequências EST Sanger do gênero *Eucalyptus* com identidades entre 82 e 96% com LTR-RTEs. Após uma segunda análise contra o banco de dados RepBase, foram selecionados 88 ESTs similares a LTR-RTEs da superfamília *Copia* e 221 ESTs similares a LTR-RTEs da superfamília *Gypsy* (Figura 11).

Uma vez que os ESTs correspondem apenas a fragmentos de LTR-RTEs, os ESTs selecionados foram utilizados para identificar elementos completos no genoma *draft* de *Eucalyptus grandis* por meio da ferramenta BLASTN. Foram obtidas 139 regiões genômicas relacionadas às LTR-RTEs das superfamílias *Copia* e 290 regiões contendo potenciais LTR-RTEs da superfamília *Gypsy* (Figura 11). Estas regiões foram novamente submetidas à comparação no banco de dados RepBase, para delimitar regiões com LTR-RTEs completos. Um total de 313 regiões genômicas (103 *Copia* e 208 *Gypsy*) apresentaram regiões maiores que 200bp com similaridade a LTR-RTEs.

Para a identificação de cópias canônicas de LTR-RTEs completos utilizou-se o programa LTR_Finder nas regiões genômicas selecionadas e foram obtidos 58 e 51 LTR-RTEs completos das superfamílias *Copia* e *Gypsy*, respectivamente. Para as regiões genômicas nas quais o programa LTR_Finder não detectou elementos completos, foi utilizado o software LTR_STRUC, adicionando-se mais quatro e oito LTR-RTEs das superfamílias *Copia* e *Gypsy*. Desta forma, foram identificados 62 LTR-RTEs pertencentes à superfamília *Copia* e 59 a superfamília *Gypsy*.

Os LTR-RTEs completos acima identificados foram novamente comparados por meio da ferramenta BLASTN contra ESTs do gênero *Eucalyptus* no banco de dados GenBank, visando confirmar que estes elementos eram relacionados a ESTs e, portanto, transcricionalmente ativos. Com isso, foram selecionados 23 elementos *Copia* e nove elementos *Gypsy* (80% de similaridade numa região igual ou maior que 200 bp).

Alguns dos LTR-RTEs completos selecionados poderiam ser dois membros de uma mesma família de retrotransposons, considerando-se a classificação de Wicker e colaboradores (2007), para eliminar as redundâncias e agrupar elementos de uma mesma família, os LTR-RTEs completos identificados foram comparados entre si, utilizando-se a ferramenta BLASTN2SEQ.

Deste modo foram identificadas seis famílias de LTR-RTEs pertencentes a superfamília *Copia* (Tabela 5) e três famílias de LTR-RTEs pertencentes a superfamília *Gypsy* (tabela 6) relacionadas a ESTs de eucalipto. A figura 11 sintetiza os passos realizados nesta análise.

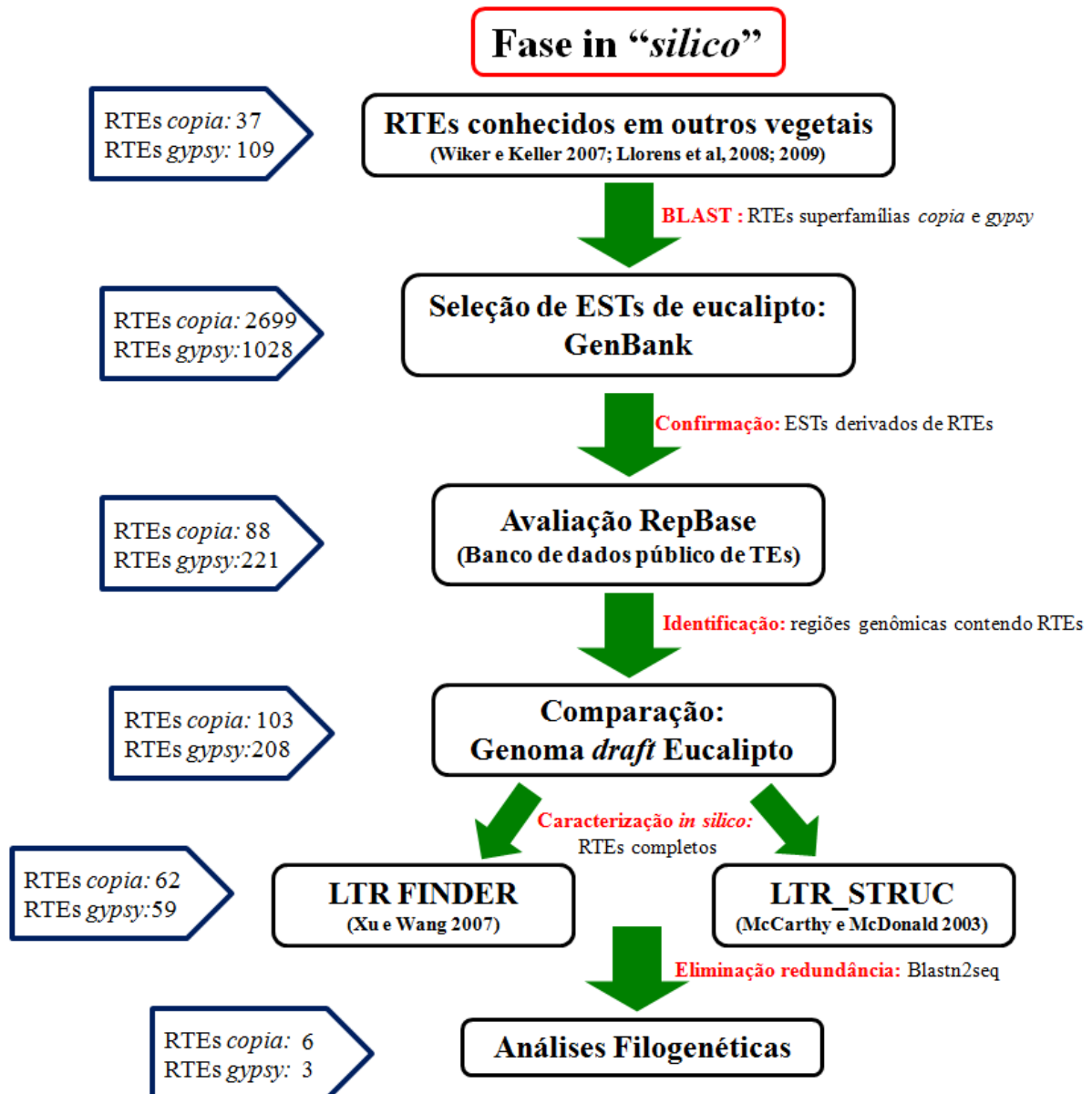


Figura 11: Etapas realizadas durante a identificação *in silico*, com o número de sequência identificadas em cada passo.

Tabela 5: LTR-RTEs pertencentes a superfamília *Copia* identificados a partir de LTR-RTEs de famílias já descritas na literatura e armazenadas no banco de dados RepBase, comparados ao genoma *draft* de *E. grandis*.

Família*	Nome recebido	1º hit – ESTs de eucalipto no dbEST - GenBank	1º Hit no EuDB (supercontig)	Coordenada	RepBase (frag/tam) ²	LTR_FINDER	LTR_STRUC
Atara	RLC_egAle_1	CT980515.1	96	749169 - 750102	1/938	sim	não
ATCopia11-116		CT985848.1	96	749169 - 750002	2/833	sim	não
ATCopia4	RLC_egAle_1	CT986411.1	31	2248975 - 2249669	9/4231	sim	sim
Atrel		CT984725.1	4	562511 - 563252	7/4442	sim	não
Atrel		CT981235.1	51	15383 - 15937	4/4807	sim	sim
Castor		CT984725.1	15	7311750 - 7312135	8/8588	sim	sim
ATCopia78	RLC_egMax_1	CT985743.1	72	178987 - 179834	3/604	sim	não
ATCopia78		CU405211.1	71	1908500 - 1908870	2/576	sim	não
ATCopia43		CT981516.1	341	152481 - 153410	2/639	sim	sim
ATCopia95		CU397599.1	267	40947 - 44865	1/321	sim	não
Alina		CT985848.1	44	3880559 - 3902147	3/3785	não	sim
ATCopia95	RLC_egBianca_1	CU397599.1	267	44865 - 40947	1/321	sim	não
Adena		CU405449.1	71	1907425 - 1907079	4/1281	sim	não
TLC		CU405449.1	70	1547564 - 1547815	18/10781	sim	sim
Bianca	RLC_egGMR_1	ES589847.1	39	30269 - 30382	3/5031	sim	sim
Bianca		ES595774.1	39	27143 - 27352	3/5031	sim	sim
Hopscotch		ES589702.1	39	30265 - 30358	3/5031	sim	sim
Ostonor	RLC_egIvana_1	CT980981.1	74	1998152 - 1998667	7/6153	sim	sim
Atara	RLC_egAle_1	CT980515.1	96	749169 - 750102	1/938	sim	não
ATCopia11-116		CT985848.1	96	749169 - 750002	2/833	sim	não
ATCopia4	RLC_egAle_1	CT986411.1	31	2248975 - 2249669	9/4231	sim	sim
Atrel		CT984725.1	4	562511 - 563252	7/4442	sim	não
Atrel		CT981235.1	51	15383 - 15937	4/4807	sim	sim

*: Dados obtidos a partir de Wicker e Keller 2007; ¹: número; ²: quantidade de fragmentos encontrados e o tamanho.

Tabela 6: LTRs-RTEs pertencente a superfamília *Gypsy* identificados a partir de LTRs-RTEs de famílias já descritas e armazenados no banco de dados RepBase, comparados ao genoma *draft* de *E. grandis*.

Família*	Nome recebido	1º hit – ESTs de eucalipto	1º Hit no EuDB	Coordenada	RepBase	LTR_FINDER	LTR_STRUC
		no dbEST - GenBank	(supercontig)		(frag/tam) ¹		
Athila4	RLG_egTat_1	CT982699.1	412	27612 - 28304	1/774	sim	não
Gloin		CT987915.1	412	27612 - 28517	1/992	sim	não
Tat		CT984201.1	1	1947501 - 1948299	2/600	sim	sim
Ogre	RLG_egTekay_1	ES588593.1	0	13833437 - 13833799	15/7144	sim	sim
Gypsy13		GO248067.1	24	833168 - 833196	3/1433	sim	não
Cereba	RLG_egCauli_1	CU398812.1	218	101367 - 101463	6/4025	sim	não
Athila4	RLG_egTat_1	CT982699.1	412	27612 - 28304	1/774	sim	não
Gloin		CT987915.1	412	27612 - 28517	1/992	sim	não
Tat		CT984201.1	1	1947501 - 1948299	2/600	sim	sim

*: Dados obtidos a partir de Lloréns et al. (2008; 2009); ¹: quantidade de fragmentos encontrados e o tamanho.

4.1.1. Identificação da ORF completa, domínios conservados e características gerais das cópias completas de referência

A partir das nove sequências de LTR-RTEs das superfamílias *Copia* e *Gypsy* transcricionalmente ativos, foi realizada a identificação da ORF de cada uma das sequências de referência utilizando-se diferentes estratégias.

Para uma das sequências de nucleotídeos dos LTRs-RTEs da superfamília *Copia* a região codificante não apresentava regiões de parada prematuros ou *frameshifts* (RLC_egBianca_1), possibilitando a identificação direta apenas pelo reconhecimento dos sítios ATG e do códon de parada. Com base nas comparações realizadas com as sequências depositadas no GenBank, e checagem manual, foram identificadas seis e três ORFs para LTR-RTEs das superfamílias *Copia* e *Gypsy*, respectivamente.

A caracterização da estrutura dos LTR-RTEs e a ordem dos domínios conservados encontrados na ORF são apresentadas nas figuras 12 e 13.

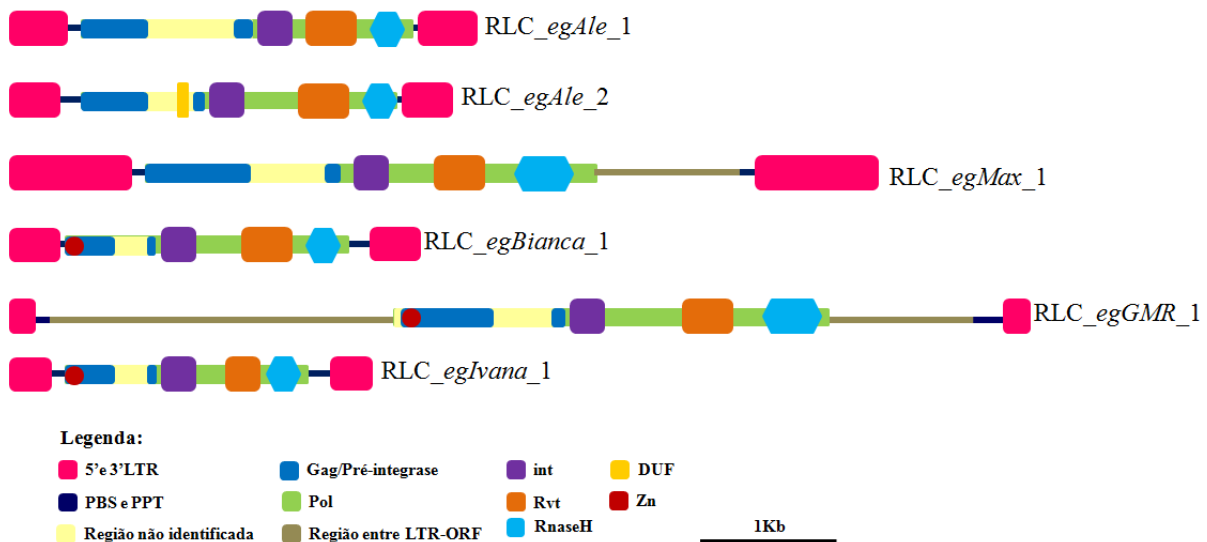


Figura 12: Estrutura hipotética de LTR-RTEs da superfamília *Copia* identificados no presente trabalho.

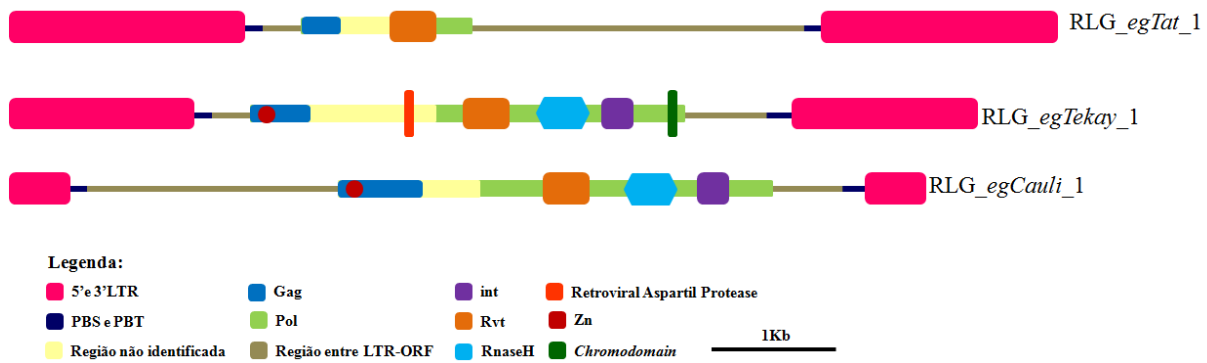


Figura 13: Estrutura hipotética de LTR-RTEs da superfamília *Gypsy* identificados no presente trabalho.

4.1.2. Número de cópias completas e distribuição genômica dos LTR-RTEs no genoma *draft* de *E. grandis*

A identidade entre a 5'LTR e a 3'LTR variou entre 86% e 99% (Tabelas 7 e 8), e demais regiões típicas de LTR-RTEs completos, tais como PBS (Figura 14), PPT e TSD estão detalhadas nas tabelas 7 e 8.

As sequências PBS e PPT guardam similaridade entre si e foram identificadas a partir de comparações com sequências consenso descritas por Wicker e Keller (2007) (Figura 14).

As sequências-referência dos LTR-RTEs foram utilizadas como sondas eletrônicas para a busca destes no genoma *draft* de *E. grandis* utilizando a ferramenta BLAT. Entre os LTRs-RTEs da superfamília *Copia* o com maior número de cópias foi o RTE RLC_egMax_1 com 290 e a que apresentou menor número foi RLC_egAle_2, com 38. Já nos LTRs-RTEs da superfamília *Gypsy*, o elemento que apresentou maior número de cópias foi RLG_egTekay_1 com 99 cópias e o com menor número foi RLG_egCauli_1, com 58.

A distribuição genômica dos LTR-RTEs avaliados foi dispersa em todos os pseudocromossomos de *E. grandis* (Figuras 15 e 16).

RLC_egAle_1	CA--TGGT-AT-CA-GAG-CT
Consenso Ale	CA--TGGT-AT-CA-GAG-CCTHBBCHHHY
RLC_egAle_2	CA--TGGT-AT-CA-GAG-CA
Consenso Ale	CA--TGGT-AT-CA-GAG-CCTHBBCHHHY
RLC_egMax_1	CA-ATTGGT-AT-CA-GAG-C-CT
Consenso Maximus	CA-ATTGGT-AT-CA-GAG-C-tNGtNtCCT
RLC_egBianca_1	AG--GGTT-TT-TC-GGTGA
Consenso Bianca	CA--CGTT-AT-CA-GCaCtHtCgctBccg
RLC_egIvana_1	AC-TTTCGT-GA-GA-AAA-TTG
Consenso Ivana	CA-ATTGGT-AT-CA-GAG-CCNCNNTNDNC

Figura 14: Comparação entre as sequências PBS das linhagens identificadas no trabalho com as respectivas sequências consenso das linhagens correspondentes (extraído de Wicker e Keller 2007).

Tabela 7: Caracterização das sequências de referência de LTR-RTEs da superfamília *Copia* utilizadas para análises de bancada.

Nome	Linhagem Evolutiva	Tamanho RTE (bp)	Tamanho LTR [#] (bp)	Id* LTRs	Tamanho ORF ¹	TSR	PBS	PPT	N ² genoma de <i>E. grandis</i>
RLC_egAle_1	<i>Ale</i>	5629	458/458	99%	4451	gttcg	catggtatcagagct	aatttgaggggggctg	48
RLC_egAle_2	<i>Ale</i>	5395	434/435	97%	4321	atggt	catggtatcagagca	ttgcgggggggatg	38
RLC_egMax_1	<i>Maximus</i>	9670	904/904	99%	5006	acatt	caattggtatcagagcct	caaaaagggggagattg	290
RLC_egBianca_1	<i>Bianca</i>	5008	424/424	99%	4099	catcat/catgct	agggttttcggtga	attgtcttcaagtgg	73
RLC_egGMR_1	<i>GMR</i>	11423	162/162	94%	4065	ctcca	cggaacaaacgttga	tgaaggtggggaga	96
RLC_egIvana_1	<i>Ivana</i>	4440	247/240	93%	3642	ttctg	acttctgtgagaaaattg	agttaagggaggga	68

*Id: identidade entre as LTRs;#: tamanho das LTRs apresentado - LTR 5'/LTR3'; 1: tamanho da ORF em pares de base; N².: número de cópias.

Tabela 8: Caracterização das sequências de referência de LTR-RTEs da superfamília *Gypsy* utilizadas para análises de bancada.

Nome	Linhagem Evolutiva	Tamanho RTE (bp)	Tamanho LTR [#] (bp)	Id* LTRs	Tamanho ORF ¹	TSR	PBS	PPT	N ² genoma de <i>E. grandis</i>
RLG_egTat_1	<i>Tat</i>	18300	2258/2199	94%	2967	gcact	tctttactgacttag	ttaattaagttttaatt	83
RLG_egTekay_1	<i>Tekay</i>	12159	656/656	97%	4551	tcaga/tcaca	atcatcaagcattcatgca	gagccatatcatata	99
RLG_egCauli_1	<i>Caulivirus</i>	11345	477/477	97%	4481	ataaa	catgtggtattagagc	agtccaagtgaggagaatg	58

*Id: identidade entre as LTRs;#: tamanho das LTRs apresentado - LTR 5'/LTR3'; 1: tamanho da ORF em pares de base; N².: número de cópias.

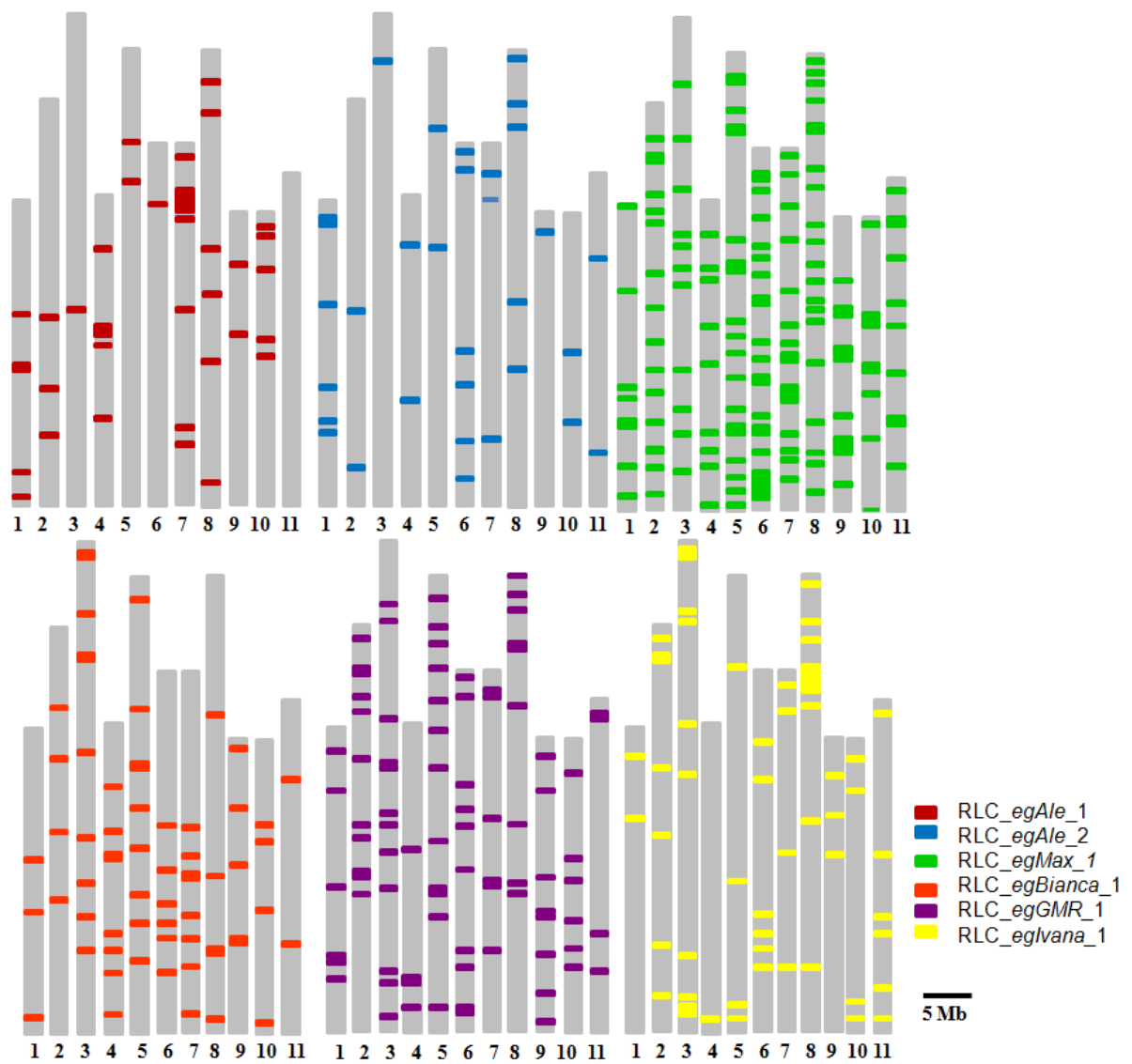


Figura 15: Distribuição genômica dos LTR-RTes pertencentes a superfamília *Copia* ao longo dos 11 pseudocromossomos de *E. grandis*.

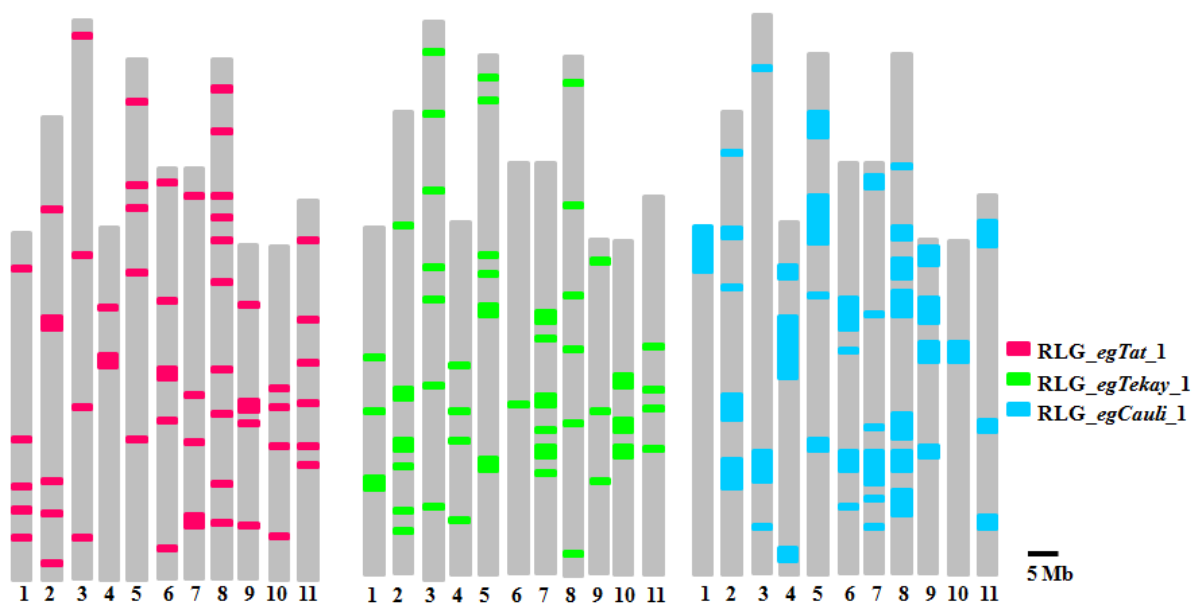
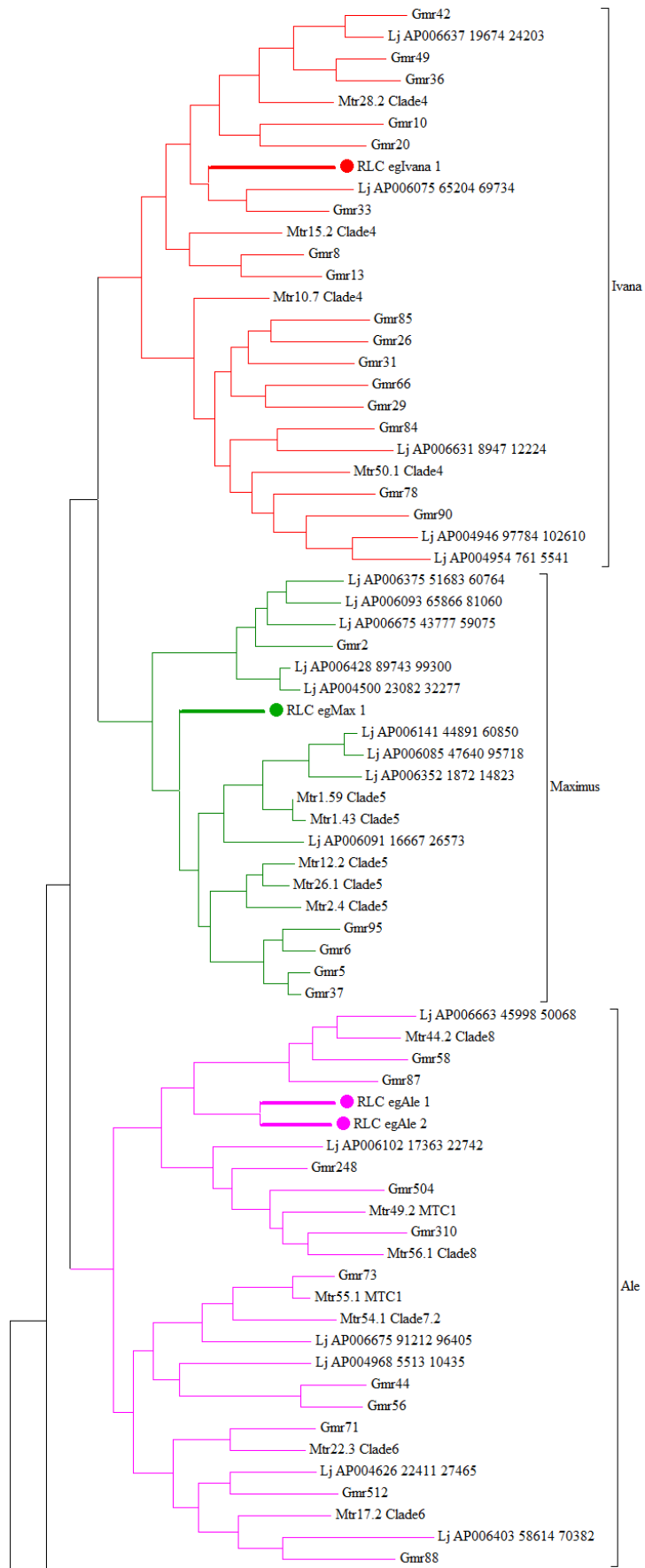


Figura 16: Distribuição genômica dos LTR-RTEs pertencentes a superfamília *Gypsy* ao longo dos 11 pseudocromossomos de *E. grandis*.

4.1.3. Nomenclatura e Classificação dos LTR-RTEs identificados

Os LTR-RTEs identificados foram classificados em linhagens evolutivas propostas para elementos das superfamílias *Copia* (Wicker e Keller 2007; Du et al. 2010) e *Gypsy* (Llorens et al. 2008; Domingues et al. 2012) presentes em vegetais. A partir das árvores filogenéticas geradas pode-se observar que os LTR-RTEs da superfamília *Copia* identificados pertencem a cinco linhagens evolutivas (Figura 17), dois deles correspondem a linhagem evolutiva *Ale* (RLC_egAle_1 e RLC_egAle_2), um a *Maximus* (RLC_egMax_1), um a linhagem *Bianca* (RLC_egBianca_1), um a *GMR* (RLC_egGMR_1) e um a linhagem *Ivana* (RLC_egIvana_1). Os LTR-RTEs da superfamília *Gypsy* foram classificados em três linhagens evolutivas (Figura 18): *Tat* (RLG_egTat_1), *Tekay* (RLG_egTekay_1) e *Caulivirus* (RLG_egCauli_1).

Com a identificação das linhagens evolutivas as quais se enquadram as famílias de LTR-RTEs foi adotada a nomenclatura sugerida por Domingues e colaboradores (2012), as quais passaram a receber o nome: RLC (*Copia*) e RLG (*Gypsy*), „*eg*’ para *E. grandis*, seguido pela indicação da linhagem evolutiva como „*Ale*’ e o número da família encontrada. Por exemplo: „RLC_egAle_1”.



Continuação na próxima página

Continuação

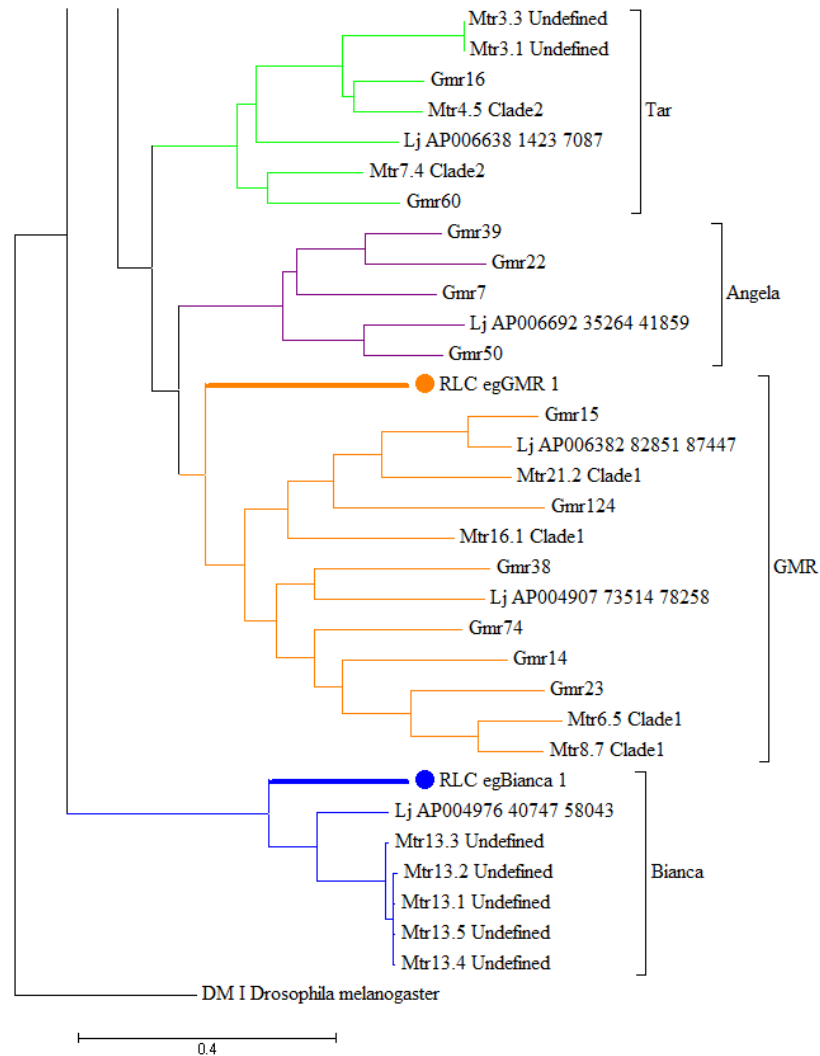


Figura 17: Relações filogenéticas a partir de sequências conservadas do domínio transcriptase reversa de LTR-RTEs pertencentes a superfamília *Copia*. A árvore foi enraizada utilizando-se o elemento *DM* de *Drosophila melanogaster* depositada no banco de dados RepBase como grupo externo. As diferentes cores nos ramos indicam as linhagens evolutivas, verde escuro - *Maximus*, rosa - *Ale*, verde claro - *Tar*, roxo - *Angela*, laranja - *GMR*, azul - *Bianca* e vermelho - *Ivana*.

Entre as famílias pertencentes à superfamília *Gypsy* foi identificado que uma das famílias não se enquadra em nenhuma linhagem evolutiva anteriormente descrita (Figura 18). A partir disso foram feitas avaliações da sequência utilizada para a construção da árvore filogenética nos bancos de dados GydB e RepBase, com estas avaliações pode-se concluir que esta família pode representar a inserção de um caulivírus integrado ao genoma de referência sequenciado.

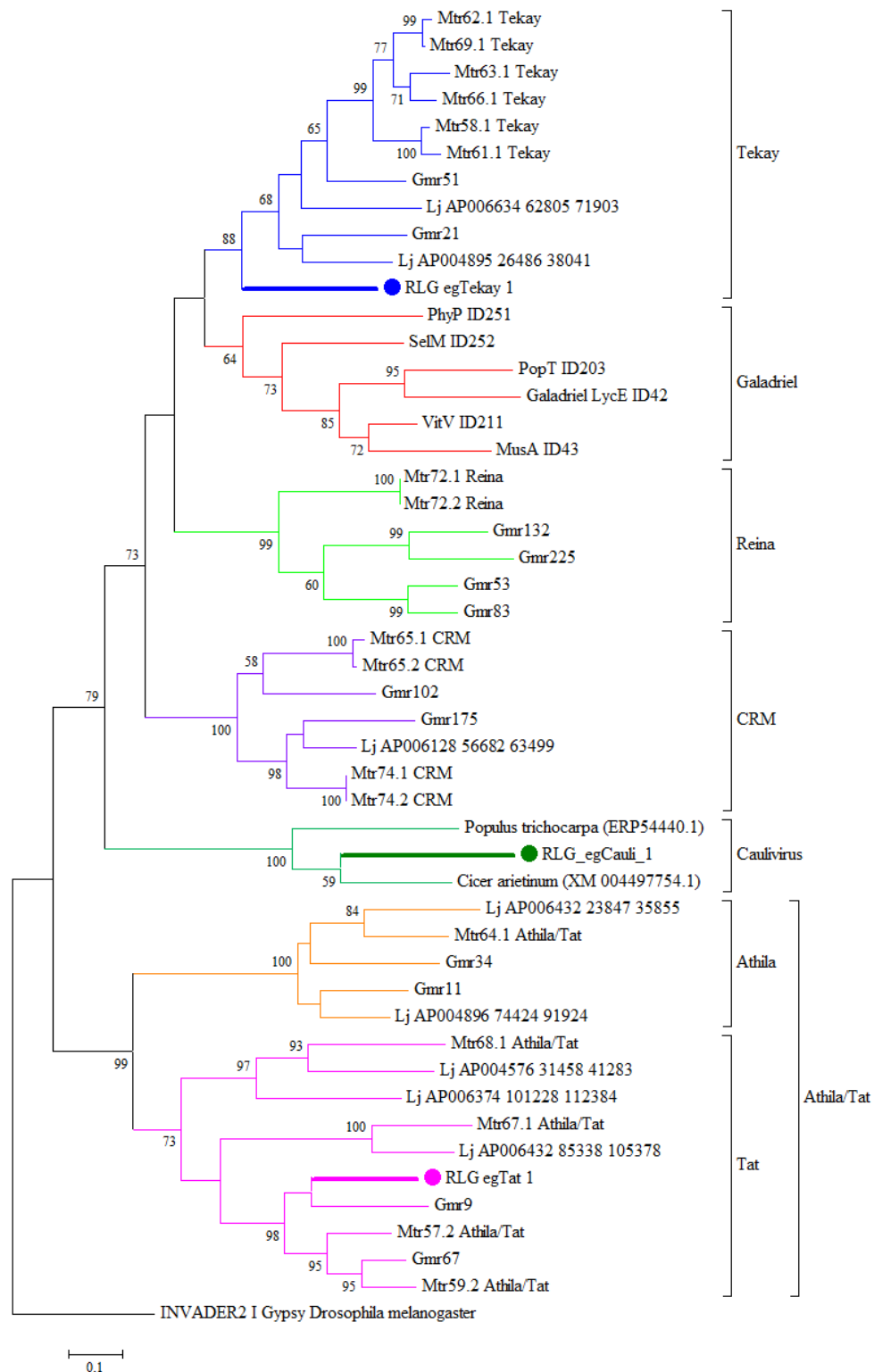


Figura 18: Relações filogenéticas a partir de sequências conservadas do domínio transcriptase reversa de LTR-RTEs pertencentes a superfamília *Gypsy* de diferentes linhagens evolutivas descritos por Du et al. (2010) e apresentadas por Neumann et al (2011), com adição dos acessos XM_004497754.1 e ERP54440.1. A árvore foi enraizada utilizando-se o elemento *Invader* de *Drosophila melanogaster*, depositada no banco de dados RepBase, como grupo externo. As diferentes cores dos ramos indicam as linhagens evolutivas: azul - *Tekay*, vermelho - *Galadriel*, verde claro - *Reina*, roxo - *CRM*, verde escuro - *Caulivirus*, laranja *Athila* e rosa – *Tat*.

4.1.4. Identificação de sequências expressas dos LTR-RTEs

Os LTR-RTEs identificados foram também utilizados como iscas para busca de sequências expressas (ESTs Sanger) em cinco espécies de eucalipto (*Eucalyptus camaldulensis*, *E. globulus*, *E. grandis*, *E. pellita* e *E. urophylla*) disponibilizadas na plataforma EUCANEXT (<http://bioinfo03.ibi.unicamp.br/eucalyptusdb/>) (Nascimento et al. 2011; Salazar et al. 2013).

Após a normalização dos dados, a superfamília *Copia* foi a mais representativa entre os ESTs em relação a *Gypsy* (Tabela 9). Entre as famílias pertencentes a superfamília *Copia*, *RLC_egBianca_1* foi identificada em ESTs de todas as espécies avaliadas, enquanto ESTs relacionados a *RLC_egIvana_1* não foram identificados em nenhuma das espécies. A espécie com maior atividade transcricional de LTR-RTEs foi *E. gunnii*.

Tabela 9: Proporção a cada 100.000 ESTs Sanger de cada LTR-RTE nas diferentes espécies de eucalipto disponíveis no banco de dados EUCANEXT (Nascimento et al. 2011; Salazar et al. 2013). Foram selecionados ESTs que atendiam as regras de Wicker et al. (2007).

Superfamília	Família	<i>E. camaldulensis</i>	<i>E. globulus</i>	<i>E. grandis</i>	<i>E. gunni</i>	<i>E. pellita</i>	<i>E. urophylla</i>	<i>E. grandis</i> Illumina	<i>E. globulus</i> 454
								RNA-seq	RNA-seq
<i>Copia</i>	RLC_egAle_1	0	0	0	15.1	0	0	X	X
	RLC_egAle_2	0	0	0	20.16	0	0	X	X
	RLC_egMax_1	1.73	0	0	30.24	0	27	-	X
	RLC_egBianca_1	3.5	0	14.1	20.16	33.82	13.45	X	X
	RLC_egGMR_1	10.4	65.63	230.2	5.04	0	0	X	X
	RLC_egIvana_1	0	0	0	0	0	0	-	X
<i>Gypsy</i>	RLG_egTat_1	1.73	0	0	20.16	0	13.45	X	X
	RLG_egTekay_1	0	0	0	5.04	0	0	X	X
	RLG_egCauli_1	5.2	0	0	5.04	0	0	-	X

X: indica que foi encontrado LTR-RTE relacionado ESTs.

4.2. Análises IRAP

4.2.1. Análises de polimorfismos

Os testes realizados utilizando-se o protocolo de Kalendar e Schulman (2006) não apresentaram sucesso, apesar das diversas alterações de protocolo realizadas. Em razão disso, passou-se a adotar o protocolo proposto por Smýkal (2006), com algumas alterações. Os fragmentos puderam ser observados em gel de agarose a 1,8%, o qual apresentou melhor qualidade de visualização (Figuras 19 a 22).

A partir dos *primers* desenhados nas 5' e 3' LTR de um mesmo LTR-RTE, pode-se observar uma grande diferença no padrão dos fragmentos IRAP obtidos (Figura 19 a 22), assim como no número de fragmentos gerados e de polimorfismos (Tabela 10). Apenas os *primers* desenhados a partir da LTR do RLC_egGMR_1 apresentaram uma proximidade no número de fragmentos gerados e polimorfismos (Tabela 10).

A análise dos 16 *primers* foram realizadas em 50 indivíduos, de cinco espécies do gênero *Eucalyptus* (10 indivíduos por espécie) (Tabela 10).

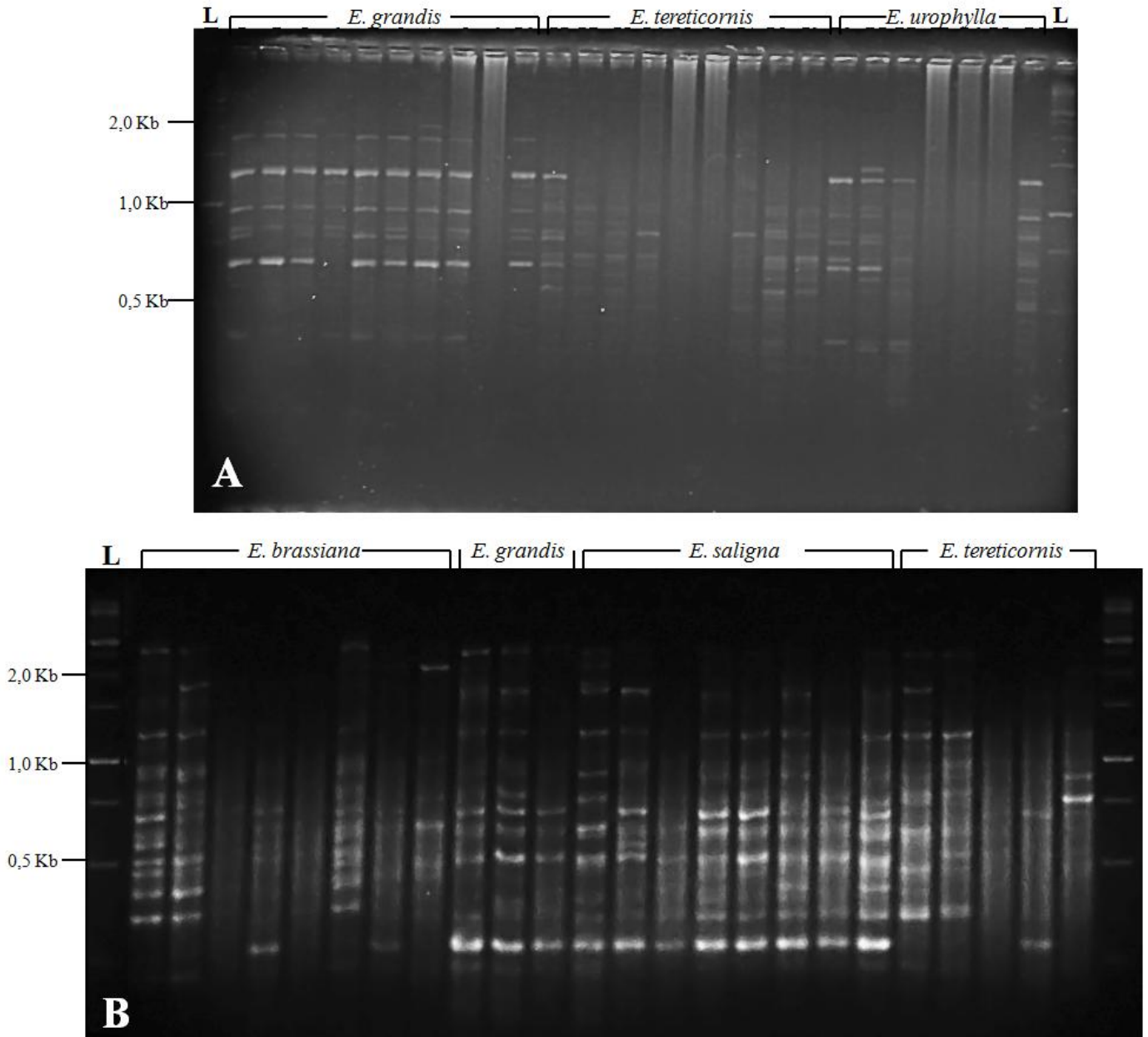


Figura 19: Análise IRAP em gel de agarose 1,8%: (A) utilizando o *primer* EgAleLTR-3; (B) utilizando *primer* EgBiancaLTR-3. L: GeneRuler 1kb DNA Ladder (Fermentas – código SM0311)

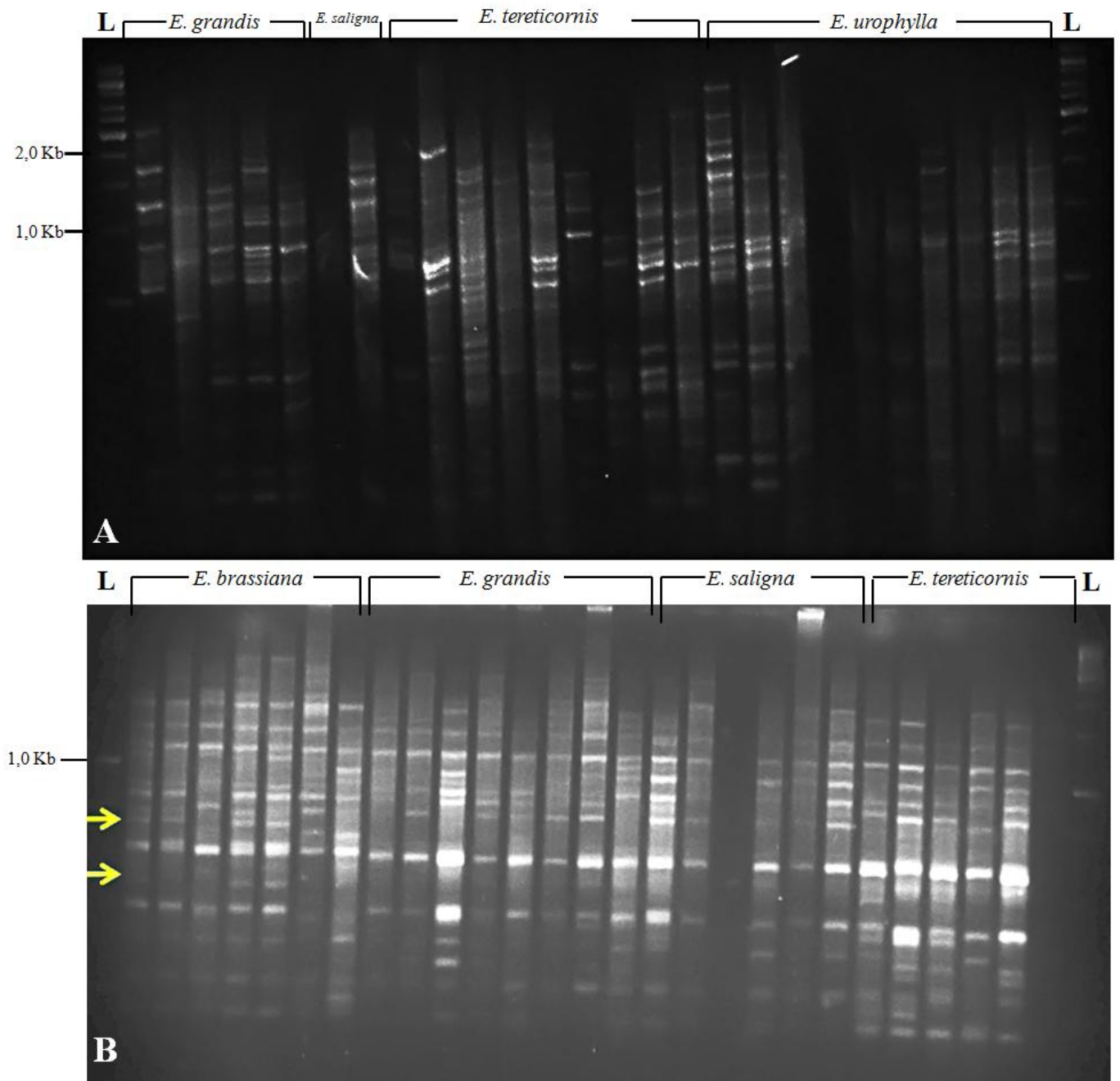


Figura 20: Análise IRAP em Gel de agarose 1,8%: (A) utilizando o *primer* EgAle2LTR-3; (B) utilizando *primer* EgAle2LTR-5. L: GeneRuler 1kb DNA Ladder (Fermentas – código SM0311). As setas (amarelas) indicam locais em que foram observados polimorfismos.

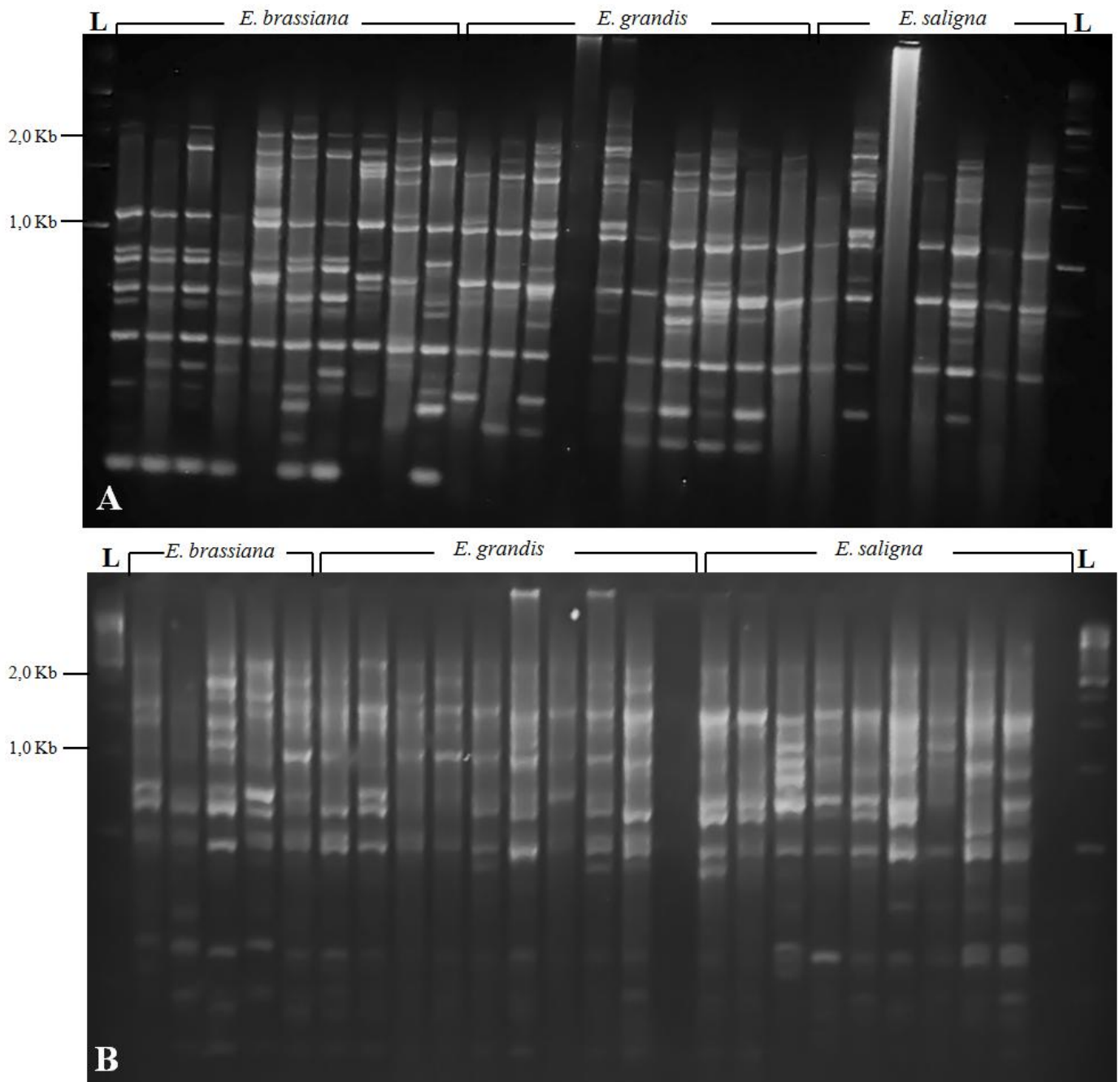


Figura 21: Análise IRAP em gel de agarose 1,8%: (A) utilizando o *primer* EgGMRLTR-5; (B) utilizando *primer* EgMaxLTR-3. L: GeneRuler 1kb DNA Ladder (Fermentas – código SM0311).

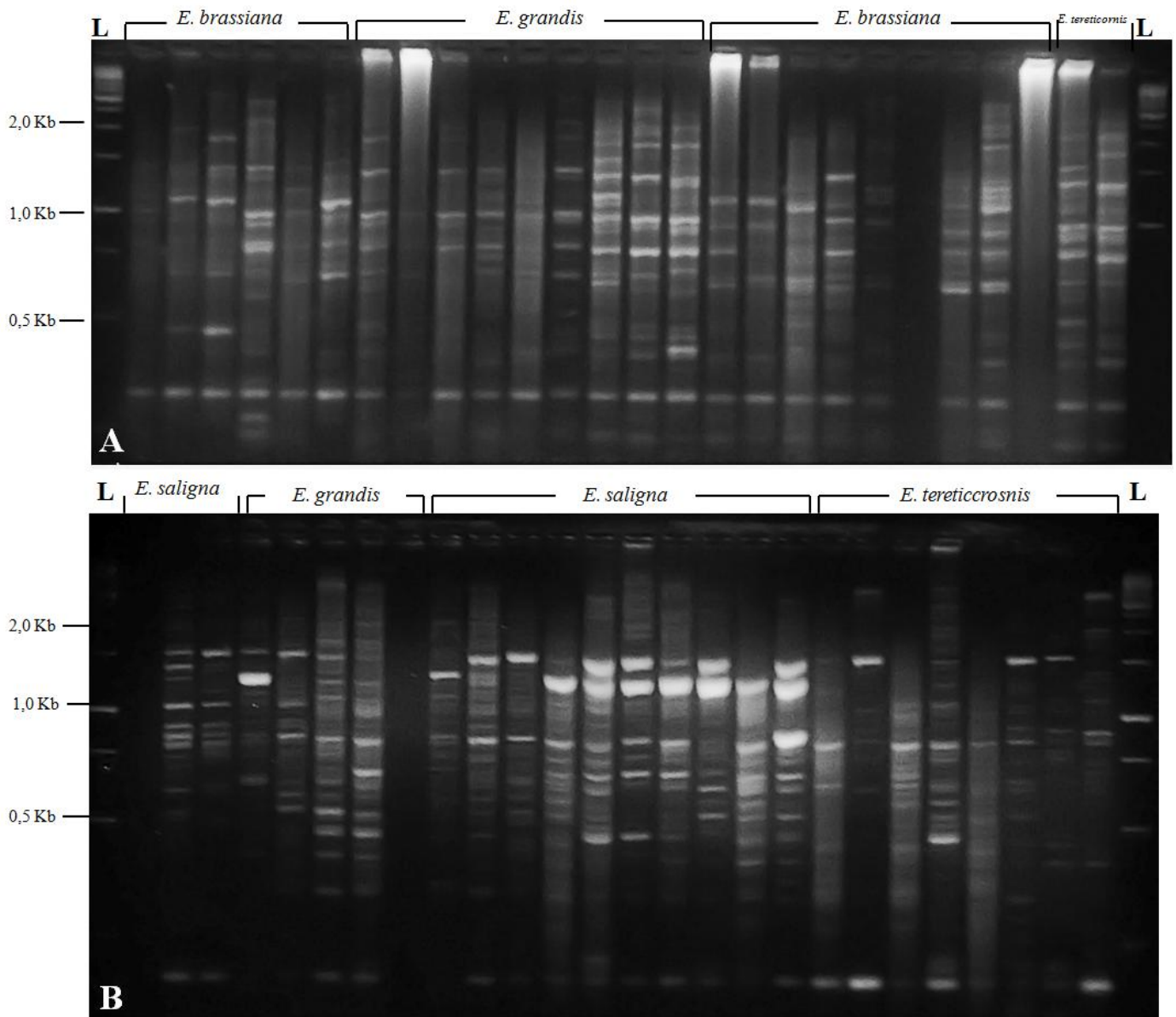


Figura 22: Análise IRAP em gel de agarose 1,8%: (A) utilizando o *primer* EgTatLTR-3; (B) utilizando *primer* EgTekayLTR-3. L: GeneRuler 1kb DNA Ladder (Fermentas – código SM0311).

Os *primers* EgIvanaLTR-5 e EgCauliLTR-5 foram eliminados das análises, pois não apresentaram padrão de fragmentos reprodutíveis, mesmo após adaptações no protocolo.

Tabela 10: Total de fragmentos e polimorfismos obtidos com *primers* IRAP.

<i>Primers</i>	Espécies	Número de fragmentos						Número de fragmentos polimórficos					
		<i>E. brassiana</i>	<i>E. grandis</i>	<i>E. saligna</i>	<i>E. tereticornis</i>	<i>E. urophylla</i>	Média ¹	<i>E. brassiana</i>	<i>E. grandis</i>	<i>E. saligna</i>	<i>E. tereticornis</i>	<i>E. urophylla</i>	Média ¹
EgAleLTR-3		16	12	16	16	16	15,2	2	3	0	0	1	1,2
EgAleLTR-5		16	22	24	22	18	20,4	2	9	6	4	3	4,8
EgAle2LTR-3		16	20	20	20	20	19,2	2	4	2	2	3	2,6
EgAle2LTR-5		17	12	15	20	15	15,8	4	5	3	3	4	3,8
EgMaxLTR-3		18	17	20	11	10	15,2	4	3	6	5	2	4
EgMaxLTR-5		15	22	22	16	12	17,4	4	5	7	3	2	4,2
EgBiancaLTR-3		24	25	22	20	25	23,2	7	5	5	1	6	4,8
EgBiancaLTR-5		21	23	22	19	22	21,4	5	8	6	4	3	5,2
EgGMRLTR-3		21	24	21	24	24	22,8	6	4	10	8	3	6,2
EgGMRLTR-5		21	20	20	16	18	19	8	6	0	3	3	4
EgIvanaLTR-3		24	16	25	29	18	22,4	7	6	8	3	0	4,8
EgTatLTR-3		15	19	20	17	20	18,2	4	4	4	6	0	3,6
EgTatLTR-5		20	22	24	20	17	20,6	7	3	5	2	3	4
EgTekayLTR-3		17	18	19	15	17	17,2	3	3	4	1	2	2,6
EgTekayLTR-5		24	16	22	22	18	20,4	3	4	4	4	2	3,4
EgCauliLTR-3		17	16	14	16	16	15,8	5	4	3	6	7	5
Média²		18,87	19	20,37	18,93	17,87		4,56	4,75	4,56	3,43	2,75	

¹: Média do número de fragmentos por *primer*, separados por espécies; ²: Média do número de fragmentos por espécie em relação aos 18 marcadores IRAP.

Um total de 1521 fragmentos foi gerado a partir dos 16 *primers* IRAP, dos quais 321 foram polimórficos. Entre os *primers* desenhados para os LTR-RTEs da superfamília *Copia*, os que possuíram maior número de fragmentos foram EgBiancaLTR-3 com média de 23,2 fragmentos por espécie, e os *primers* com menor número de fragmentos foram EgAleLTR-3 e EgMaxLTR-3, ambos com média de 15,2 fragmentos por espécie de eucalipto. O *primer* mais polimórfico foi EgGMRLTR-3, com média de 6,2 fragmentos polimórficos por espécie analisada. A espécie que apresentou o maior número de polimorfismos foi *E. saligna* (Tabela 10).

Dos *primers* desenhados para os LTR-RTEs da superfamília *Gypsy*, EgTatLTR-5 apresentou o maior número de fragmentos, com média de 22,4 fragmentos por espécie de eucalipto, dentre estas *E. saligna* com maior número de fragmentos, e o *primer* com maior número de polimorfismos foi EgTatLTR-3, com média de 3,6 fragmentos por espécie de eucalipto, no entanto a espécie *E. urophylla* não apresentou nenhum fragmento polimórfico para este *primer*. Já EgTekayLTR-3 foi o *primer* com menor número de polimorfismos, com média de 2 fragmentos por espécie, e todas as espécies possuíam fragmentos polimórficos (Tabela 10).

A identificação de polimorfismos possibilitou a caracterização de inserções espécie(s)-específica(s), as quais somente foram selecionadas se houvesse a mesma variante em no máximo duas espécies.

Dos *primers* desenhados nos LTR-RTEs da superfamília *Copia*, EgAle2LTR-3 foi o com maior número de polimorfismos. Já EgGMRLTR-5 apresentou fragmentos polimórficos de mesmo tamanho para duas espécies diferentes, *E. tereticornis* e *E. saligna* (Tabela 10). Dos cinco *primers* desenhados para os elementos da superfamília *Gypsy*, somente EgCauliLTR-3 produziu fragmentos a serem analisados para desenvolvimento de marcadores espécie específicos em duas espécies de eucalipto diferentes, *E. brassiana* e *E. urophylla* (Tabela 10).

Com isso, a partir das análises detalhadas de cada *primer* IRAP, foram selecionados 16 fragmentos para clonagem e sequenciamento, visando o desenvolvimento de marcadores espécie-específicos (Tabela 11).

Tabela 11: Fragmentos polimórficos selecionados entre os *primers* desenhados nos LTR-RTEs das superfamílias *Copia* e *Gypsy*.

<i>Primer</i>	Tamanho fragmento (bp)	Espécie
EgAle2LTR-3	1100	<i>E. grandis</i>
	950	<i>E. grandis</i>
	1150	<i>E. urophylla</i>
EgAle2LTR-5	700	<i>E. brassiana</i>
	1250	<i>E. saligna</i>
EgMaxLTR-3	750	<i>E. saligna</i>
EgMaxLTR-5	470	<i>E. saligna</i>
EgBiancaLTR-5	680	<i>E. grandis</i>
EgGMRLTR-3	630	<i>E. saligna</i>
EgGMRLTR-5	800	<i>E. tereticornis</i>
	980	<i>E. tereticornis</i> e <i>E. saligna</i>
EgTatLTR-5	760	<i>E. grandis</i>
	780	<i>E. saligna</i>
EgTekayLTR-5	630	<i>E. saligna</i>
EgCauliLTR-3	650	<i>E. brassiana</i>
	830	<i>E. urophylla</i>

A partir da seleção dos fragmentos, três bandas de mesmo tamanho para cada espécie foram excisadas e purificadas, e para cada um destes fragmentos foram selecionadas três colônias para sequenciamento, resultando em nove sequências para análise por fragmento polimórfico.

4.3. Análises REMAP

4.3.1. Amplificação de marcadores REMAP

Os 18 *primers* desenhados para a técnica de IRAP foram testados conjuntamente com 10 *primers* microsátélites. Destas combinações duas apresentam resultados passíveis de análise - Micro 1 e 12 (Tabela 2), que foram utilizados nas reações de PCR baseadas no protocolo de Smýkal (2006), com adaptações. Os perfis de amplificação são mostrados nas figuras 23 a 26.

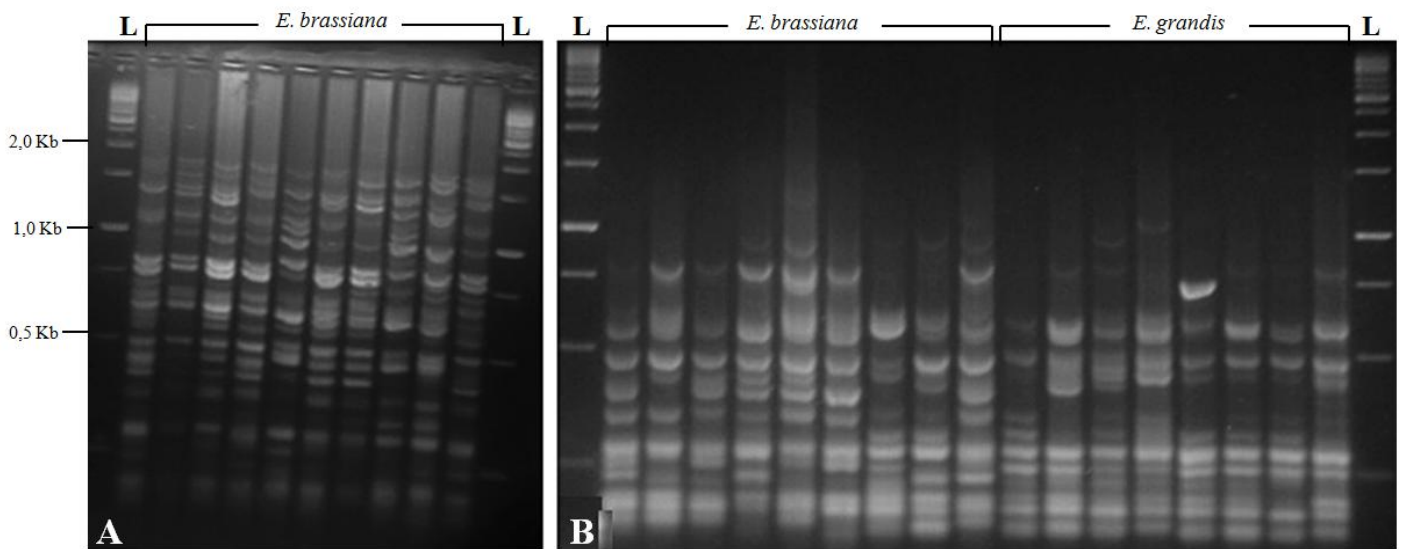


Figura 23: Análises REMAP em gel de agarose 1,8% corado com brometo de etídio. A: Combinação dos *primers* Micro1 e EgAle2LTR-3. B: Combinação dos *primers* Micro1 e EgTatLTR-5. L: GeneRuler 1kb DNA Ladder (Fermentas – código SM0311).

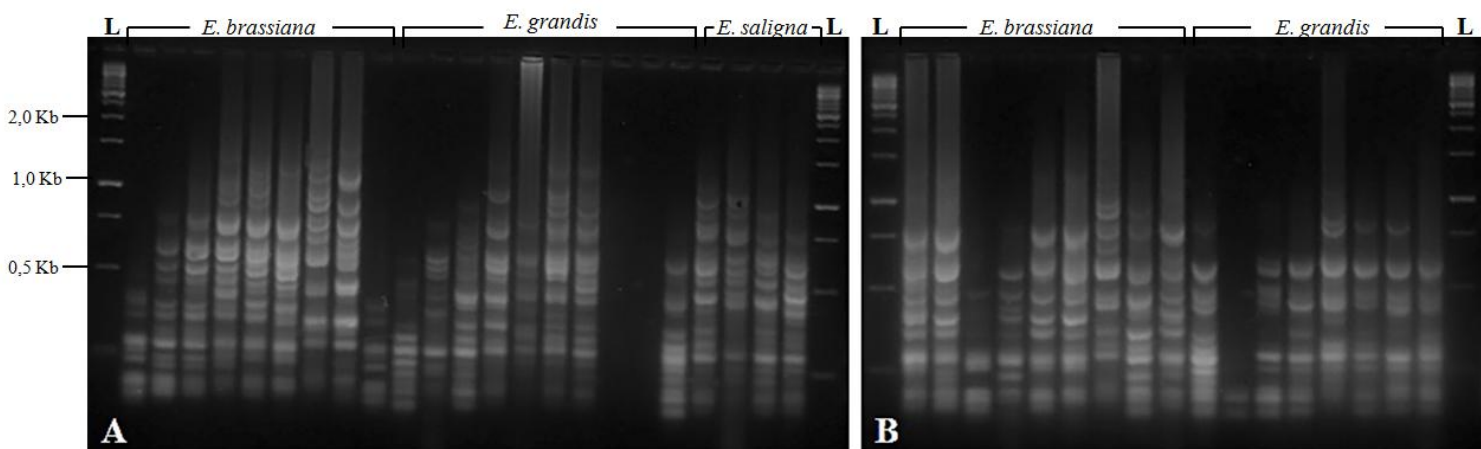


Figura 24: Análises REMAP em gel de agarose 1,8% corado com brometo de etídio. A: Combinação dos *primers* Micro1 e EgCauliLTR-5, com indivíduos *E. brassiana*, *E. grandis* e *E. saligna*; B: Combinação dos *primers* Micro1 e EgCauliLTR-3, com indivíduos *E. brassiana* e *E. grandis*. L: GeneRuler 1kb DNA Ladder (Fermentas – código SM0311).

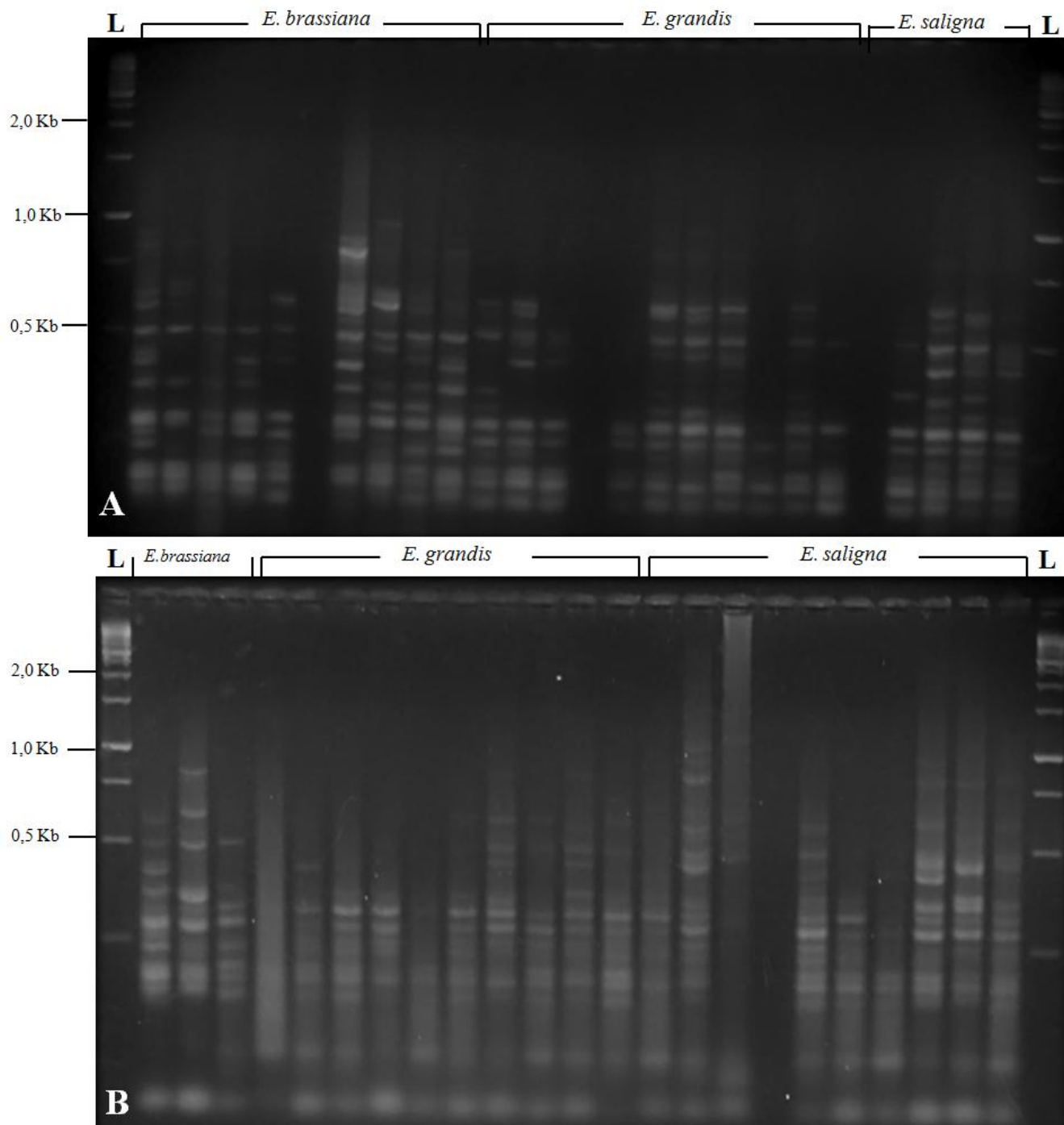


Figura 25: Análises REMAP em gel de agarose 1,8% corado com brometo de etídio. A: Combinação dos *primers* Micro1 e EgAleLTR-3. B: Combinação dos *primers* Micro1 e EgBiancaLTR-3. L: GeneRuler 1kb DNA *Ladder* (Fermentas – código SM0311).

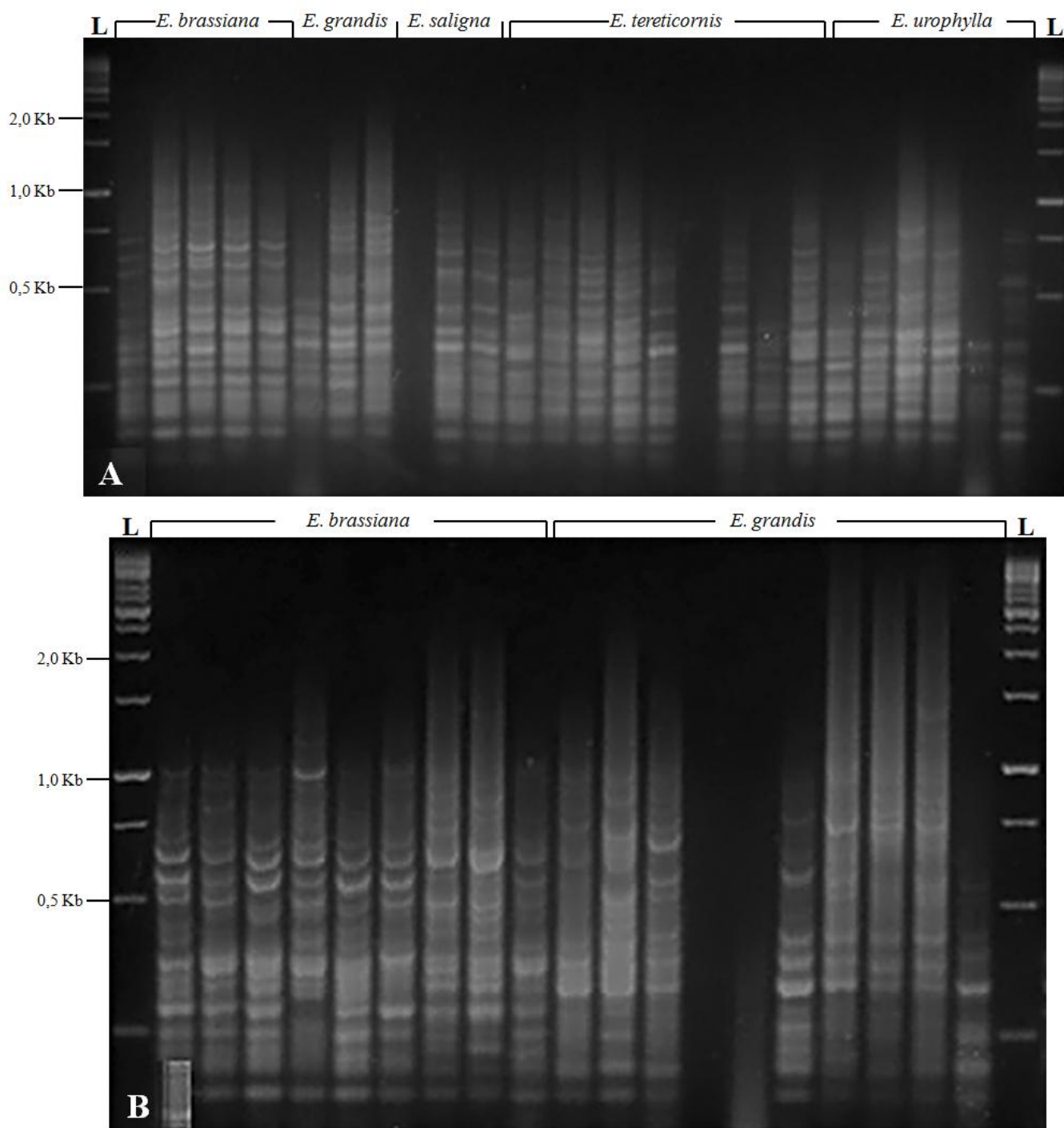


Figura 26: Análises REMAP em gel de agarose 1,8% corado com brometo de etídio. A: Combinação dos *primers* Micro12 e EgMaxLTR-5. B: Combinação dos *primers* Micro12 e EgTatLTR-5. L: GeneRuler 1kb DNA *Ladder* (Fermentas – código SM0311).

Mesmo após adaptações de protocolo, para o *primer* microssatélite Micro 1 não se obteve sinais satisfatórios de amplificação com os *primers* EgAleLTR-3, EgAle2LTR-3, EgAle2LTR-5 e EgCauliLTR-5, que foram excluídos. Assim, as análises foram feitas por meio da combinação de 15 *primers* IRAP com este *primer* de microssatélite. Do mesmo modo a amplificação com o iniciador Micro 12, em conjunto com os *primers* IRAP, apresentou problemas com EgAle2LTR-3, EgMaxLTR-3, EgBiancaLTR-5, EgTekayLTR-3, que foram excluídos das análises. Dessa forma, 13 *primers* IRAP foram avaliados com este *primer* microssatélite.

A combinação de *primers* IRAP com os *primers* Micro 1 e 12 resultou em grande parte em fragmentos com tamanho menor que 1kb, com exceção do *primer* EgAle2LTR-3 (figura 23A). 1116 e 794 fragmentos foram geradas a partir das combinação com os *primers* Micro 1 e 12 respectivamente, dos quais 298 e 194 foram observados como polimórficos (Tabela 12).

Entre as combinações com o *primer* Micro 1, EgAle2LTR-3 e EgCauliLTR-5 apresentaram o maior número de fragmentos, com média de 20,6 e 15,4 fragmentos por espécie de eucalipto, respectivamente (Tabela 12). Já os que apresentaram maior números de polimorfismos foram EgAle2LTR-3 e EgTatLTR-5, com 5,2 e 5,4 fragmentos em média por espécie de eucalipto, respectivamente (Tabela 12).

Dentre as combinações de *primers* avaliadas para elementos da superfamília *Copia* com o *primer* Micro 12, EgAleLTR-3 teve o maior número de fragmentos, com médias de 18 fragmentos por espécie de eucalipto, respectivamente. (Tabela 12), e o *primer* mais polimórfico foi EgAleLTR-5, com média de 5,2 fragmentos polimórficos por espécie de eucalipto (Tabela 12). O menor número de fragmentos e bandas polimórficas foi observado na combinação com os *primers* baseados na superfamília *Gypsy*, o *primer* EgCauliLTR-3 foi o que teve maior número de bandas entre os da superfamília *Gypsy*, com média de 13,8.

Tabela 12: Total de fragmentos e polimorfismos obtidos com os *primer* IRAP em conjunto com os *primers* Micro 1 e 12.

Micro	Número de Fragmentos		Número de Fragmentos Polimórficos											
	Primers	Sp*	<i>E. brassiana</i>	<i>E. grandis</i>	<i>E. saligna</i>	<i>E. tereticornis</i>	<i>E. urophylla</i>	Média ¹	<i>E. brassiana</i>	<i>E. grandis</i>	<i>E. saligna</i>	<i>E. tereticornis</i>	<i>E. urophylla</i>	Média ¹
Micro1	EgAleLTR-3		17	12	13	14	12	13,6	8	4	2	1	1	3,2
	EgAle2LTR-3		23	19	20	20	21	20,6	8	5	4	6	3	5,2
	EgAle2LTR-5		23	18	19	12	14	17,2	6	3	5	2	2	3,6
	EgMaxLTR3		16	20	20	21	15	18,4	4	5	4	4	3	4
	EgMaxLTR-5		20	18	16	16	17	17,4	6	5	2	3	1	3,4
	EgBiancaLTR-3		19	13	13	10	12	13,4	7	5	5	3	4	4,8
	EgBiancaLTR-5		15	13	13	10	12	12,6	4	4	1	0	2	2,2
	EgGMRLTR-3		16	15	15	14	15	15	2	2	7	6	4	4,2
	EgGMRLTR-5		12	11	13	10	7	10,6	2	1	2	0	0	1
	EgIvanaLTR-3		11	11	11	12	12	11,4	3	3	2	6	5	3,8
	EgIvanaLTR-5		14	15	15	12	11	13,4	4	4	3	5	3	3,8
	EgTatLTR-3		20	19	15	14	12	16	6	4	6	4	4	4,8
	EgTatLTR-5		15	17	14	14	12	14,4	3	7	8	8	1	5,4
	EgCauliLTR-3		17	12	15	15	10	13,8	6	5	7	6	2	5,2
EgCauliLTR-5		16	19	15	15	12	15,4	6	5	5	5	3	4,8	
Micro12	EgAleLTR-3		21	18	20	16	15	18	3	4	10	2	4	4,6
	EgAleLTR-5		10	10	12	10	8	10	4	5	3	2	4	3,6
	EgAle2LTR-5		16	12	10	10	10	11,6	3	4	3	2	0	2,4
	EgMaxLTR-5		17	12	13	18	10	14	4	2	5	3	4	3,6
	EgBiancaLTR-5		14	15	15	15	16	15	3	1	2	0	1	1,4
	EgGMRLTR-3		11	10	13	12	11	11,4	5	1	4	2	2	2,8
	EgGMRLTR-5		15	10	12	10	12	11,8	3	1	3	4	0	2,2
	EgIvanaLTR-5		14	14	12	14	11	13	4	5	2	6	2	3,8
	EgAleLTR-3		10	10	10	10	8	9,6	3	5	2	4	1	3
	EgAleLTR-5		12	12	11	11	7	10,6	2	5	4	1	0	2,4
	EgTekayLTR-5		13	12	13	14	7	11,8	3	1	3	3	0	2
	EgCauliLTR-3		10	11	14	10	10	11	6	7	7	1	2	4,6
	EgCauliLTR-5		7	11	11	14	12	11	3	4	5	3	2	3,4
		Média ²		15,1	13,89	14,03	13,32	11,82	-	4,10	3,89	4,14	3,28	2,14

*Sp: Espécie de eucalipto; ¹: Média do número de fragmentos por *primer* em relação a cinco espécies; ²: Média do número de fragmentos por espécie em relação aos 18 *primers* IRAP.

As análises das combinações de *primers* REMAP possibilitaram a seleção 14 fragmentos para clonagem e sequenciamento, visando o desenvolvimento de marcadores espécie específicos (Tabela 13).

Tabela 13: Fragmentos polimórficos selecionados a partir da combinação dos *primers* Micro1 e 12 com os *primers* IRAP.

Microssatélite	Primers IRAP	Tamanho (bp)	Espécie(s)
Micro1	EgAleLTR-3	900	<i>E. brassiana</i>
	EgAle2LTR-3	580	<i>E. brassiana</i>
	EgAle2LTR-5	740	<i>E. brassiana</i>
	EgMaxLTR-3	650	<i>E. saligna</i>
		580	<i>E. urophylla</i>
	EgGMRLTR-3	1050	<i>E. saligna</i>
	EgIvanaLTR-3	420	<i>E. tereticornis</i>
	EgTatLTR-3	740	<i>E. brassiana</i> e <i>E. grandis</i>
	EgCauliLTR-5	380	<i>E. tereticornis</i>
510		<i>E. brassiana</i> e <i>E. saligna</i>	
Micro12	EgAleLTR-3	680	<i>E. brassiana</i>
	EgMaxLTR-5	440	<i>E. tereticornis</i> e <i>E. urophylla</i>
	EgGMRLTR-5	490	<i>E. brassiana</i> e <i>E. tereticornis</i>
	EgCauliLTR-5	500	<i>E. brassiana</i>

O maior número de fragmentos polimórficos foi identificado nas análises com o Micro 1, dos quais três amplificaram em duas espécies. Já entre os fragmentos polimórficos identificados com o Micro 12, somente os *primers* EgAleLTR-3 e EgIvanaLTR-5, apresentaram fragmentos polimórficos com uma relação espécie específica (Tabela 13).

4.4. Similaridade Genética – IRAP e REMAP

Com o intuito de inferir a similaridade genética entre as cinco espécies de eucalipto avaliadas por meio das técnicas IRAP e REMAP foi construído um dendrograma UPGMA (Figura 27).

A topologia do dendrograma dos dados IRAP e REMAP indicou que as cinco espécies de eucalipto apresentaram grande distância genética ($\sim 0,90$; Figura 27), e que *E. tereticornis* e *E. urophylla* foram as com menor distância genética ($\sim 0,60$; Figura 27).

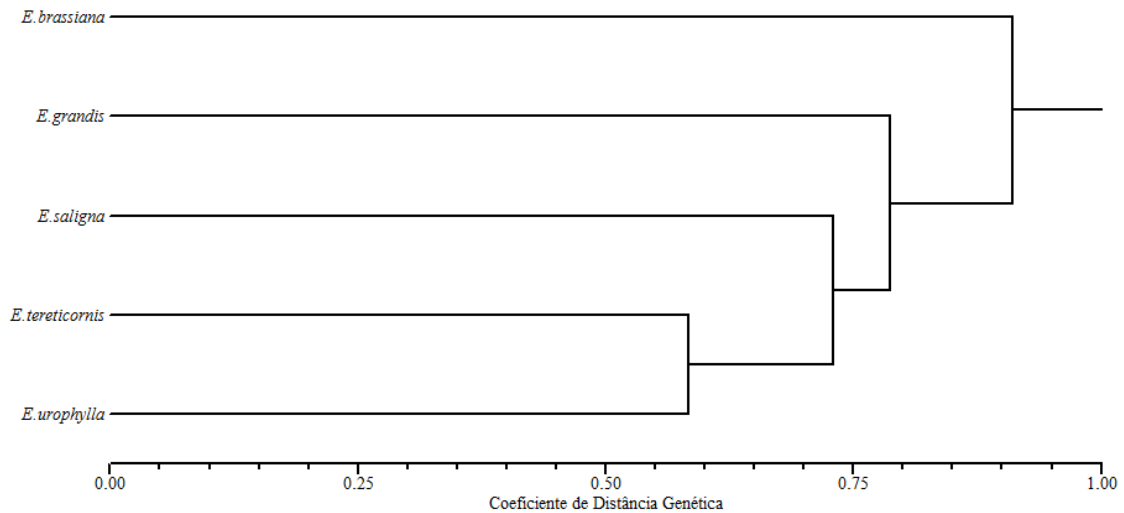


Figura 27: Dendrograma UPGMA de distância genética com os dados obtidos por meio das técnicas IRAP e REMAP.

4.5. Análises dos fragmentos polimórficos IRAP e REMAP

Das 47 sequências obtidas pelo sequenciamento dos 16 fragmentos polimórficos gerados a partir da técnica IRAP, 41 apresentaram identidade com alguma região do genoma *draft* de *E. grandis* (Figura 28).

Além da validação citada, as sequências foram alinhadas com os LTR-RTEs das quais foram provenientes. Com isso, duas sequências foram excluídas, por não apresentarem relação com qualquer elemento selecionado neste trabalho.

Posteriormente as sequências foram agrupadas e classificadas quanto a sua especificidade para desenvolvimento de marcadores RBIP.

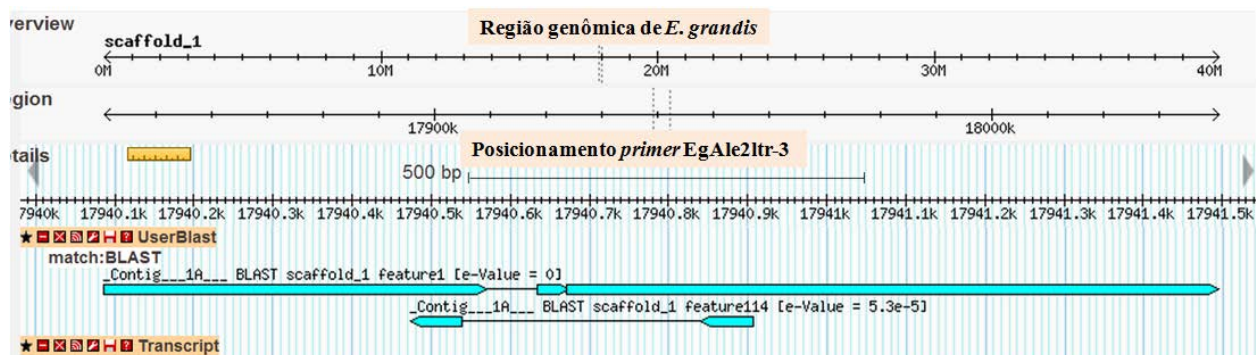


Figura 28: Região polimórfica identificada a partir do *primer* CoEu4-3 da amostra de *E. grandis* no genoma *draft* de *E. grandis*.

A partir das combinações entre o *primer* Micro 1 e os *primers* IRAP foi obtido um total de 64 sequências, das quais 24 foram selecionadas após seleções baseadas a busca de similaridade entre o *primer* e o LTR-RTE correspondente.

Das 38 sequências obtidas a partir das combinações entre o Micro 12 e os *primers* IRAP, foram selecionadas 14 com base nos critérios colocados acima.

Posteriormente, as sequências foram agrupadas de acordo com o intervalo observado para posterior classificação quanto a ser espécie ou genótipo específica, cujas análises seguiram para confecção dos marcadores RBIP.

4.6. Desenvolvimento e análise de marcadores RBIP

Um total de 22 marcadores polimórficos, provenientes das análises IRAP e REMAP, foram convertidos em marcadores RBIP (Tabela 14).

Dos 22 marcadores desenvolvidos, 11 foram provenientes de IRAP e 11 de REMAP. Vinte marcadores são provenientes de elementos da superfamília *Copia* e quatro são provenientes de elemento da superfamília *Gypsy*. Um resumo dos elementos que apresentaram polimorfismos utilizados para desenvolvimento de marcadores RBIP é apresentada na Tabela 14.

Tabela 14: Marcadores obtidos a partir dos polimorfismos encontrados através das técnicas IRAP e REMAP.

Técnica Original	Superfamília	RTE relacionado	Nome	Tamanho (bp)
IRAP	<i>Copia</i>	RLC_egGMR_1	SGMR1	460
		RLC_egAle_2	BAle2.1	698
		RLC_egAle_2	BAle2.2	737
		RLC_egMax_1	SMax1	447
		RLC_egGMR_1	TGMR1	838
		RLC_egAle_2	UAle2.1	571
		RLC_egAle_2	EAle2.1	665
		RLC_egAle_2	EAle2.2	534
		RLC_egAle_2	EAle2.3	733
	<i>Gypsy</i>	RLG_egTat_1	ETat1	416
RLG_egCauli_1		BCauli1	440	
REMAP - Micro 1	<i>Copia</i>	RLC_egAle_2	BAle2.3	-
		RLC_egAle_2	BAle2.4	-
	<i>Gypsy</i>	RLG_egCauli_1	TCauli1	-
		RLG_egCauli_1	SCauli1	*
		RLG_egCauli_1	SCauli2	-
Micro 12	<i>Copia</i>	RLC_egGMR_1	BGMR1	*
		RLC_egAle_1	BAle2.5	-
		RLC_egAle_1	BAle2.6	-
		RLC_egMax_1	UMax1	*
		RLC_egMax_1	TMax1	*
		RLC_egIvana_1	BAle2.7	-

* Fragmentos inespecíficos; - não apresentaram fragmentos.

Para nove dos 22 marcadores desenvolvidos não foram obtidos sinais de amplificação, mesmo após mudanças de protocolo de amplificação: BAle2.1, BAle2.2, EAle2.3, TCauli1, BAle2.4, SCauli2, BAle2.5, BAle2.6 e BAle2.7. Sete *primers* apresentaram inespecificidade na geração de fragmentos: SGMR1, UAle2.1, EAle2.2, ETat1, BAle2.3, SCauli1, BGMR1 e TMax1. Já nas análises com os *primers* BCauli1 e SMax1 os fragmentos obtidos foram de mesmo tamanho em todas as espécies de eucalipto, sem polimorfismo observável.

As famílias RLG_egMax_1, RLC_egGMR_1 e RLC_egAle_2 apresentaram *primers* candidatos para o desenvolvimento de marcadores moleculares nas espécies de eucalipto.

RLC_egGMR_1 apresentou polimorfismos de presença e ausência de um único fragmento (Figura 29B). A ausência de fragmento em *E. grandis* esta relacionada a presença somente da sequência do *primer* TGMR1_F (F - *forward*), o que também pode ter ocorrido em *E. saligna* (Figura 29A e B).

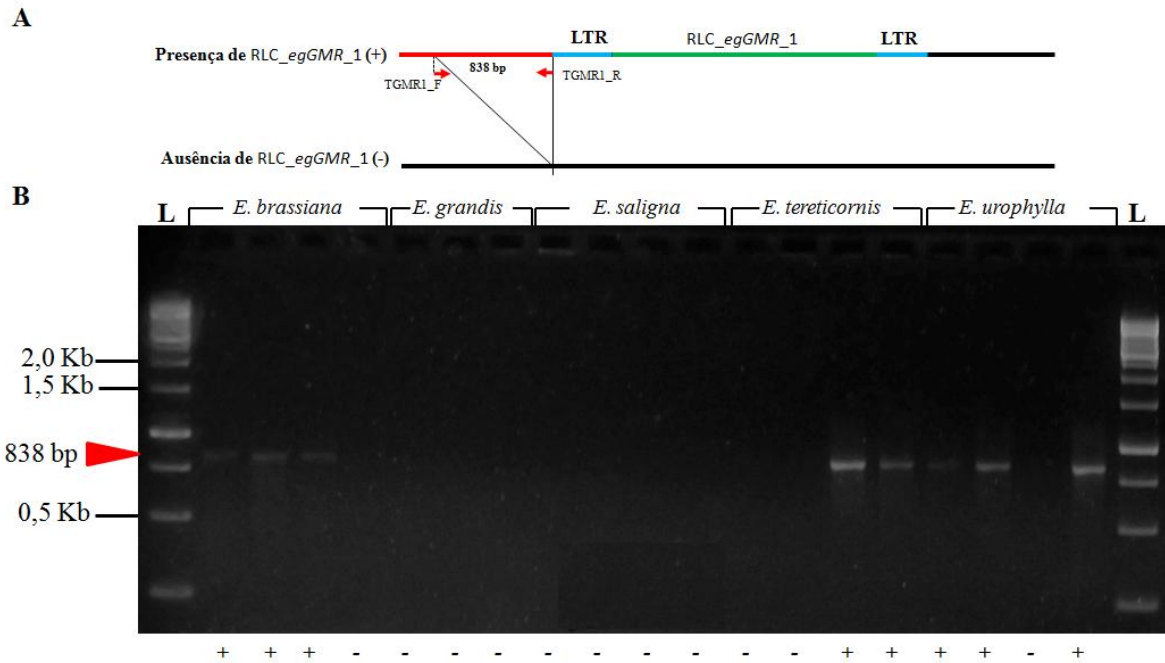


Figura 29: Detecção da família RLC_egGMR_1. A: Ilustração do *primer* RBIP desenhado. Setas vermelhas e azul indicam a posição e direção dos *primers* usados. TGMR1_F X TGMR1_R: conjunto de *primers* que flanqueiam a região genômica a esquerda (F - *forward*) e a direita (R - *reverse*). B: Presença ou ausência de RLC_egGMR_1 nas cinco espécies de eucalipto. L: GeneRuler 1kb DNA Ladder (Fermentas – código SM0311).

RLC_egAle_2 apresentou fragmentos de mesmo tamanho em *E. brassiana*, *E. tereticornis* e *E. urophylla* (Figura 30A). A variação em indivíduos de uma mesma espécie pode estar relacionada a mutações na região adjacente a RLC_egAle_2, visto que ambos os *primers* foram identificados no genoma de *E. grandis*.

RLC_egMax_1 ao ser isolado de fragmento obtido por meio da técnica REMAP, em que foi desenhado o *primer* Umax1 usado em conjunto com o *primer* EgMaxLTR5 (desenhado a partir da 5°LTR) (Tabela 14), apresentou 3 padrões de fragmentos polimórficos - 300, 500 e 700 bp (Figura 31).

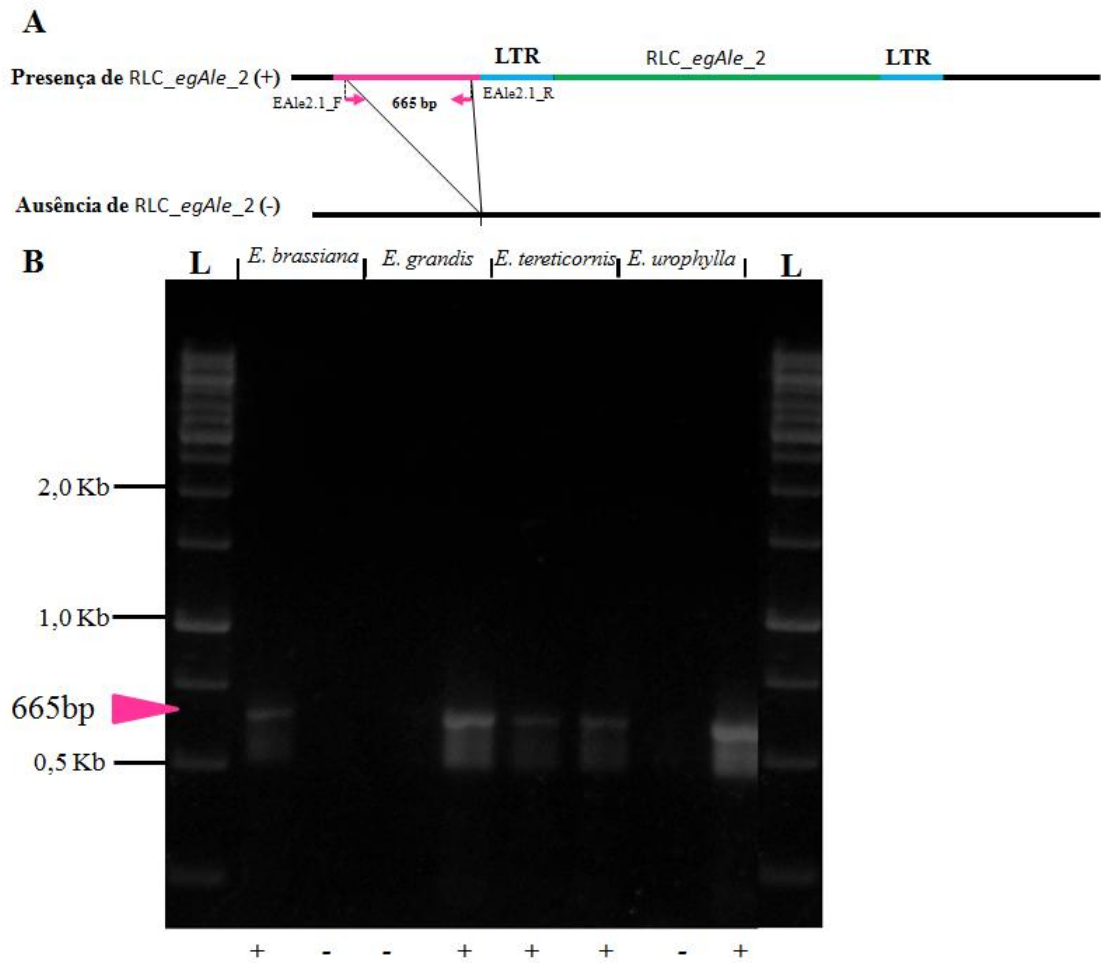


Figura 30: Detecção da família RLC_egAle2_1. A: Ilustração do *primer* RBIP desenhado. Setas rosas indicam a posição e direção dos *primers* usados. EAle2.1_F X EAle2.1_R: conjunto de *primers* que flanqueiam a região genômica a esquerda (F) e a direita (R). B: Presença ou ausência de RLC_egAle2_1 nas cinco espécies de eucalipto. L: GeneRuler 1kb DNA Ladder (Fermentas – código SM0311).

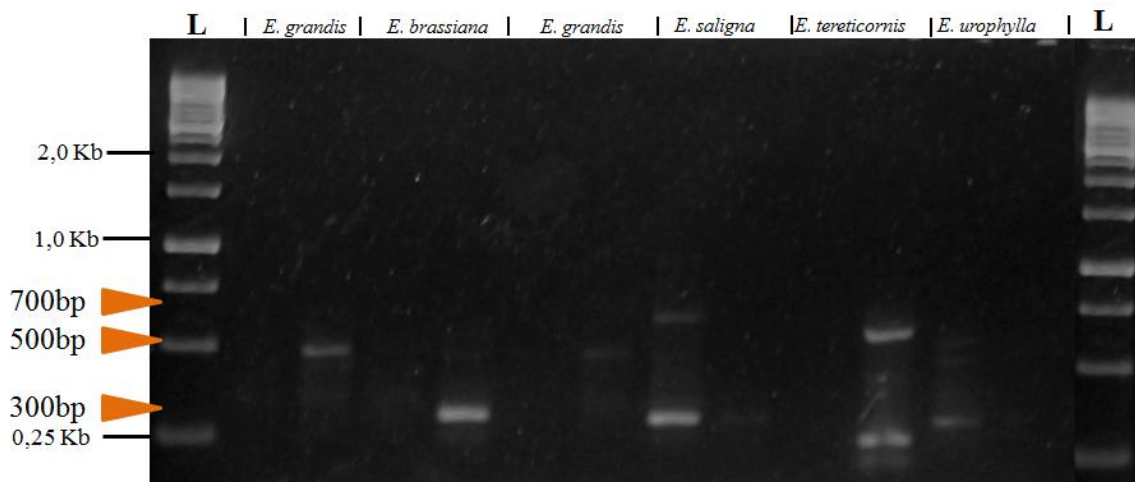


Figura 31: Detecção da família RLC_egMax_1 em cinco espécies de eucalipto com tamanhos de fragmentos diferentes indicados (setas laranja). L: GeneRuler 1kb DNA Ladder (Fermentas – código SM0311).

4.7. Quantificação de famílias de LTR-RTEs por qPCR

Seis famílias de LTR-RTEs foram quantificadas, e para cada uma das famílias se desenhou dois pares de *primers*: um par específico a LTR e outro para uma região interna dos LTR-RTEs (Figura 10 e Tabela 3).

Para validação dos *primers* desenhados, os mesmos foram utilizados para amplificação em PCR convencional, com clonagem e sequenciamento dos fragmentos obtidos, garantindo-se assim a especificidade dos mesmos. Posteriormente, foi calculada a eficiência de amplificação dos mesmos a partir de diluições seriadas do DNA plasmidial contendo o inserto de interesse. Foi gerada uma curva padrão (Figura 32), a qual é produzida pela regressão dos valores de C_T (valor de fluorescência acima do *threshold*) e o logaritmo do número de cópias (Tabela 3). Os valores do logaritmo do número de cópias do plasmídeo foram calculados por meio da massa, concentração e tamanho do plasmídeo (Tabela 3), em função da primeira diluição.

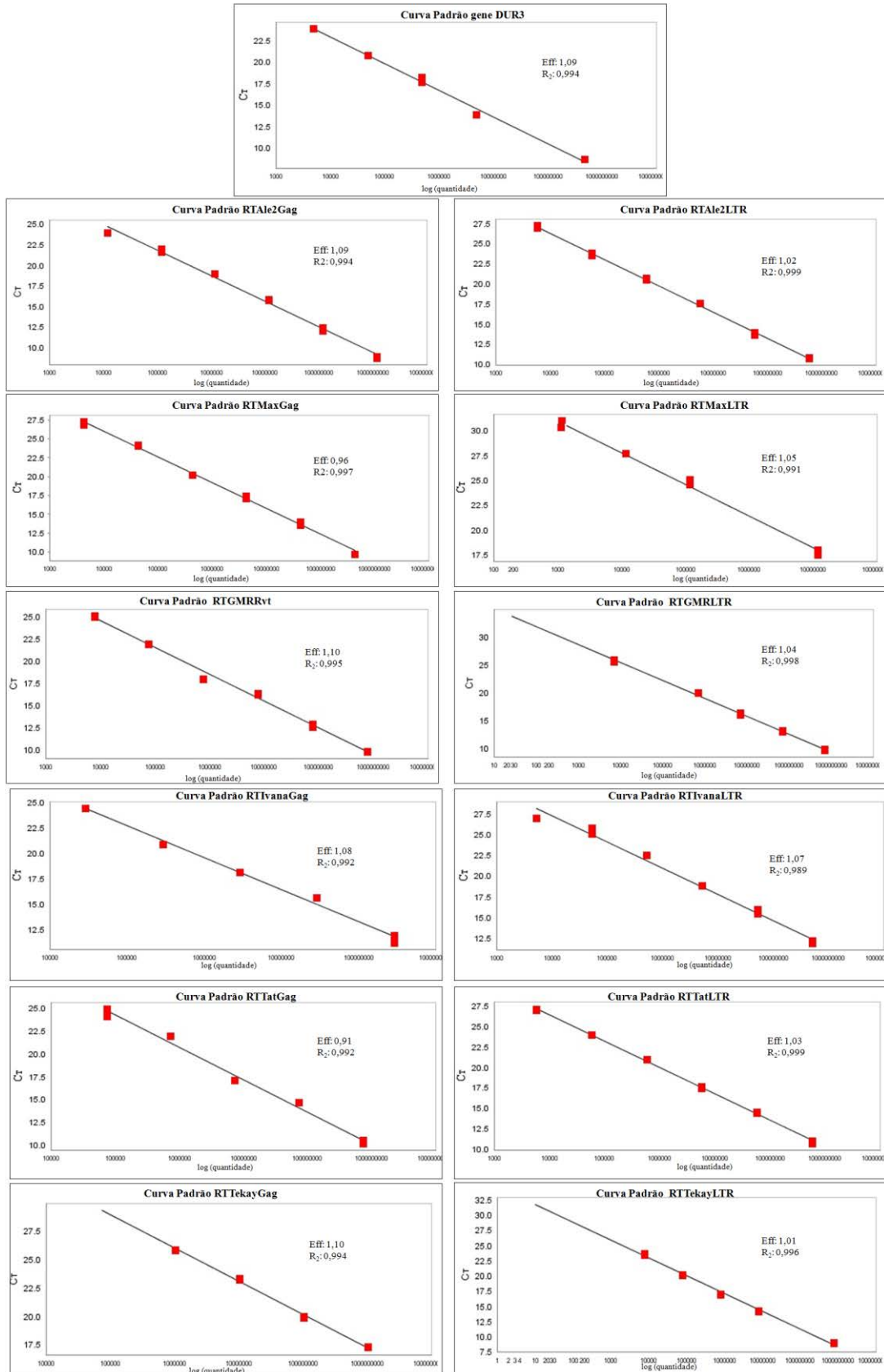


Figura 32: Eficiência de amplificação do qPCR feito para os pares de *primers* utilizados na quantificação de LTR-RTes. Um conjunto de quatro a seis diluições foram usadas, cujos valores de C_T no eixo Y indicam o valor do *threshold*, enquanto que o eixo X indica a quantidade de DNA. Em todos os casos a regressão (R₂) foi acima de 0,99.

A quantificação absoluta dos elementos utilizou a espécie *E. grandis* como calibrador (RQ=1), por já haver informações de distribuição genômica para esta espécie baseada nas informações do genoma. Além disso, foi calculada a razão entre a LTR e a região interna para estimar a proporção de elementos completos em relação às LTRs (Figura 33 a 38).

As figuras 33 a 38 indicam a quantidade relativa entre a região interna e a LTR no genoma dos genitores (*E. grandis* e *E. urophylla*) e os valores do híbrido F₁ observado e esperado.

Para a família RLC_egAle_2 o valor de RQ foi maior em LTR em relação a Gag, o que poderia sugerir um maior número de cópias defectivas em relação a completas (Figura 33A e B). O padrão LTR: Gag no híbrido F₁ foi próximo ao número de cópias de *E. grandis* (Figura 33C).

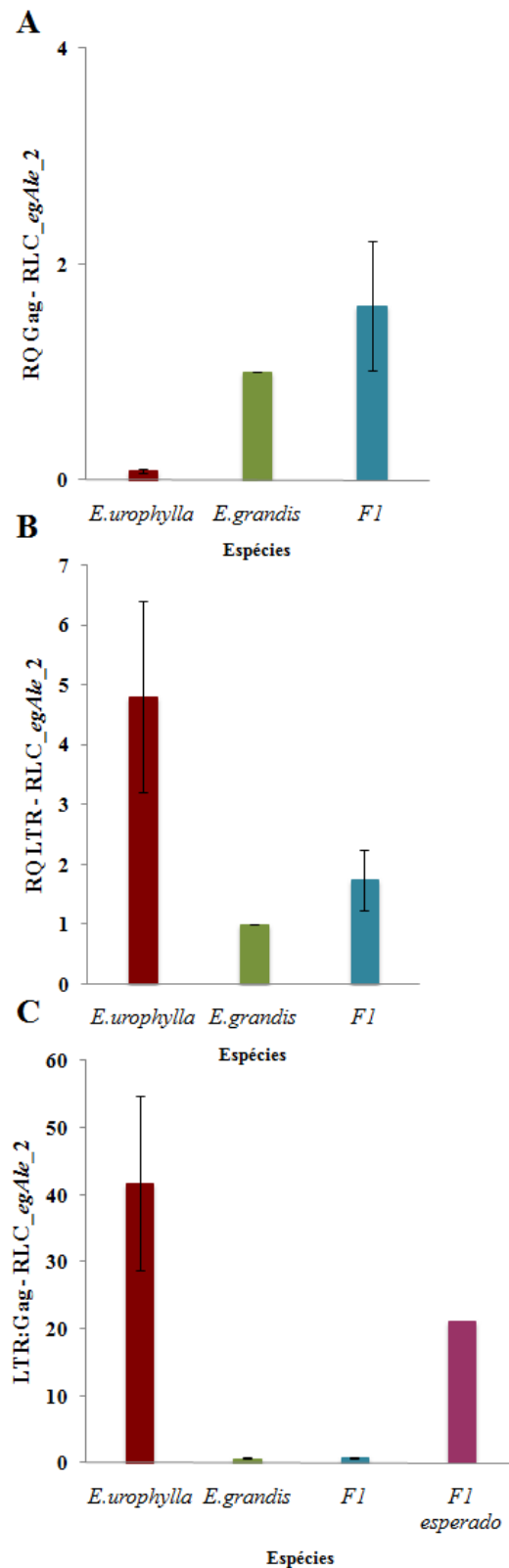


Figura 33: Análises de qPCR do RLC_egAle_2 nos parentais *E. grandis* e *E. urophylla* e no híbrido (F₁). Os resultados de *E. grandis* foram utilizados como calibrador. (A) RQ da região interna (Gag) do RLC_egAle_2. (B) RQ da LTR do RLC_egAle_2. (C) Razão LTR:Gag do RLC_egAle_2. Nas barras estão indicados o erro padrão (n=9).

Por outro lado, para RLC_egMax_1, o híbrido F₁ apresentou um RQ próximo ao de *E. urophylla*. Os valores de RQ no híbrido F₁ (Figura 34A e B) foram maiores para o domínio interno do que na LTR.

Para a família RLC_egGMR_1 foi observado um RQ maior que o do calibrador *E. grandis* para a transcriptase reversa e menor para as LTRs (Figuras 35A e B). Essa diferença de valores de RQ não foi observada na relação LTR/Rvt, na qual o híbrido F₁ apresentou um valor intermediário entre as espécies genitoras (Figura 35C).

Já em RLC_eglvana_1, a razão de RQ no híbrido F₁ foi intermediária aos genitores para Gag e LTR, mas isto não foi observado na razão Gag: LTR (Figura 36).

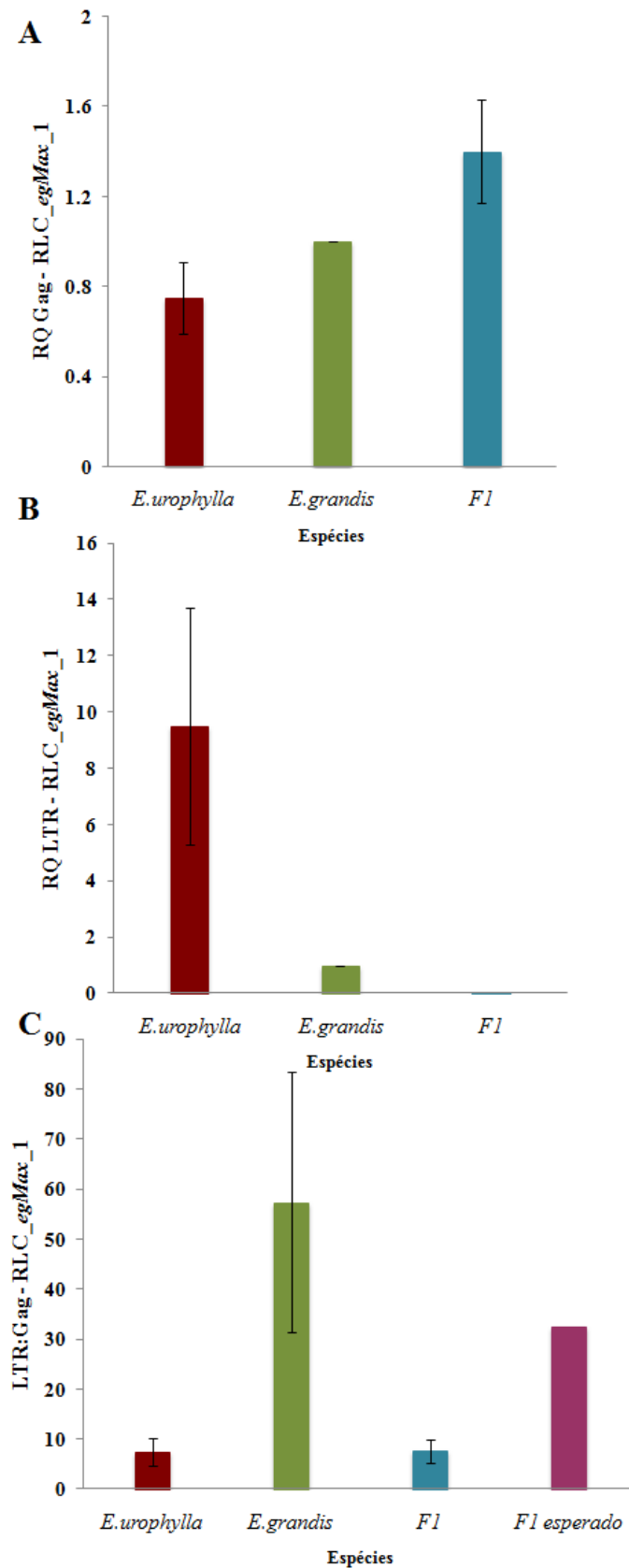


Figura 34: Análises de qPCR do RTE RLC_{egMax_1} nos parentais *E. grandis* e *E. urophylla* e no híbrido F₁. Os resultados de *E. grandis* foram utilizados como calibrador. (A) RQ da região interna (Gag) do RTE RLC_{egMax_1}. (B) RQ da LTR do RTE RLC_{egMax_1}. (C) LTR:Gag do RTE RLC_{egMax_1}. Nas barras estão indicados o erro padrão (n=9).

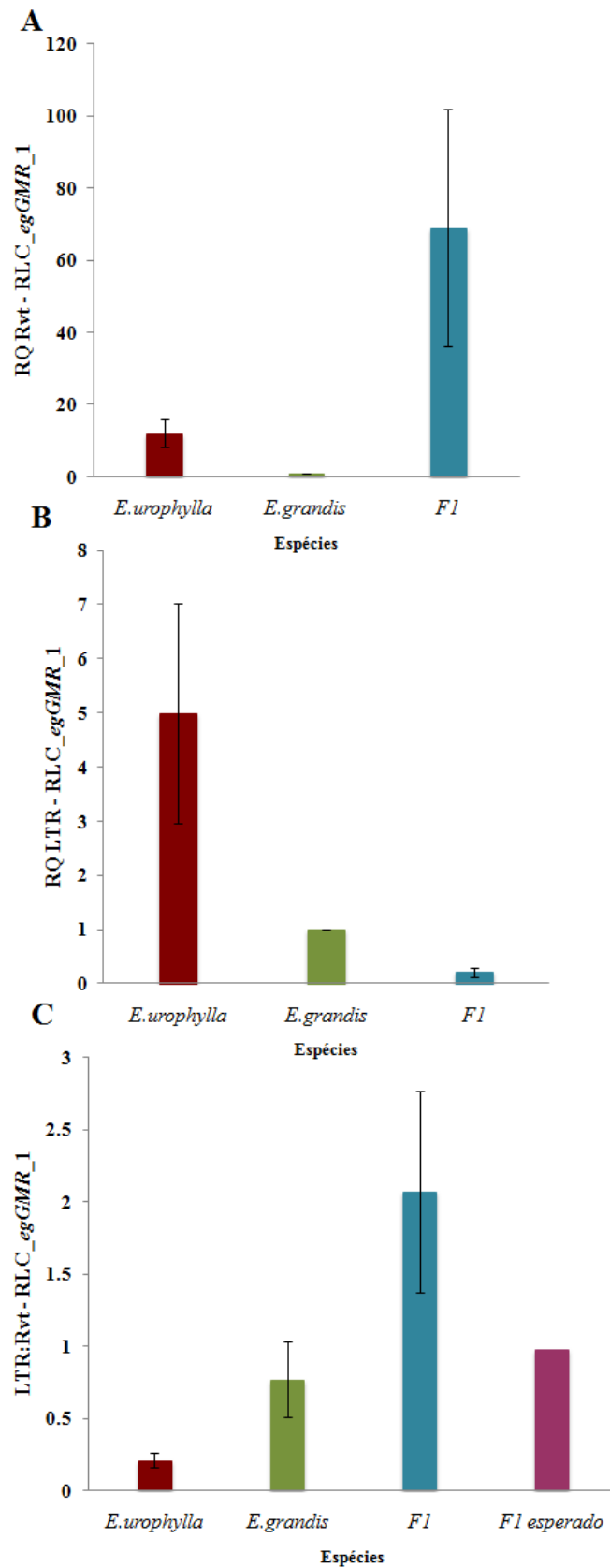


Figura 35: Análises de qPCR do RLC_egGMR_1 nos parentais *E. grandis* e *E. wrophylla* e no híbrido (F₁). Os resultados de *E. grandis* foram utilizados como calibrador. (A) RQ da região interna (Rvt) do RLC_egGMR_1. (B) RQ da LTR do RLC_egGMR_1. (C) LTR:Rvt do RLC_egGMR_1. Nas barras estão indicados o erro padrão (n=9).

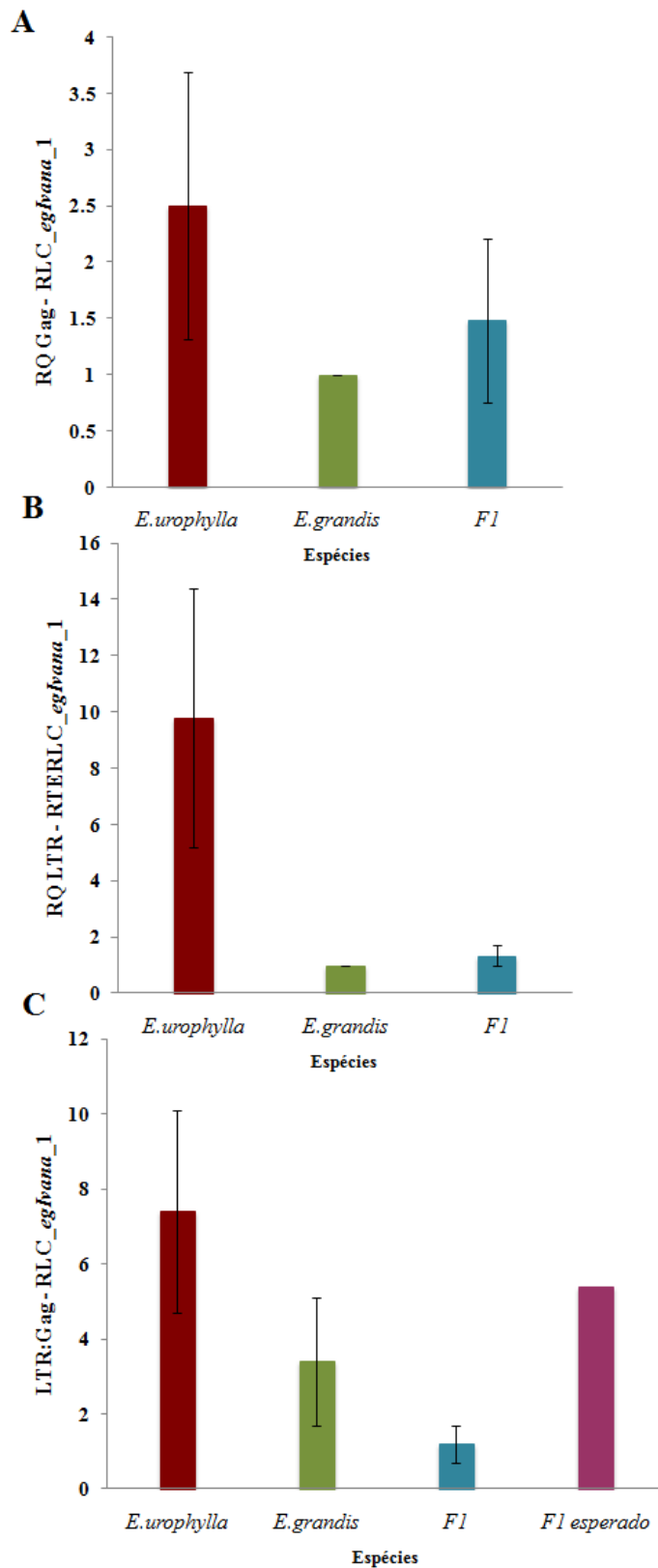


Figura 36: Análises de qPCR do RLC_eglvana_1 nos parentais *E. grandis* e *E. urophylla* e no híbrido (F₁). Os resultados de *E. grandis* foram utilizados como calibrador. (A) RQ da região interna (Gag) do RLC_eglvana_1. (B) RQ da LTR do RLC_eglvana_1. (C) LTR:Gag do RLC_eglvana_1. Nas barras estão indicados o erro padrão (n=9).

A família RLG_egTat_1 apresentou um padrão similar ao de RLC_egIvana_1, com um valor intermediário de RQ para o domínio interno e a LTR, não observando-se o mesmo na razão LTR:domínio interno (Figura 37)

As razões entre LTR:Gag para a família RLG_egTekay_1 foram de 1:0,7 em *E. urophylla*, 1:0,6 em *E. grandis* e 1:0,43 no híbrido F₁ (Figura 38C). A proporção observada sugere um número de regiões internas menor ao de LTRs (Figura 38A e B).

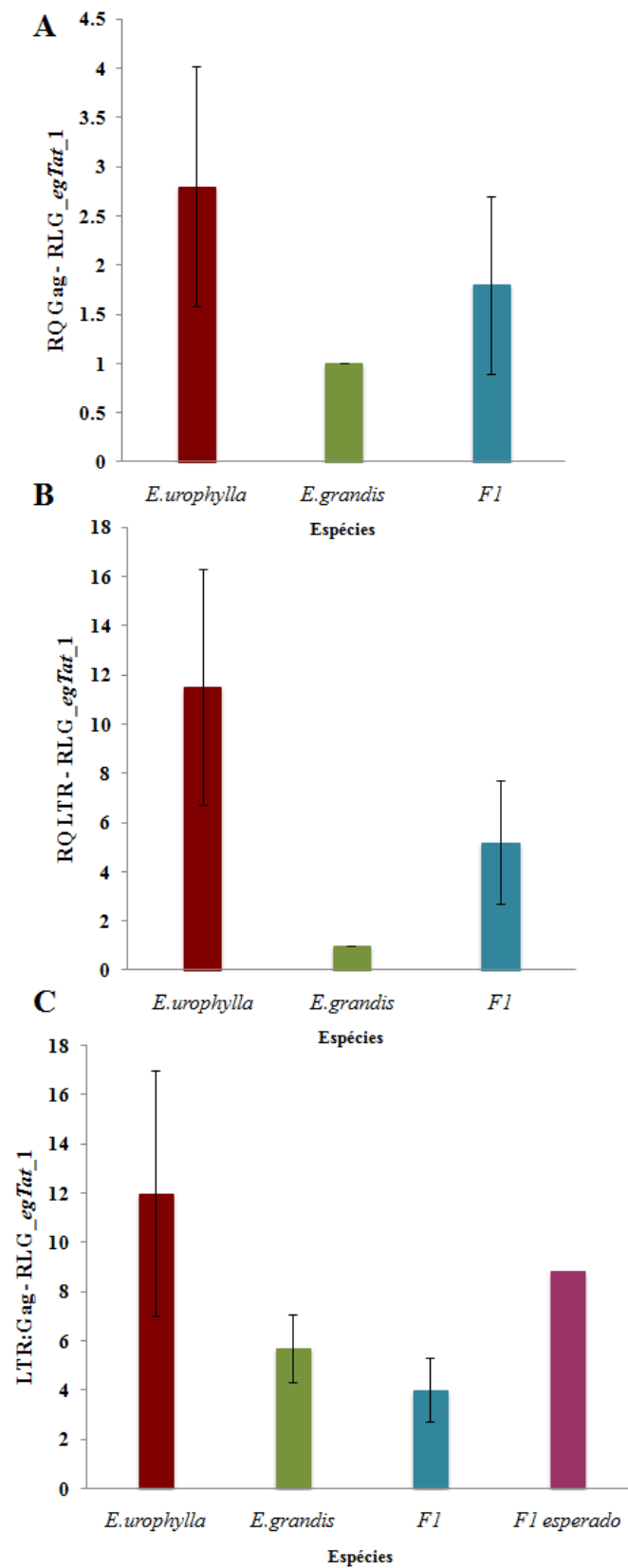


Figura 37: Análises de qPCR do RLG_egTat_1 nos parentais *E. grandis* e *E. wrophylla* e no híbrido (F₁). Os resultados de *E. grandis* foram utilizados como calibrador. (A) RQ da região interna (Gag) do RLG_egTat_1. (B) RQ da LTR do RLG_egTat_1. (C) LTR:Gag do RLG_egTat_1. Nas barras estão indicados o erro padrão (n=9).

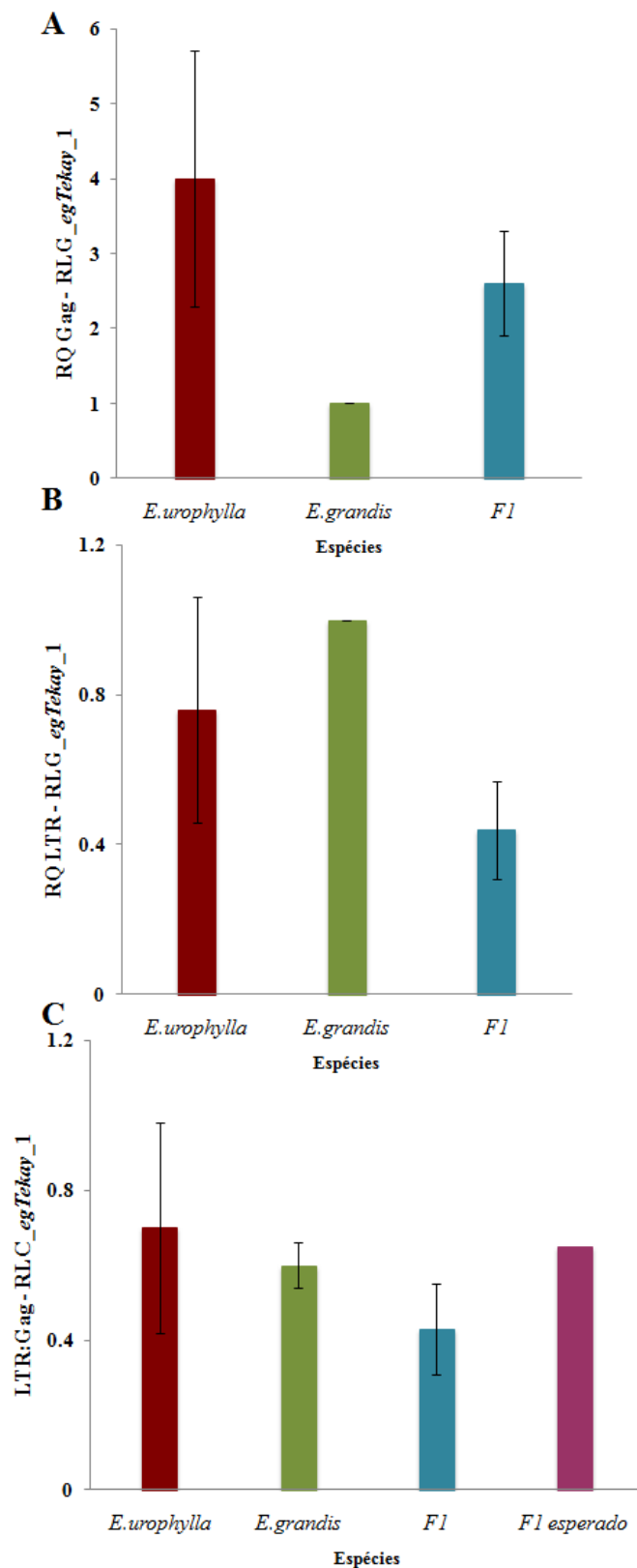


Figura 38: Análises de qPCR do RLG_egTekay_1 nos parentais *E. grandis* e *E. urophylla* e no híbrido (F₁). Os resultados de *E. grandis* foram utilizados como calibrador. (A) RQ da região interna (Gag) do RLG_egTekay_1. (B) RQ da LTR do RLG_egTekay_1. (C) LTR:Gag do RLG_egTekay_1. Nas barras estão indicados o erro padrão (n=9).

4.8. Análise transcricional de LTR-RTEs

Foi detectada atividade transcricional de LTR-RTEs nos três tecidos avaliados em cinco espécies do gênero *Eucalyptus* (Figura 39 a 47) e sob estresse osmótico induzido por PEG em *E. grandis* (Figura 48).

A família *RLC_egAle_1* apresentou os maiores valores de expressão relativa no caule de *E. tereticornis*, seguido pelas folhas do híbrido F₁. Em raízes e folhas sua expressão foi baixa nas demais espécies, com exceção de raiz em *E. saligna* e folha em *E. brassiana* (Figura 39).

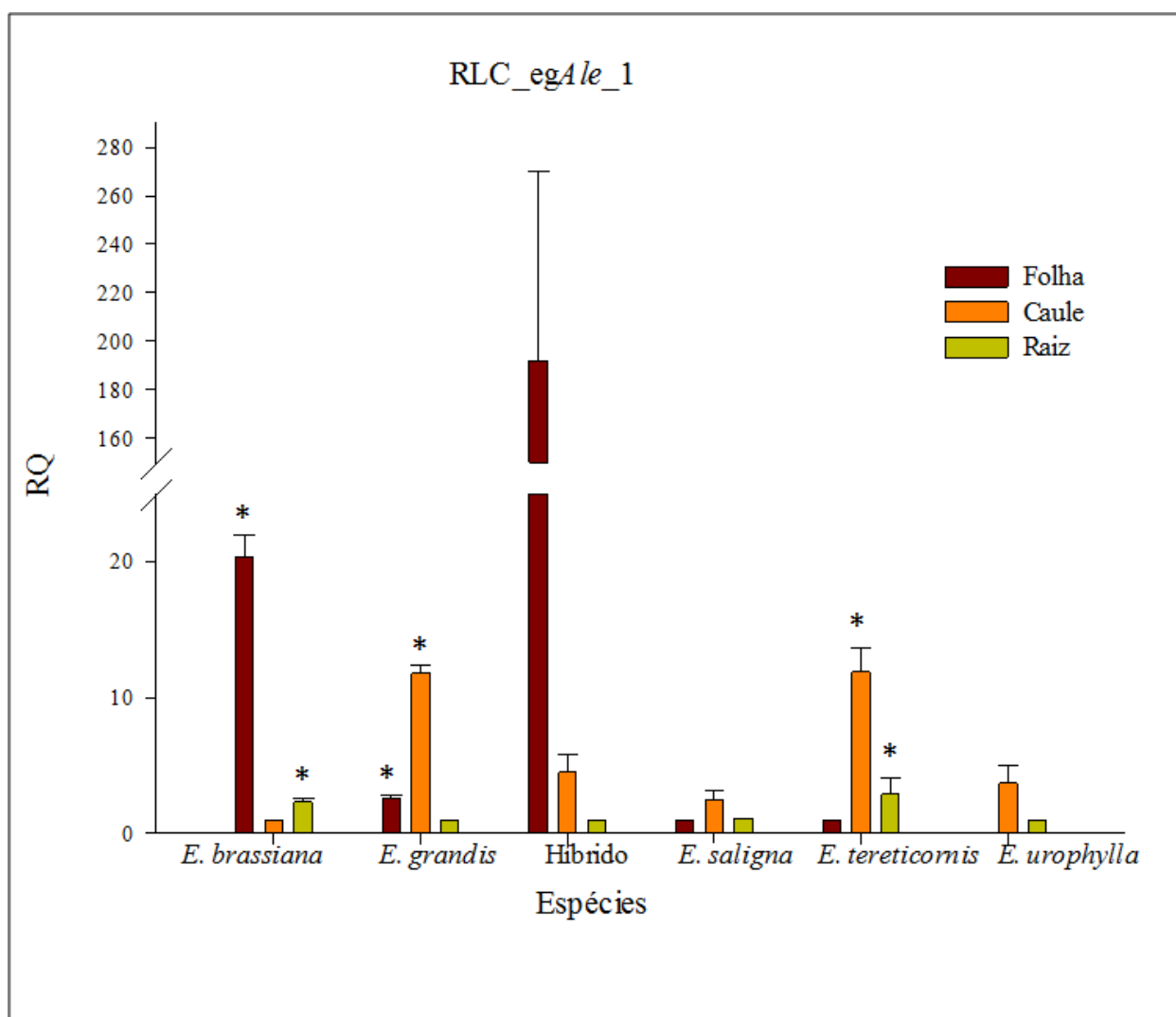


Figura 39: Expressão relativa da família *RLC_egAle_1* em diferentes tecidos de cinco espécies de eucalipto e o híbrido F₁. Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, ANOVA seguido de teste LSD).

A família *RLC_egAle_2* apresentou maior atividade transcricional em folhas de *E. urophylla*. Em caule, de maneira geral, a expressão relativa foi baixa (Figura 40). No Híbrido F_1 a atividade transcricional foi maior em folha e raiz (Figura 40).

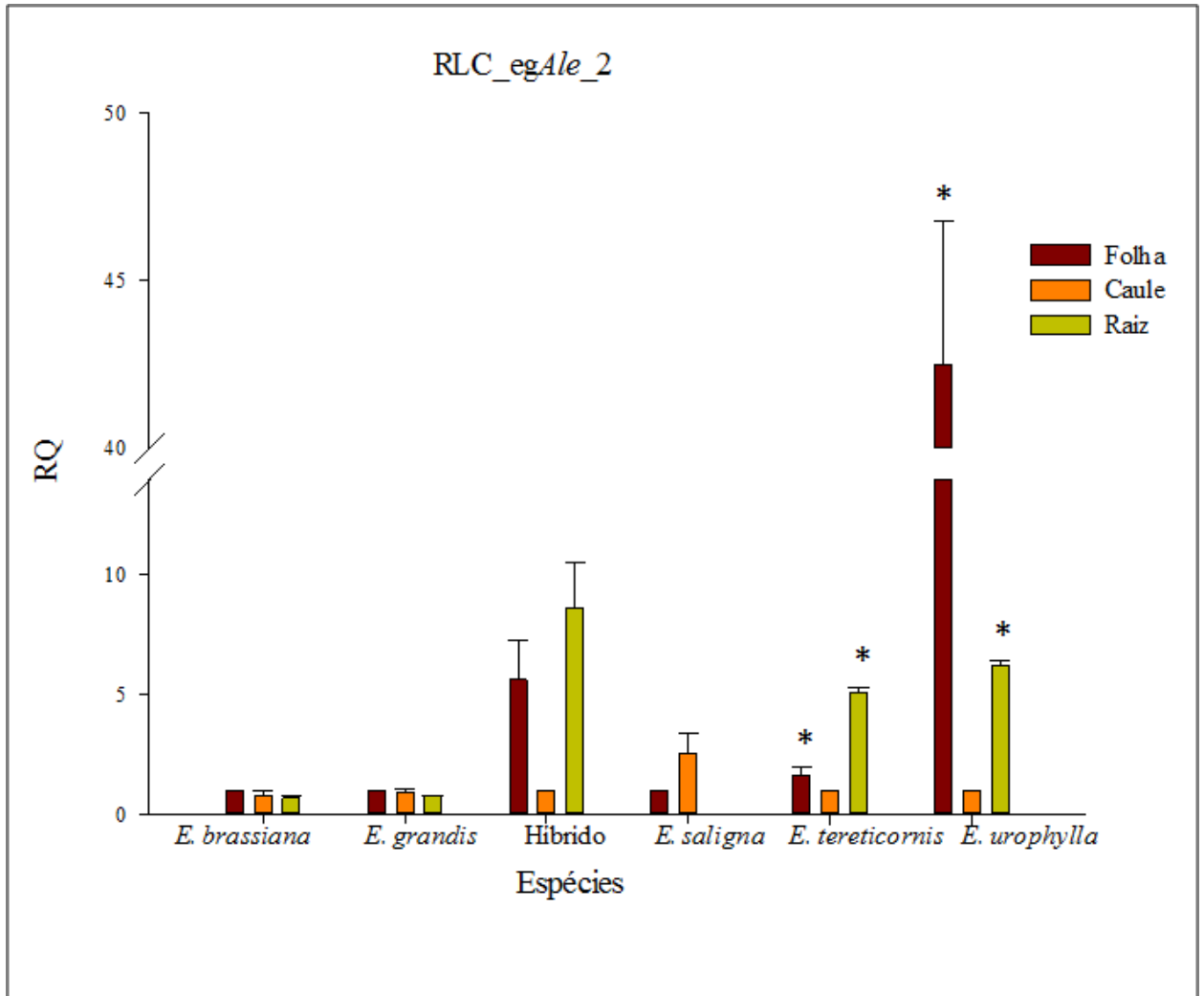


Figura 40: Expressão relativa da família *RLC_egAle_2* em diferentes tecidos de cinco espécies de eucalipto e o híbrido F_1 . Nas barras estão indicados o erro padrão ($n=6$) (* $p \leq 0,05$, ANOVA seguido de teste LSD).

Em *RLC_egMax_1* foi observado uma atividade transcricional mais elevada em raízes de *E. brassiana*, *E. grandis* e no híbrido F_1 . Em caule a expressão relativa foi baixa em *E. saligna* e *E. urophylla*. Já em raiz, o híbrido F_1 apresentou valores altos, maiores do que em seu genitor *E. grandis* enquanto que em *E. urophylla* foi baixo (2,98). Em folha foi observada expressão relativa em *E. tereticornis* (Figura 41).

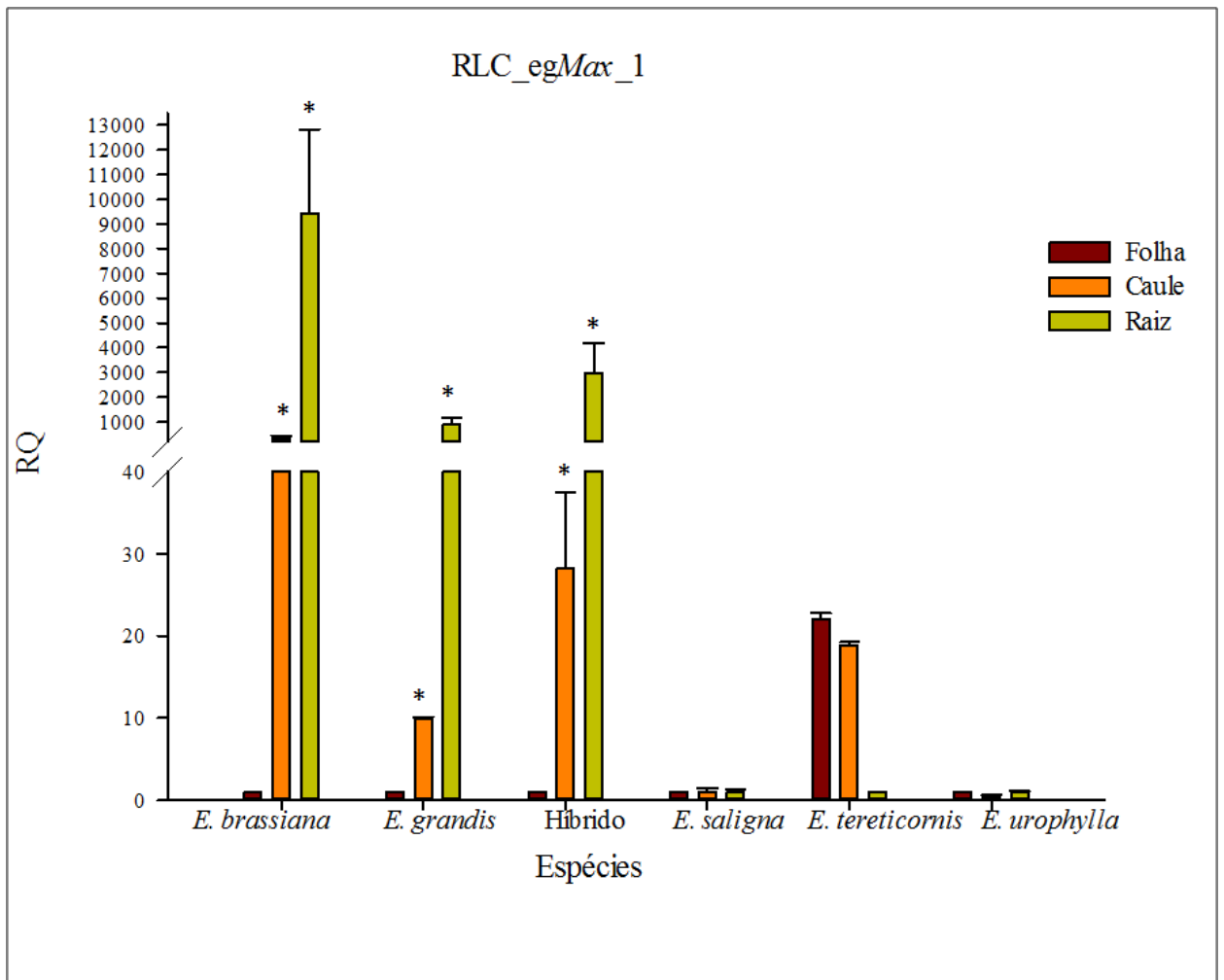


Figura 41: Expressão relativa da família RLC_egMax_1 em diferentes tecidos de cinco espécies de eucalipto e o híbrido F₁. Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, ANOVA seguido de teste LSD).

RLC_egBianca_1 apresentou atividade transcricional alta em folhas em *E. brassiana* e no híbrido F₁. Já em raiz houve maior expressão no híbrido F₁ e ausente em *E. brassiana* e *E. urophylla* (Figura 42).

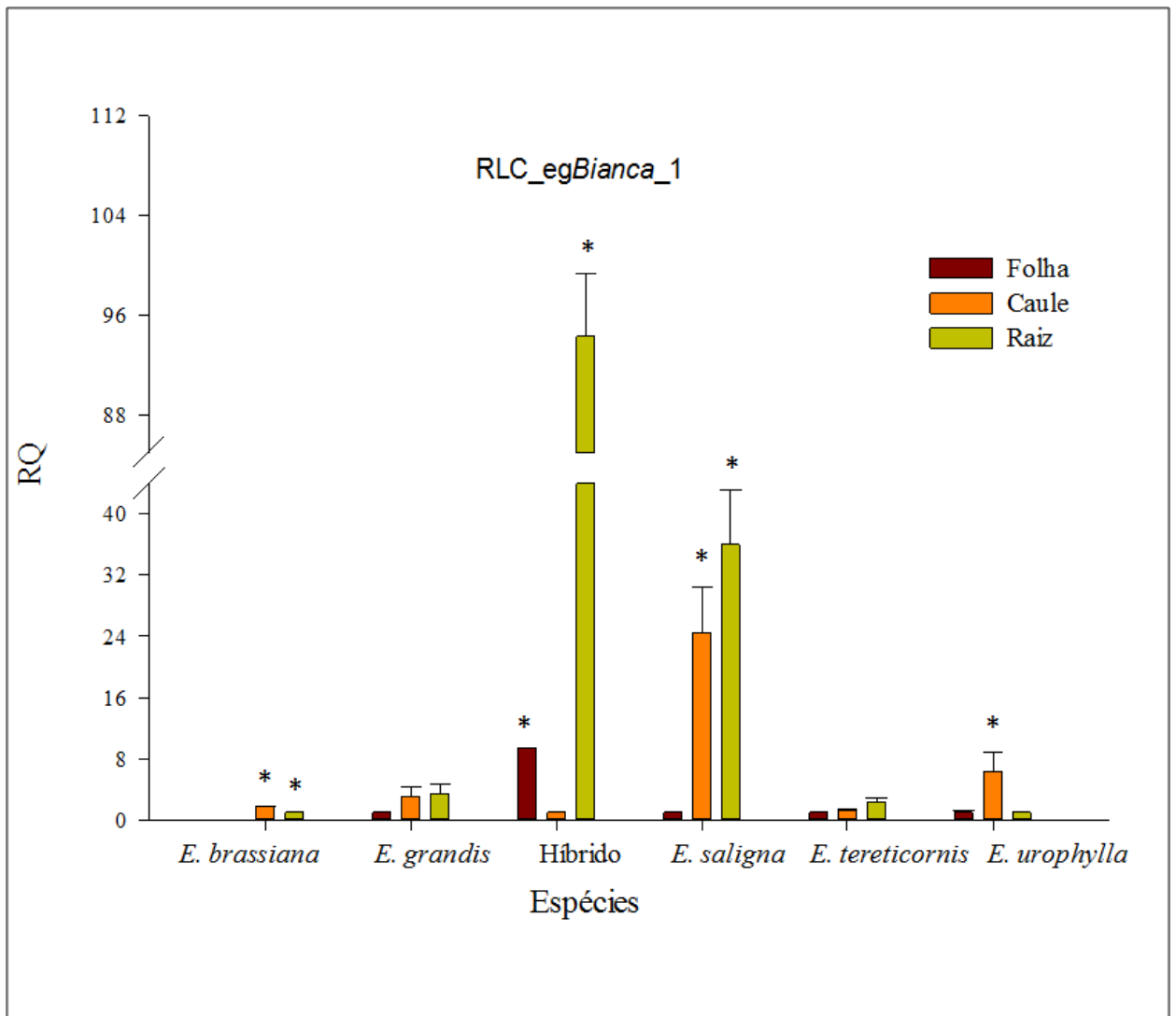


Figura 42: Expressão relativa da família *RLC_egBianca_1* em diferentes tecidos de cinco espécies de eucalipto e o híbrido F₁. Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, ANOVA seguido de teste LSD).

A atividade transcricional de *RLC_egGMR_1* foi detectada no caule de todas as espécies de eucalipto avaliadas e no híbrido F₁, com maior expressão observado no híbrido F₁ e em seus genitores (*E. grandis* e *E. urophylla*) (Figura 43). Em raiz a transcrição desta família foi observada em *E. grandis* e *E. urophylla* (15,6 e 7,6), enquanto nas outras espécies e no híbrido F₁ foram observados valores baixos (Figura 43). Em folha não houve expressão deste LTR-RTE (Figura 43).

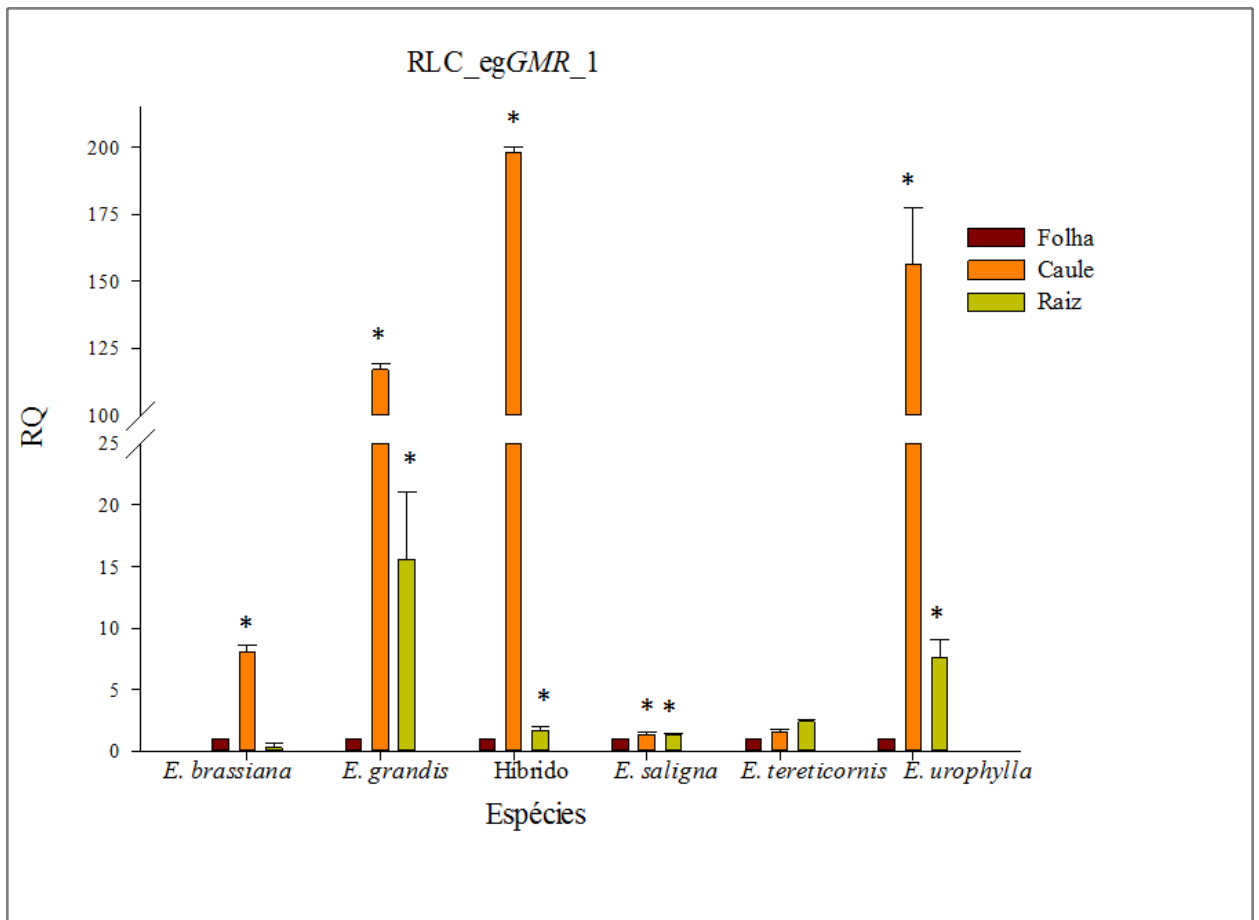


Figura 43: Expressão relativa da família RLC_egGMR_1 em diferentes tecidos de cinco espécies de eucalipto e o híbrido F₁. Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, ANOVA seguido de teste LSD).

A família RLC_egIvana_1 apresentou expressão em todos os tecidos avaliados, mas com intensidades distintas entre cada espécie e no híbrido F₁ (Figura 44). A raiz foi o tecido com maior atividade transcricional em *E. saligna* e no híbrido F₁ (Figura 44). Os menores valores de expressão relativa foram observados em caule, com exceção do híbrido F₁ (Figura 44).

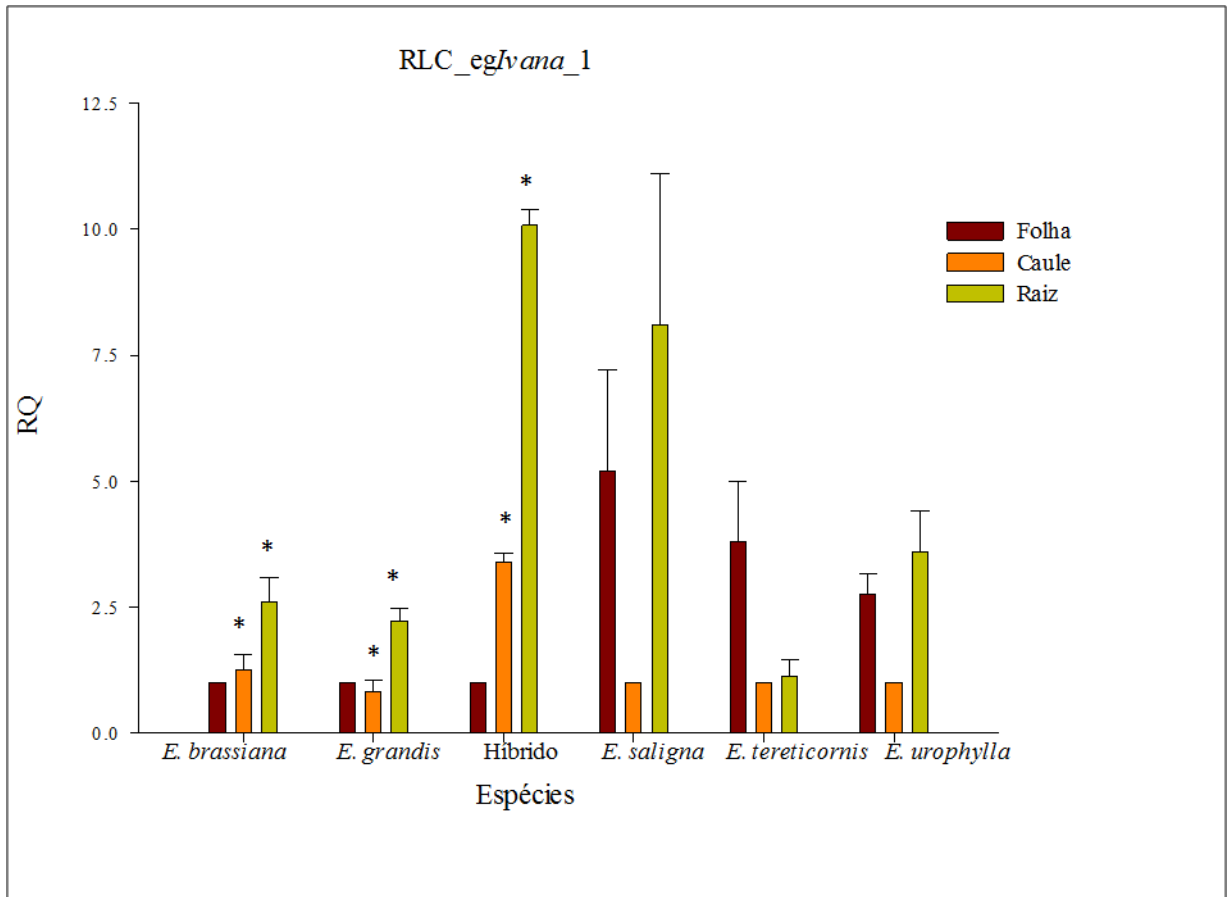


Figura 44: Expressão relativa da família RLC_eglvana_1 em diferentes tecidos de cinco espécies de eucalipto e o híbrido F₁. Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, ANOVA seguido de teste LSD).

A família RLG_egTat_1 apresentou atividade transcricional em folha em duas espécies, e em caule somente não foi detectado atividade transcricional em *E. saligna*, e em raízes a expressão foi identificada apenas no híbrido F₁ (Figura 45).

Já na família RLG_egTekay_1 apresentou atividade transcricional nos tecidos de todas espécies e no híbrido F₁, com exceção de *E. saligna* e de caule e folha de *E. grandis* (Figura 46). *E. tereticornis* apresentou as maiores taxas de expressão em caule e raiz (Figura 46).

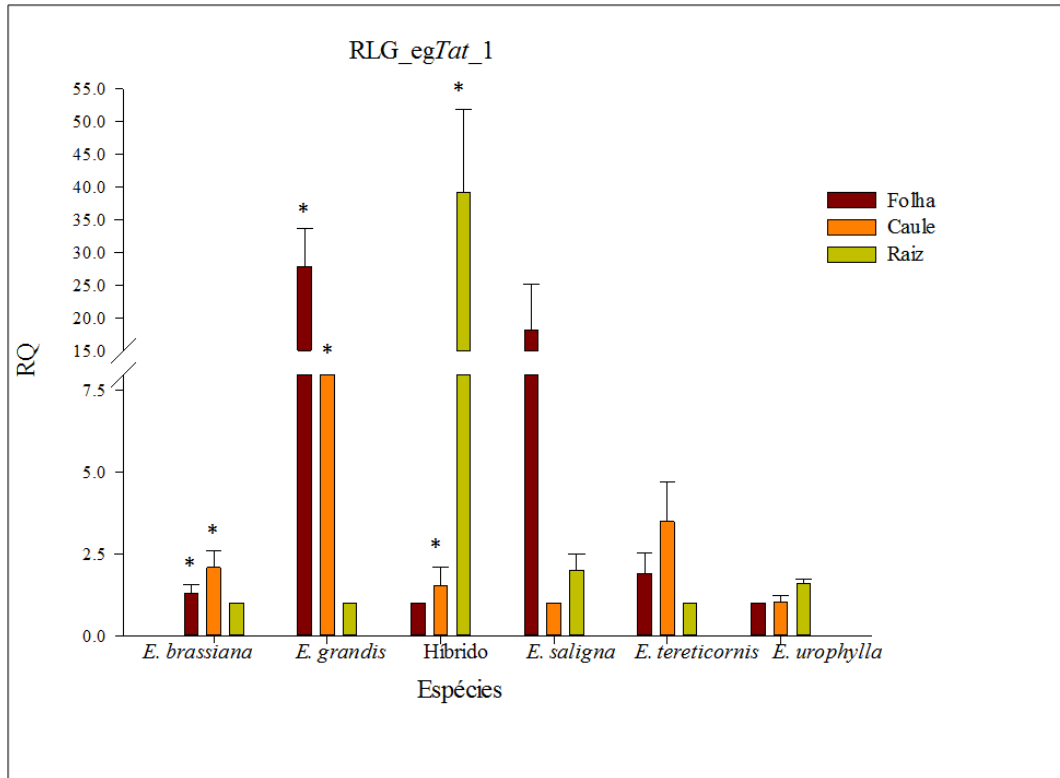


Figura 45: Expressão relativa da família RLC_egTat_1 em diferentes tecidos de cinco espécies de eucalipto e o híbrido F₁. Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, ANOVA seguido de teste LSD).

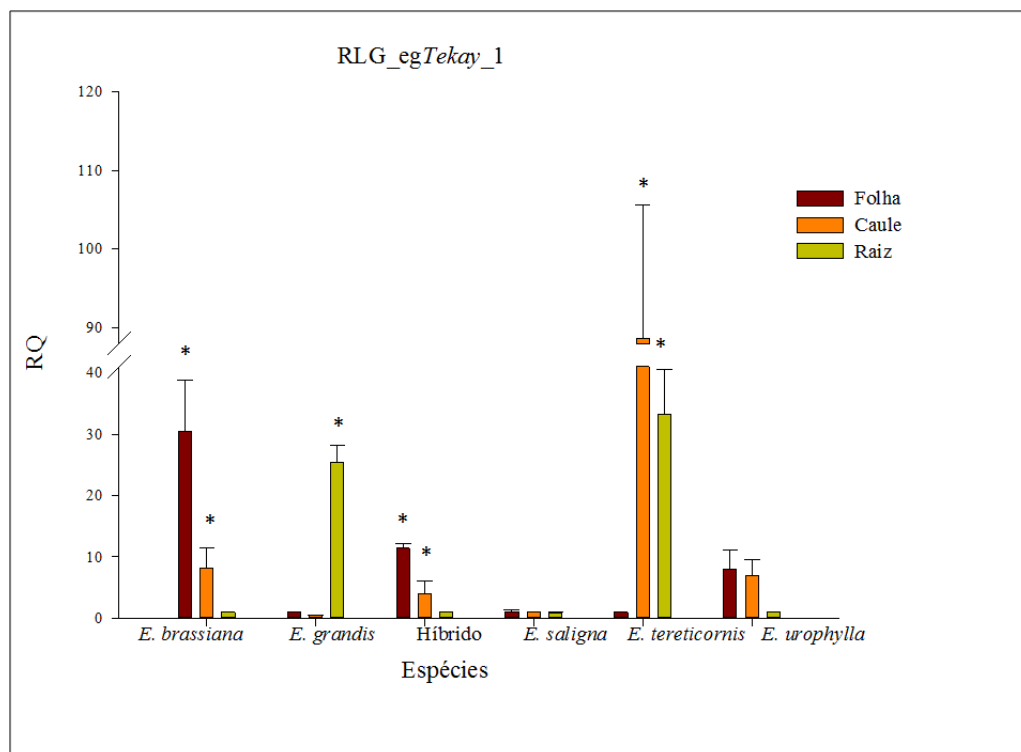


Figura 46: Expressão relativa da família RLC_egTekay_1 por meio de RT-qPCR em diferentes tecidos de cinco espécies de eucalipto e o híbrido F₁. Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, ANOVA seguido do teste LSD).

RLG_egCauli_1 apresentou os maiores valores de atividade transcricional entre os LTR-RTEs pertencentes a superfamílias *Gypsy* descritos no trabalho em caule e raiz do híbrido F₁ (Figura 47). Em folha somente foi observada expressão relativa em *E. saligna*, do mesmo modo que em raiz (Figura 47).

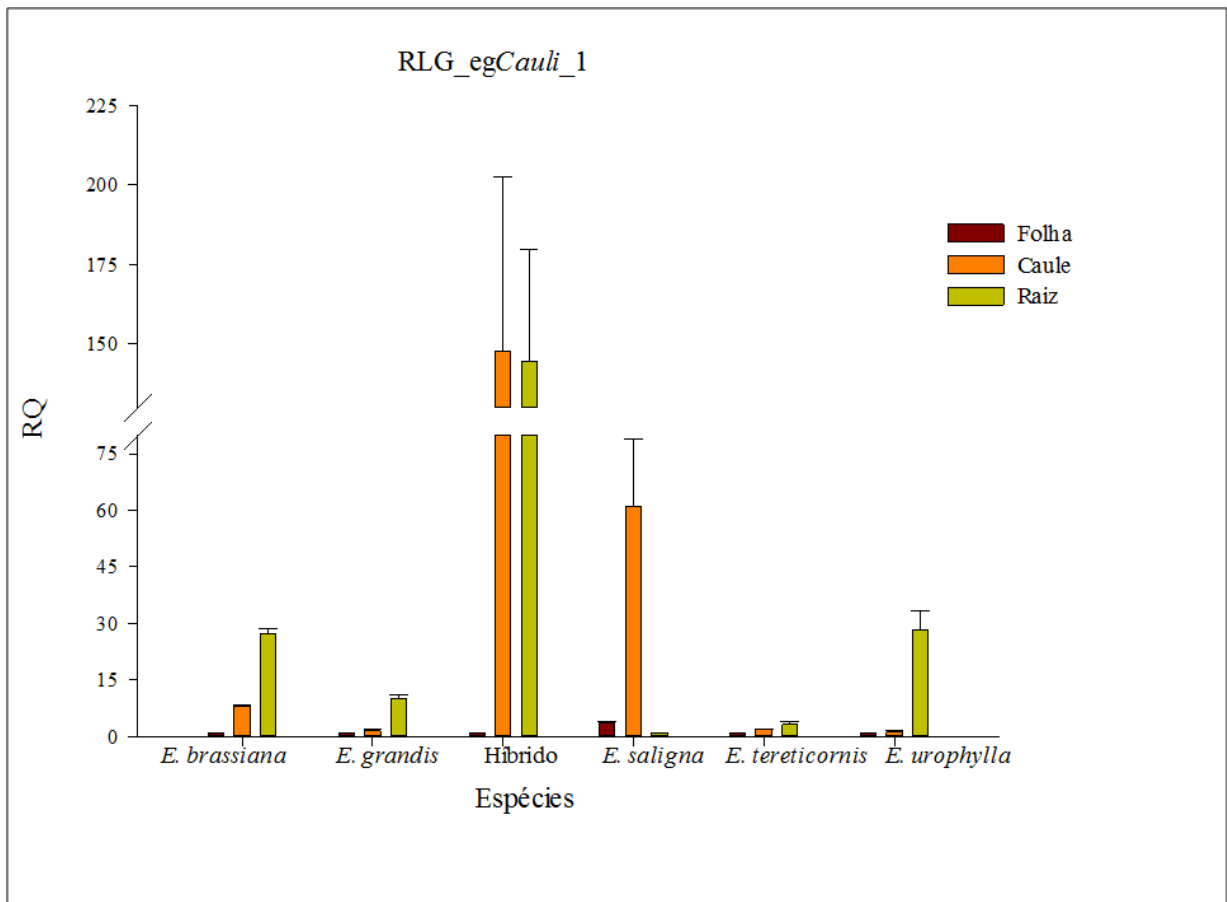


Figura 47: Expressão relativa da família RLC_egCauli_1 em diferentes tecidos de cinco espécies de eucalipto e o híbrido F₁. Nas barras estão indicados o erro padrão (n=6) (*p ≤ 0,05, ANOVA seguido de teste LSD).

Para cinco famílias de LTR-RTEs caracterizadas neste trabalho (RLC_egAle_1, RLC_egMax_1, RLC_egBianca_1, RLC_egIvana_1, RLG_egTat_1, RLG_egTekay_1), foi observado aumento da atividade transcricional nas raízes após seis h de aplicação, diminuindo após 24 h. Somente a expressão de RLC_egAle_1 aumentou após 24 horas e a atividade transcricional de RLC_egAle_2, RLC_egBianca_1 e RLG_egCauli_1 diminuiu após a aplicação do tratamento osmótico (Figura 48).

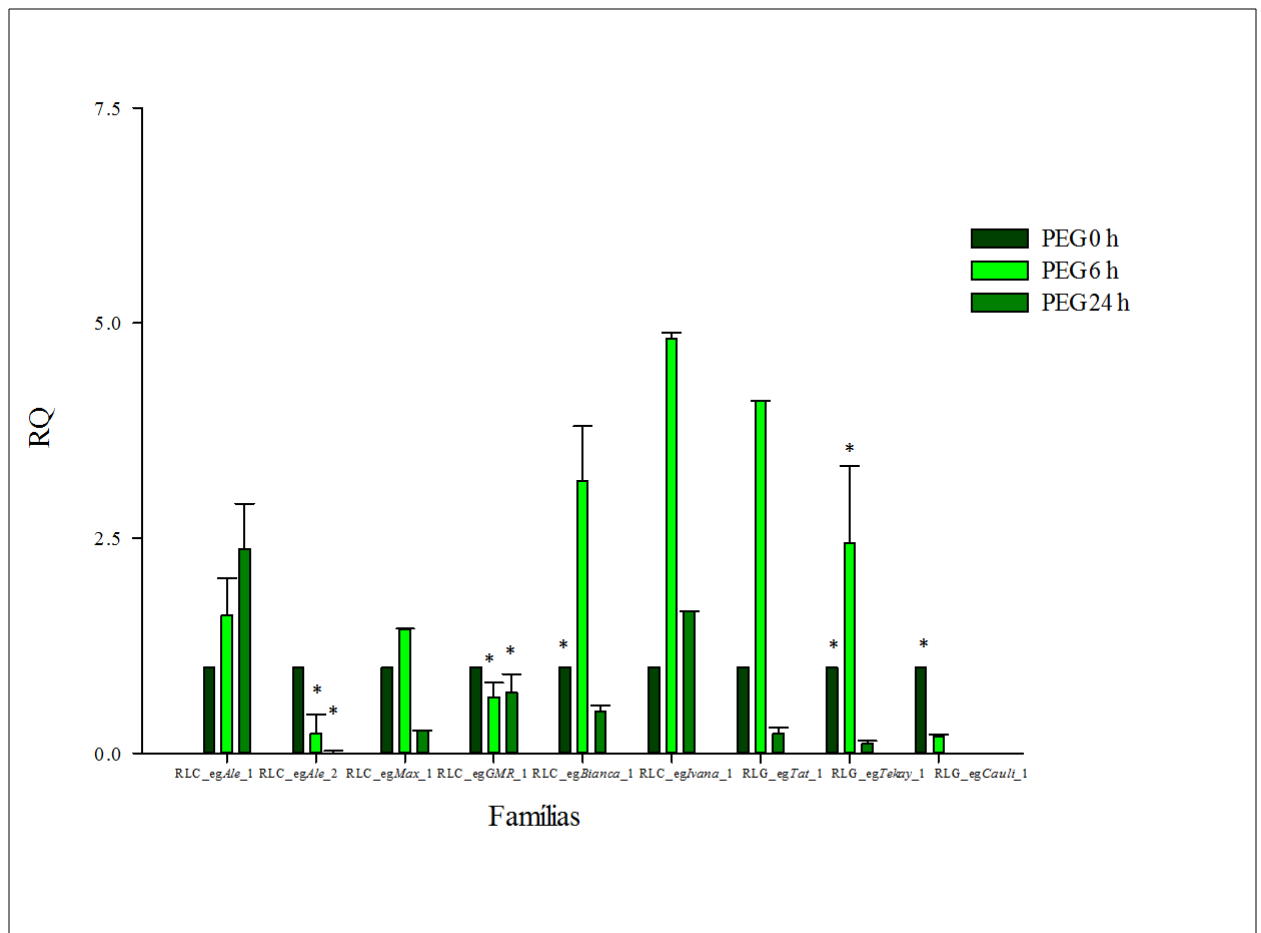


Figura 48: Expressão relativa das famílias de LTR-RTEs por meio de RT-qPCR em plantas de *E. grandis* estressadas osmoticamente com PEG após 6 e 24h. Nas barras estão indicados o erro padrão (n=6) (* $p \leq 0,05$, teste ANOVA seguido de teste LSD).

5. Discussão



Os LTR-RTEs são conhecidos pela sua natureza ubíqua em todos os genomas eucariotos investigados até o momento. Em vegetais, as superfamílias *Copia* e *Gypsy* são as predominantes entre estes elementos, e podem influenciar diretamente na variação do tamanho do genoma entre espécies de um mesmo gênero. Além disso, eles são responsáveis por alterações na estrutura do genoma e podem também agir na regulação de genes.

O presente trabalho buscou identificar LTR-RTEs no genoma de eucalipto que tivessem alguma evidência de atividade transcricional. Para isso, foi feita uma busca inicial em ESTs Sanger do gênero *Eucalyptus* utilizando LTR-RTEs já caracterizados em outras plantas. A partir das sequências parciais de EST, foram obtidos LTR-RTEs transcionalmente ativos completos presentes no genoma *draft* de *E. grandis*, utilizando-se programas de identificação específica deste grupo de elementos. Para validação dos elementos completos, foram localizadas as regiões conservadas características da ORF de LTR-RTEs (Figura 12 e 13). A partir dos LTR-RTEs completos pôde-se desenvolver marcadores IRAP, REMAP e RBIP. A quantificação de seis famílias ocorreu em três genótipos de eucalipto: *E. grandis*, *E. urophylla* e o híbrido F₁ das duas espécies. Para as análises transcricionais, foram utilizados três tecidos de cinco espécies de eucalipto (*E. brassiana*, *E. grandis*, *E. saligna*, *E. tereticornis* e *E. urophylla*) e em um híbrido F₁ (*E. grandis* x *E. urophylla*), além de plantas de *E. grandis* submetidas a estresse osmótico com PEG.

5.1. LTR-RTEs caracterizados em eucalipto

Diversos estudos em plantas com base em dados genômicos em larga escala demonstram que, dentro das superfamílias *Copia* e *Gypsy*, existem linhagens evolutivas, que são compartilhadas entre monocotiledôneas e dicotiledôneas (Wicker et al. 2007; Du et al. 2010; Domingues et al. 2012; Piednoël et al. 2013), e foram formadas antes mesmo da divisão do reino Plantae, há aproximadamente 140 - 150 milhões de anos (Chaw et al, 2004). Posteriormente, com base na similaridade das sequências dentro das linhagens, os LTR-RTEs são subdivididos em famílias (Wicker et al. 2007).

Nas nove famílias de LTR-RTEs descritas neste trabalho foram identificadas as LTRs, marcadas pela região iniciadora 5'-TG-3' e região de término 5'-CA-3', imediatamente após a 5'LTR e a 3'LTR estão as regiões PBS e PPT, respectivamente, que são conservadas

ao longo das espécies (Wicker e Keller 2007), e regulam de forma *cis* o início da transcrição e da replicação dos LTR-RTEs (Zhao e Ma 2013).

As famílias pertencentes à superfamília *Copia* foram caracterizadas quanto ao posicionamento da integrase, Rvt e RnaseH, que garantem a retrotransposição e integração das famílias de LTR-RTE ao genoma (Figura 12). Também se identificou os mesmos domínios nas famílias pertencentes a superfamília *Gypsy*, e em RLG_egTat_1 somente se identificou a Rvt (Figura 13).

Dentre os LTR-RTEs da superfamília *Copia*, os elementos da linhagem evolutiva *Ale* (RLC_egAle_1 e RLC_egAle_2) apresentaram tamanhos semelhantes de LTRs, com aproximadamente 500 bp e tamanho de cerca de 5 Kb (Figura 12 e tabela 5), o que está de acordo com o proposto por Wicker e Keller (2007), que caracterizou os LTR-RTEs da linhagem *Ale* como os menores entre os identificados em arroz, *Arabidopsis* e Triticaceae.

O maior elemento completo da superfamília *Copia* pertence a linhagem *Maximus* (RLC_egMax_1). O padrão de RLC_egMax_1 é característico da linhagem a qual pertence, com longas LTRs, além de possuir uma grande região não codificante interna e apresentar uma segunda ORF a jusante (*downstream*) dos domínios gag-INT-RT (Wicker e Keller 2007). Embora não haja função conhecida para esta ORF2, há uma grande similaridade entre a mesma e o gene *env* de retrovírus (Wilhelm e Wilhelm 2001).

A família RLC_egBianca_1 pertencente a linhagem evolutiva *Bianca* é a que possui maior distância em relação a outras linhagens da superfamília *Copia*, assim como observado por Wicker e Keller (2007).

As regiões PBS e PPT identificadas nos LTR-RTEs pertencentes a superfamília *Copia* apresentaram tamanho próximo a 20 bp, assim como identificadas por Wicker e Keller (2007), que indicaram uma possível variação no tamanho entre 20 e 25 bp. É conhecido que famílias de uma mesma linhagem possuam composição semelhantes nas regiões de PBS e PPT (Wicker e Keller 2007). Nas famílias RLC_egAle_1 e RLC_egAle_2 pertencentes a linhagem *Ale*, 14 nucleotídeos foram idênticos em PBS, e 11 na região PPT, indicando uma maior variabilidade ao longo de uma mesma linhagem evolutiva. A similaridade entre estas regiões muitas vezes também é encontradas entres famílias de linhagens diferentes, mas que estejam intimamente relacionadas (Wicker et al. 2007).

As diferentes famílias pertencentes a superfamília *Copia* apresentam similaridade a outras regiões PBS já identificadas em linhagens descritas em outras plantas (Figura 14). Nas famílias pertencentes as linhagens evolutivas *Ale* e *Maximus* foram as que guardaram

maior similaridade com os motivos PBS e PPT descritos por, Wicker e Keller (2007), enquanto nas linhagens *Bianca* e *Ivana* essa similaridade foi menor (Figura 14).

Os três LTR-RTEs da superfamília *Gypsy* estão distribuídos em três linhagens evolutivas que apresentaram número de cópias variável. Para os LTR-RTEs da superfamília *Gypsy*, pouco se conhece sobre a relação entre linhagem evolutiva e número de cópias as quais pertencem.

As linhagens de LTR-RTEs pertencentes a superfamília *Gypsy* podem ser subdivididas a partir da presença de um domínio *Chromodomain* (*Chromatin Organization Modifiers*) em: *Athila/Tat* (que não apresentam *Chromodomain*) e *Chromoviridae*, que são subdivididos em *Tekay*, *Reina*, *Galadriel* e *CRM* e possuem *Chromodomain* (Lloréns et al. 2009).

As famílias identificadas se enquadraram no perfil característico da superfamília *Gypsy*, com LTR-RTEs maiores que 6 kb, a diferença no tamanho do elemento RLG_egTat_1 pode estar relacionado ao tamanho da região espaçadora entre a LTR e as regiões Gag/pol (Figura 13). As LTRs foram maiores que 500 bp, com exceção de RLG_egCauli_1 (Tabela 8).

Dentre as famílias caracterizadas, RLG_egTat_1, pertencente linhagem a *Tat*, é a com maior similaridade com retrovírus, linhagem que muitas vezes possui um gene que codifica uma proteína transmembrana (Wright e Voytas 2002; Du et al. 2010).

A família RLG_egTekay_1 foi a única a ser identificado um *Chromodomain*, o qual possui geralmente de 40-50 bp e é responsável pela geração de uma proteína que auxilia no remodelamento da cromatina (Koonin et al. 1995). Frequentemente este domínio está integrado ao domínio C-terminal da integrase, recebendo o nome de chromo-integrase. As observações sugerem que este possa controlar a integração do LTR-RTE ao genoma (Novikova et al. 2009). Este *Chromodomain* possivelmente é membro do grupo II (Gorinsek et al. 2004; Gao et al. 2008), o qual é encontrado em LTR-RTEs de plantas da linhagem evolutiva *Tekay*, caso do RLG_egTekay_1.

RLG_egCauli_1 apresentou características de LTR-RTE com LTRs completas, regiões PBS e PPT, Gag/pol (Figura 13 e Tabela 8) e também múltiplas cópias completas no genoma de *E. grandis*, no entanto suas análises filogenéticas não demonstraram homologia com as linhagens evolutivas já descritas na literatura (Du et al. 2010; Neumann et al. 2011). A similaridade com retrovírus endoparasita de plantas, caulimovírus, e a existência de diversos relatos sobre a integração e manutenção de caulivirus em genomas vegetais (Harper et al.,

2002; Pahalawatta et al., 2008; Squires et al., 2011), abrem a perspectiva de estudos nessa linha no gênero *Eucalyptus*.

As análises *in silico* contra ESTs públicos de cinco espécies do gênero *Eucalyptus* mostraram que todas as famílias de LTR-RTEs apresentaram ESTs em *E. globulus*. RLC_egBianca_1 foi o único transcricionalmente ativo todas as espécies analisadas, com exceção de *E. gunni*, enquanto que RLC_egIvana_1 somente foi identificado em *E. globulus* (Tabela 9). Nesse sentido, é provável que a composição da LTR e a distribuição genômica destes elementos seja bastante distinta, abrindo margem a caracterização comparativa destes elementos em futuros trabalhos.

5.2. Número de cópias dos LTR-RTEs

As análises quantitativas dos LTR-RTEs podem ser feitas por diferentes frentes, que vão desde a quantificação em porções sequenciadas utilizando-se programas de alinhamento, como já realizado em diversas espécies vegetais (Moisy et al. 2008; Cossu et al. 2011; Setta et al. 2012; Yin et al. 2013), ou pela quantificação por qPCR (Manetti et al. 2009; Kraitshtein et al. 2010; Yaakov e Kashkuch 2012; Baruch e Kashkuck 2012; Yaakov et al. 2013; Yaakov e Kashkuch 2013). A variação no número de LTR-RTEs é um dos fatores mais importantes para se compreender como estes estão se proliferando nos genomas vegetais.

Os dados gerados a partir da ferramenta do BLAT apresentaram um número de cópias entre as duas superfamílias de LTR-RTEs amplamente variável, inclusive dentro de uma mesma superfamília, com valores entre 38 e 290. Todos os LTR-RTEs da superfamília *Gypsy* apresentaram menos que 100 cópias completas no genoma.

É comum em diversos genomas vegetais casos em que poucas famílias de LTR-RTEs podem ser responsáveis por uma porção substancial de um genoma (Hawkins et al. 2006), como em pinhão-manso, onde três famílias de LTR-RTEs compreendem grande parte do total do genoma (Alipour et al. 2013). As famílias de LTR-RTEs identificadas neste trabalho estão dispersas por todo genoma de *E. grandis*, correspondendo a uma porção significativa do mesmo (Figuras 15 e 16). RLC_egMax_1 foi a família mais representativa (Tabela 5 e Figura 15).

Em populações naturais de genitores de espécies diferentes e uma geração F₁ híbrida é esperado que os valores de número de cópias de LTR-RTEs sejam intermediários

aos dos genitores (Ungerer e Kawakami 2013). Além disso, a hibridação interespecífica tem sido apresentada como um evento que leva a especiação e a evolução de genomas (Rieseberg et al. 2003), visto a instabilidade gerada durante a segregação dos cromossomos (Zhao e Ma 2013), que leva a ativação da retrotransposição de LTR-RTEs (Kashkush et al. 2002; Shapiro 2010). As famílias de LTR-RTEs investigadas no presente trabalho apresentaram no híbrido F₁ número de cópias intermediário ao dos genitores e também um número de cópias maior que a dos parentais, indicando uma atividade de retrotransposição de algumas famílias no genótipo avaliado. As variações quanto ao posicionamento dos LTR-RTEs contribuem para a diversidade genética e induzem polimorfismos nos híbridos formados (Kraitshtein et al. 2010).

A razão entre as LTRs e a quantidade de elementos completos e defectivos em plantas pode ser variável como apresentado em soja de 1:8 (LTR:elemento completo), implicando em uma baixa taxa de LTR-RTEs perdidos (Wawrzynski et al. 2008). Em *Solanum sp.* foi de 0,5:1, que indica a existência de LTR-RTEs truncados ou LTRs que divergiram (Manetti et al. 2009).

E. grandis apresentou os maiores valores quanto ao número de cópias completas em relação a defectivas nas famílias RLC_egMax_1 e RLG_egTat_1, do mesmo modo que já observado nas análises *in silico*. Em *E. urophylla* o maior número de cópias completas foi em RLC_egAle_2 e RLG_egTat_1.

As famílias RLC_egAle_2, RLC_egMax_1 e RLC_egGMR_1 apresentaram maior número de cópias completas no híbrido F₁ em relação ao seus genitores (Figuras 33 a 35A), o que indica que os mesmos podem ter tido sua atividade de retrotransposição ativada em F₁ assim como proposto por Kawakami e colaboradores (2010), que identificou um aumento de 3,7 vezes em girassóis híbridos em relação aos genitores, e como Wang e colaboradores (2010) ao observar a atividade de transposição de duas famílias ser retomada após o evento de hibridação.

Embora a hibridação possa ser um evento desencadeador da retrotransposição dos LTR-RTEs, três famílias apresentaram RQ para região interna intermediário ao dos genitores, em RLC_egIvana_1, RLG_egTat_1 e RLG_egTekay_1 (Figuras 36 a 38A), o que indica que o pareamento cromossômico no híbrido não sofreu alteração nestes casos, ocorrendo a segregação dos cromossomos de forma eficiente, sem a perda de material genético. Deste modo não são todas as famílias que passam a ter sua retrotransposição ativada a partir de um mecanismo estressor em um mesmo indivíduo, podendo ocorrer até mesmo a não ativação de

nenhuma família, assim como observado em plantas F₁ obtidas a partir do cruzamento entre genótipos diferentes de girassol (Ungerer e Kawakami 2013).

Em *E. urophylla* somente a família RLG_egTekay_1 apresentou maior número de cópias completas em relação a defectivas (figura 37A e B). Assim os dados sugerem que em *E. grandis* o número de famílias com cópias completas seja mais representativo.

5.3. A diversidade e similaridade genética de espécies do gênero *Eucalyptus* por meio das técnicas IRAP e REMAP

Os LTR-RTEs estão entre os maiores geradores de diversidade genética e, ao mesmo tempo, servem como ferramenta para detecção de alterações genômicas associadas a sua atividade. Além disso, em razão de seu dinamismo e dispersão, os LTR-RTEs podem ser utilizados como marcadores moleculares (Smýkal et al. 2006; Branco et al. 2007; Kumar et al. 2009; Carvalho et al. 2010; Kalendar 2011; Schulman et al. 2012), e os diferentes padrões de inserção compartilhados entre diferentes genótipos podem ser aplicados no estabelecimento de *pedigrees* e filogenias.

Este é o primeiro estudo em que se aplica o uso das técnicas IRAP e REMAP em eucalipto, e foram desenhados 18 *primers* para IRAP e 36 combinações em REMAP, que foram avaliadas em cinco espécies do gênero *Eucalyptus*. A maioria das combinações apresentou resultados passíveis de análise em todas as espécies estudadas, o que indica que as famílias de LTR-RTEs isoladas no presente trabalho estão dispersas dentro do gênero *Eucalyptus*, e que os mesmos podem ser utilizados para análises inter ou intraespecíficas, de diversidade, variabilidade e construção de mapas genéticos. Podem também ser aplicados na identificação de genótipos e de parentais (Ungerer et al. 2006; D'Onofrino et al. 2010; Carvalho et al. 2010; Campbell et al. 2011; Carvalho et al. 2012; Bento et al. 2013; Newaskar et al. 2013), além de análises fenotípicas, sistemáticas e ecologia molecular (Schulman et al. 2012).

As técnicas IRAP e REMAP resultaram em 1521 e 1910 bandas, respectivamente, totalizando 3431 fragmentos. O número de fragmentos gerados é alto se comparado a outras espécies arbóreas. Em laranja, marcadores IRAP geraram de 3 a 8 fragmentos e em REMAP apresentou em algumas combinações uma única banda (Biswas et al. 2010) e em *Pinus massoniana* com 153 fragmentos observados em IRAP (Fan et al. 2013).

O número de polimorfismos encontrados em média entre os *primers* foi maior na técnica IRAP (18) em relação ao REMAP (17,5) o que é relativamente comum como observados em outros vegetais como em caqui (Du et al. 2010) e trigo (Carvalho et. 2010; 2012). A técnica IRAP demonstrou ser mais eficiente em relação a REMAP, a despeito do grande número de microssatélites dispersos no genoma.

Nas ampliações IRAP foram observados fragmentos que chegavam a mais de 2,5 kb (Figuras 16 a 21). Já os resultados observados em REMAP geraram fragmentos em grande parte menores que 1 kb, o que indica uma grande proximidade entre as LTRs e as regiões microssatélite, sugerindo que os SSRs estão de maneira geral mais próximos que inserções em tandem de dois elementos (caso da amplificação detectada pelo IRAP). Além disso, a multiplicidade de bandas por espécie de eucalipto (Figura 21 a 24) sugere eventos de inserção das diferentes famílias de LTR-RTEs em regiões de microssatélite [(CT)₁₀G] e [(AG)₁₀T].

Em ambas as técnicas foi observada a geração de múltiplos fragmentos polimórficos, o que demonstra que os LTR-RTEs estão de fato em padrão disperso no genoma, assim como observado em caqui e alfafa (Du et al. 2010; Mandoulakani et al. 2012), e podem estar envolvidos em mecanismos de diversidade genética dentro do gênero *Eucalyptus*. O que corrobora com os dados observados através da quantificação das famílias por qPCR.

Embora todas as espécies de eucalipto apresentassem um número semelhante de fragmentos gerados e também de polimorfismos observados, *E. urophylla* apresentou um número de bandas amplificadas e polimórficas ligeiramente menor (2% e 1%, respectivamente – Tabelas 8 e 9), em relação as demais, o que pode indicar um menor número de LTR-RTEs em estudo no presente trabalho, o que também foi observado na quantificação das famílias por qPCR.

Ao comparar os dados gerados pelas técnicas IRAP e REMAP com resultados obtidos em análises RAPD realizadas no Centro de Análise Genômicas (CAGEN), em que se utilizou o mesmo material vegetal, foi constatado 23% a mais de fragmentos polimórficos (Aletéa C. Ávila Madacki e Celso Luis Marino, comunicação pessoal). Este tipo de observação já foi verificada em estudos da mesma ordem em outras plantas (Jawdat et al. 2010; Campbell et al. 2011; Santana et al. 2012). Assim, o uso das técnicas moleculares baseadas nos LTR-RTEs identificou um número de polimorfismos superior que possibilita análises diferenciadas entre as espécies.

Os *primers* desenvolvidos poderão auxiliar em futuros trabalhos realizados em plantas do gênero *Eucalyptus*, como já utilizado em outras espécies (Schulman 2007; Kalendar et al. 2011; Du et al. 2013) e também em espécies não relacionadas como realizado por Teo e colaboradores (2005) ao usar *primers* LTR-RTEs obtidos a partir uma família de Gramineae na identificação de cultivares de banana.

As espécies avaliadas neste trabalho são classificadas como pertencentes ao subgênero *Symphyomyrtus* (Steane et al. 2011). Este é o primeiro estudo em que estas cinco espécies do gênero *Eucalyptus* são comparadas quanto a sua distância genética e que a espécie *E. saligna* é incluída a este tipo de análises. A distribuição das espécies em relação a sua distância genética não foi semelhante a distribuição das espécies proposta por Steane e colaboradores (2011) ao usarem sequências conservadas de gene ribossomal.

A adição de *E. saligna* as análises possivelmente realinhou o posicionamento das espécies no dendrograma, cuja distância foi menor entre *E. tereticornis* e *E. urophylla*, com *E. grandis* posicionado-se mais distante. O uso de marcas menos conservadas por Steane e colaboradores (2011) também resultou em alterações da disposição dos cladogramas.

5.4. Marcadores Moleculares gerados a partir dos LTR-RTEs

Os polimorfismos identificados ao se utilizar as técnicas IRAP e REMAP foram avaliados para produzir marcadores tipo RBIP, que se baseiam na integração do elemento a um locus particular (Kalendar et al. 2011). O grande interesse em se produzir *primers* a partir da técnica RBIP, é por fornecerem dados sobre a dispersão e o padrão de inserção de um elemento em uma região genômica em particular (Kim et al. 2012; Baruch e Kashkush 2012; Yaakov et al. 2013).

Marcadores do tipo RBIP já foram utilizados para a identificação de cultivares de pêra japonesa (Kim et al. 2012), espécies em bancos de germoplasma de ervilha (Jing et al. 2012), acessos de mamona (Alipour et al 2013). As identificações são feitas pela presença ou ausência da inserção dos LTR-RTEs, o que ocorreu a milhões de anos atrás, no entanto é necessário o conhecimento das regiões flanqueantes a 5' e 3' LTRs (Kalendar 2011).

Os marcadores Ale2.1 e TGMR1 foram obtidos a partir de locus específicos a inserção das famílias RLC_egAle_2 e RLC_egGMR_1. O *primer* TGMR1 possibilitará a identificação espécie específica, visto que a presença de fragmento foi variável, com ausência

em indivíduos *E. grandis* e *E. saligna*. Já Ale2.1 poderá ser aplicado para se identificar variações intraespecíficas, as quais são de grande valia em análises populacionais e de diversidade.

Em café, marcadores RBIPs gerados a partir da técnica REMAP possibilitaram inferir em quais genótipos foi fixado o evento de inserção, isto é, antes ou após a diferenciação das espécies em diferentes acessos (Hamon et al. 2011). A partir do *primer* Umax1 foram observadas variações no tamanho dos locus específicos a inserção da família RLC_egMax_1, o que pode estar relacionado a variações quanto ao número de microsatélites [(AG)₁₀T] adjacentes, pois o mesmo foi obtido pela técnica REMAP. Pode-se sugerir que este marcador seja caracterizado como codominante por apresentar fragmentos de diferentes tamanhos para um locus único (Kumar e Hirochika 2001), sua aplicação se daria junto a análises de diversidade, populacionais, genótipo específica, além de análises filogenéticas por apresentar inserção irreversível (Kalendar et al. 2011).

O uso de marcas loco - específicas é facilmente executável por uma PCR seguida de uma visualização em gel de agarose, submetido a eletroforese ou até mesmo por meio de alto rendimento com o uso do TaqMan e chip de DNA (Alipour et al. 2013). Deste modo, podem auxiliar em estudos evolutivos e quanto a diferenciações inter e intraespecíficas.

5.5. Análise transcricional dos LTR-RTEs

Uma característica dos LTR-RTEs é sua regulação de forma *cis*, em que a região reguladora é associada a um sinal que irá ativar sua transcrição. No entanto muitos são os mecanismos que impedem que os LTR-RTEs permaneçam ativos, como a metilação do DNA e a modificação da cromatina, porém um “choque” por hibridação e/ou por estresse ambiental facilitam o seu rompimento (Michalak 2009; Lisch 2009; Kawakami et al. 2010; Wang et al. 2010; Yaakov et al. 2013).

Assim podemos demonstrar que os LTR-RTEs aqui estudados estão transcionalmente ativos e que podem influenciar na evolução de diversos componentes de regulação do genoma de eucalipto.

Neste estudo os experimentos realizados por RT-qPCR demonstraram que as famílias pertencentes às superfamílias *Copia* e *Gypsy*, possuem atividade transcricional em

diferentes tecidos de cinco espécies do gênero *Eucalyptus* e no híbrido F₁. Os membros da superfamília *Gypsy* apresentaram atividade mais uniforme entre os tecidos, embora os LTR-RTEs com maior expressão relativa sejam de famílias pertencentes a superfamília *Copia*.

Em carvalho (*Quercus suber*), foi identificada atividade transcricional diferenciada de LTR-RTEs da superfamília *Gypsy* nos tecidos e diferentes fases de desenvolvimento (embriões, raiz principal e secundárias, folhas jovens e velhas e pólen). A maior atividade foi identificada em raízes secundárias e em pólen, o que estava relacionado ao estresse celular e ambiental (Rocheta et al. 2012). No presente trabalho caule e raiz foram os tecidos com maior atividade transcricional nas famílias de LTR-RTEs, o que pode ser relacionado com uma alta atividade de divisão celular nestes tecidos, assim como já observado em diferentes famílias pertencentes a *Copia* e *Gypsy* em milho (Vicent 2010).

Já em folhas a atividade transcricional das famílias foi menor e em alguns casos ausente, como na família RLC_egAle_1 em *E. urophylla*. Em folhas é comum que os estresses sejam os principais desencadeadores de atividade transcricional. Já foi relatado que o estresse biótico pode ser um desencadeador da ativação de famílias pertencentes a *Copia* e *Gypsy* em folhas de cevada (Alzohairy et al. 2012).

A ativação de LTR-RTEs em híbridos é observada em monocotiledôneas e dicotiledôneas como arroz (Wang et al. 2010); tabaco (Petit et al. 2010) e girassol (Kawakami et al. 2010; Ungerer et al. 2013), este evento natural leva a uma reestruturação dos cromossomos por recombinação, que pode desencadear alterações genômicas, o que é um dos maiores fatores da ativação dos elementos nos híbridos (Shapiro 2010).

Folhas de híbrido F₁ proveniente do cruzamento de *E. grandis* e *E. urophylla* apresentaram um única família, RLC_egBianca_1, com atividade transcricional aumentada em relação seus genitores. Em caule o controle epigenético pode ter sido quebrado em três famílias: RLC_egIvana_1, RLC_egMax_1 e RLG_egCauli_1, que possuíram expressão aumentada no híbrido em relação aos genitores. A maioria das famílias (RLC_egMax_1, RLC_egBianca_1, RLC_egIvana_1, RLG_egTat_1 e RLG_egCauli_1) teve sua atividade transcricional aumentada em raízes secundárias do híbrido F₁. Contudo duas famílias, RLC_egAle_2 e RLG_egTat_1, apresentaram uma diminuição na atividade transcricional em F₁.

O uso de agentes estressantes na avaliação transcricional de LTR-RTEs é amplamente utilizado na literatura (Grandbastien et al. 2005; He et al. 2011; He et al. 2012), mas este é o primeiro estudo em que a atividade transcricional de LTR-RTEs é avaliada em plantas submetidas a estresse osmótico por PEG.

A indução de estresse osmótico somente não induziu o sinal de duas famílias, *RLC_egBianca_1* e *RLG_egCauli_1*. Já a família *RLC_egAle_1* teve sua atividade transcricional aumentada com o passar do tempo. As outras famílias apresentaram seu pico transcricional após 6 h de aplicação diminuindo em 24 h. Rodrigues e colaboradores (2013) identificaram o aumento gradual de expressão de uma aquaporina em *E. grandis* tratado com PEG.

6. Conclusão



- Os LTR-RTEs identificados apresentam polimorfismo de inserção ocupando grande parte do genoma de *E. grandis*, em que as famílias estão distribuídas em diferentes linhagens evolutivas.

- O grande número de fragmentos observados nas técnicas IRAP e REMAP é devido a sua distribuição no genoma das diferentes espécies de eucalipto, estando os LTR-RTEs próximos a regiões de microssatélites do que de outros LTR-RTEs relacionados.

- A atividade transcricional dos LTR-RTEs não apresentou um padrão de expressão nos diferentes tecidos, espécies de eucalipto e no híbrido, vindo a ser ativados por mecanismo de estresse abiótico.

- Por ser um trabalho pioneiro na caracterização de LTR-RTEs no genoma de *Eucalyptus* sp., a identificação de novas famílias e novas classes de elementos poderão ajudar ainda mais na compreensão da organização do genoma nestas espécies.

7. Referências



AGGARWAL, D.; KUMAR, A.; REDDY, M. S. Shoot organogenesis in elite clones of *Eucalyptus tereticornis*. **Plant Cell Tissue and Organ Culture**. v. 102, p. 45 – 52, 2010.

AGREN, J. A. e WRIGHT, S. I. Co-evolution between transposable elements and their hosts: a major factor in genome size evolution? **Chromosome Research**. v. 19, p. 777 – 786, 2011.

ALIPOUR, A.; TSUCHIMOTO, S.; SAKAI, H.; OHMIDO, N.; FUKUI, K. Structural characterization of *Copia*-type retrotransposons leads to insights into the marker development in a biofuel crop, *Jatropha curcas* L. **Biotechnology for Biofuels**. v.6, p. 129 – 142, 2013.

ALZOHAIRY, A. M.; YOUSEF, M. A.; EDRIS, S.; KERTI, B.; GYULAI, G.; SABIR, J. S. M.; RADWAN, N. A.; BAESHEN, M. N.; BAESHEN, N. A.; BAHIELDIN, A. Detection of LTR Retrotransposons Reactivation induced by *in vitro* Environmental Stresses in Barley (*Hordeum vulgare*) via RT-qPCR. **Life Science Journal**,v. 9 (4), p. 5019 – 5027, 2012.

ARAUJO, P. G.; ROSSI, M.; JESUS, E. M.; SACCARO JR, N. L.; KAJIHARA, D; MASSA, R.; FELIX, J. M.; DRUMMOND, R. D.; FALCO, M. C.; CHABREGAS, S. M.; ULIAN, E. C.; MENOSSI, M.; SLUYS, M.-A. V. Transcriptionally active transposable elements in recent hybrid sugarcane. **The Plant Journal**. v. 44, p. 707 – 717, 2005.

ATCHISON, E. Chromosome Number in the Myrtaceae. **American Journal of Botany**. v. 34 (3), p. 159 – 164, 1947.

BACCI JÚNIOR, M.; SOARES, R. B.S; TAJARA, E.; AMBAR, G.; FISCHER, C. N.; GUILHERME, I. R.; COSTA, E. P.; MIRANDA, V. F. O. Identification and frequency of transposable elements in *Eucalyptus*. **Genetics and Molecular Biology**. v. 28, n. 3, p. 634 – 639, 2005.

BARUCH, O. e KASHKUSH, K. Analysis of copy-number variation, insertional polymorphism, and methylation status of the tiniest class I (TRIM) and class II (MITE) transposable element families in various rice strains. **Plant Cell Reports**. v. 31, p. 885 – 893, 2012.

BENNETZEN, J. L. Transposable elements, gene creation and genome rearrangement in flowering plants. **Current Opinion in Genetics and Development**. v. 15, p. 621 – 627, 2005.

BENTO, M.; TOMÁS, D.; VIEGAS, W.; SILVA, M. Retrotransposons Represent the Most Labile Fraction for Genomic Rearrangements in Polyploid Plant Species. **Cytogenetic Genome Research**. doi: 10.1159/000353308. 2013.

BERTIOLI, D. J.; VIDIGAL, B.; NIELEN, S.; RATNAPARKHE, M. B.; LEE, T. H.; LEAL-BERTIOLI, S. C. M.; KIM, C.; GUIMARÃES, P. M.; SEIJO, G.; SCHWARZACHER, T.; PATERSON, A. H.; HESLOP-HARRISON, P.; ARAUJO, A. C. G. The repetitive component of the A genome of peanut (*Arachis hypogaea*) and its role in remodelling intergenic sequence space since its evolutionary divergence from the B genome. **Annals of Botany**. v. 112, p. 545 – 559. 2013.

BISWAS, M. K.; BAIG, M. N. R.; CHENG, Y.-J.; DENG, X.-X. Retrotransposon based genetic similarity within the genus *Citrus* and its relatives. **Genetic Resources and Crop Evolution**. v. 57, p. 963 – 972, 2010.

BOEKE, J. D. e CORCES, V. G. Transcription and reverse transcription of retrotransposons. **Annual Review of Microbiology**. v. 43, p. 403 - 434. 1989.

BRANCO, C. J. S.; VIEIRA, E. A.; MALONE, G.; KOPP, M. M.; MALONE, E.; BERNARDES, A.; MISTURA, C. C.; CARVALHO, F. I. F.; OLIVEIRA, C. A. IRAP and REMAP assessments of genetic similarity in rice. **Journal Applied Genetics**. v. 48, n. 2, p. 107 – 113, 2007.

BROOKER, M. I. H. e KLEINIG, D. A. Field guide to eucalypts. Ed. Inkata. v. 3. 1994.

BUSTIN, S. A. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. **Journal of Molecular Endocrinology**. v. 25, p. 169 – 193, 2000.

BUTELLI, E.; LICCIARDELLO, C.; ZHANG, Y.; LIU, J.; MACKAY, S.; BAILEY, P.; REFORGIATO-RECUPERO, G.; MARTINA, C. Retrotransposons Control Fruit-Specific,

Cold-Dependent Accumulation of Anthocyanins in Blood Oranges. **The Plant Cell**. v. 24, p. 1242 - 1255, 2012.

BUTI, M.; GIORDANI, T.; VUKICH, M.; PUGLIESI, C.; NATALI, L.; CAVALLINI, A. Retrotransposon-related genetic distance and hybrid performance in sunflower (*Helianthus annuus* L.). **Euphytica**. v. 192, p. 289 – 303, 2013.

CADLE-DAVIDSON, M. M. e OWENS, C. L. Genomic amplification of the *Gret1* retroelement in white-fruited accessions of wild *Vitis* and interspecific hybrids. **Theoretical Applied Genetics**. v. 116, p. 1079 – 1094, 2008.

CAMPBELL, B. C.; LEMARE, S.; PIPERIDIS, G.; GODWIN, I. D. IRAP, a retrotransposon-based marker system for the detection of somaclonal variation in barley. **Molecular Breeding**. v. 27, n.2, p. 193 – 206, 2011.

CARCAMO, C.; PROVVEDO, I.; ARROYO-GARCÍA, R. Detection of polymorphism in ancient Tempranillo clones (*Vitis vinifera* L.) using microsatellite and retrotransposon markers. **Iranian Journal of Biotechnology**. v. 8, n.1, 2010.

CARVALHO, A. M. e NAHUZS, M. A. R. The valuation of the *Eucalyptus grandis* x *urophylla* hybrid wood through the production of small dimension sawnwood, pulpwood and fuelwood. **Scientia Florestalis**. v. 59, p. 61 – 76, 2001.

CARVALHO, A.; GUEDES-PINTO, H.; MARTINS-LOPES, P.; LIMA-BRITO, J. Genetic variability of Old Portuguese bread wheat cultivars assayed by IRAP and REMAP markers. **Annal Applied Biology**. v. 156, p. 337 – 345, 2010.

CARVALHO, A.; GUEDES-PINTO, H.; LIMA-BRITO, J. E. Genetic Diversity in Old Portuguese Durum Wheat Cultivars Assessed by Retrotransposon-Based Markers. **Plant Molecular Biology Reporter**. v. 30, p. 578 – 589, 2012.

CASACUBERTA, E. e GONZÁLEZ, J. The impact of transposable elements in environmental adaptation. **Molecular Ecology**. v. 22, p. 1503 – 1517, 2013.

CASTRO, I.; D'ONOFRIO C.; MARTÍN, J. P.; ORTIZ, J. M.; LORENZIS, G.; FERREIRA, V.; PINTO-CARNIDE, O. Effectiveness of AFLPs and Retrotransposon-Based Markers for the Identification of Portuguese Grapevine Cultivars and Clones. **Molecular Biotechnology**. v. 52, p. 26 – 39, 2012.

CHANG, S.; PURYEAR, J.; CAIRNEY, J. A simple and efficient method for isolating RNA from pine trees. **Plant Molecular Biology Reprtr.** v.11, p. 113 – 116, 1993.

CHAW, S. M.; CHANG, C. C.; CHEN, H. L.; LI, W. H. Dating the Monocot–Dicot Divergence and the Origin of Core Eudicots Using Whole Chloroplast Genomes. **Journal Molecular Evolution**. v. 58, p. 424 – 441, 2004.

COSSU, R. M.; BUTI, M.; GIORDANI, T.; NATALI, L.; CAVALLINI, A. A computational study of the dynamics of LTR retrotransposons in the *Populus trichocarpa* genome. **Tree Genetics & Genomes**. v. 8, n.1, p. 61 – 75, 2011.

CUI, X.; JIN, P.; CUI, X.; GU, L.; LU, Z.; XUE, Y.; WEI, L.; QI, J.; SONG, X.; LUO, M.; AN, G.; CAO, X. Control of transposon activity by a histone H3K4 demethylase in Rice. **PNAS**. v. 110 (5), p. 1953 – 1958, 2013.

D'ONOFRIO, C.; DE LORENZIS, G.; GIORDANI, T.; NATALI, L.; CAVALLINI, A.; SCALABRELLI, G. Retrotransposon-based molecular markers for grapevine species and cultivars identification. **Tree Genetics & Genomes**. v. 6, p. 451 – 466, 2010.

DE MICHELE, R.; LOQUE, D.; LALONDE, S.; FROMMER, W. B. Ammonium and urea transporter inventory of the *Selaginella* and *Physcomitrella* genomes. **Frontiers in Plant Science**. v. 3, n. 62, 2012.

DOI, B. T.; BINKLEY, D.; STAPE, J. L. Does reverse growth dominance develop in old plantations of *Eucalyptus saligna*? **Forest Ecology and Management**. v. 259, p. 1815 – 1818, 2010.

DOMINGUES, D. S.; FARRO, A. P. C.; COSCRATO, V. E.; MELLO, E. J.; ODA, S.; MARINO, C. L. Identification of RAPD and Scar Marker related to early flowering in *Eucalyptus grandis*. **Ciência Florestal, Santa Maria**. v. 16, n. 3, p. 251 – 260, 2006.

DOMINGUES, D. S.; CRUZ, G. M. K.; METCALFE, C. J.; NOGUEIRA, F. T. S.; VICENTINI, R.; ALVES, C. S.; VAN SLUYS, M.-A. Analysis of plant LTR-retrotransposons at the finescale family level reveals individual molecular patterns. **BMC Genomics**. v. 13, n. 137, 2012.

DONG, Z.; YANG, X.; LI, Y.; ZHANG, K.; LOU, H.; AN, X.; DONG, L.; GU, Y. Q.; ANDERSON, L. D.; LIU, X.; QIN, H.; WANG, D. Haplotype Variation of Glu-D1 Locus and the Origin of Glu-D1d Allele Conferring Superior End-Use Qualities in Common Wheat. **PLOS ONE**. v. 8, e74859, 2013.

DOOLITTLE, W. F. e SAPIENZA, C. Selfish genes, the phenotype paradigm and genome evolution. **Nature**. v. 284, p. 601 – 603, 1980.

DU, J.; TIAN, Z.; HANS, C. S.; LATEN, H. M.; CANNON, S. B.; JACKSON, S. A.; SHOEMAKER, R. C.; MA, J. Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. **The Plant Journal**. v. 63, p. 584 – 598, 2010.

DU, X. Y.; HU, Q. N.; ZHANG, Q.L.; WANG, Y. B.; LUO, Z. R. Transferability of retrotransposon primers derived from Persimmon (*Diospyros kaki* Thunb.) across other plant species. **Genetics and Molecular Research**. v. 12, n. 2, p.1781 – 1795, 2013.

ECHENIQUE, V.; STAMOVA, B.; WOLTERS, P.; LAZO, G.; CAROLLO, V. L.; DUBCOVSKY, J. Frequencies of Ty1-*copia* and Ty3-*gypsy* retroelements within the Triticeae EST databases. **Theoretical Applied Genetic**. v. 104, p. 840 – 844, 2002.

EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**. v. 32, n.5, p. 1792 - 1797, 2004.

EL BAIDOURI, M. e PANAUD, O. Comparative Genomic Paleontology Across Plant kingdom Reveals The Dynamics Of TE-driven Genome Evolution. **Genome Biology and Evolution**. v. 5, p. 954 – 965, 2013.

ESTEP, M. C.; DE BARRY, J. D.; BENNETZEN, J. L. The dynamics of LTR retrotransposon accumulation across 25 million years of panicoid grass evolution. **Heredity**. v. 110, p. 194 – 204, 2013.

FAN, F.; WEN, X.; DING, G.; CUI, B. Isolation, identification, and characterization of genomic LTR retrotransposon sequences from masson pine (*Pinus massoniana*). **Tree Genetics & Genomes**. doi: 10.1007/s11295-013-0631-y. 2013.

FERREIRA, M. E. e GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análises genéticas**. 3. ed. Brasília: Embrapa-Cenargen, 1998. 220p.

FESCHOTTE, C.; JIANG, N.; WESSLER, S. R. Plant transposable elements: where genetics meets genomics. **Nature Reviews Genetics**. v. 3, p. 329 – 341, 2002.

FESCHOTTE, C. Transposable elements and the evolution of regulatory networks. **Nature Reviews Genetics**. v. 9, p. 397 – 405, 2008.

FINNEGAN, D. J. Eukaryotic transposable elements and genome evolution. **Trends in Genetics**. v. 5, p. 103 – 107, 1989.

FINNEGAN, D. J. Retrotransposons. **Current Biology**. v. 22, n.11, p. 432 – 437, 2012.

FLAVELL, A.J.; KNOX, M.R.; PEARCE, S.R.; ELLIS, T.H. Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analyses. **Plant Journal**. v. 16, p. 643 - 650, 1998.

GAIOTTO, F. A.; BRAMUCCI, M.; GRATTAPAGLIA, D. Estimation of outcrossing rate in a breeding population of *Eucalyptus urophylla* with dominant RAPD and AFLP markers. **Theoretical and Applied Genetics**. v. 95, p. 842 – 849, 1997.

GAN, S.; SHI, J.; LI, M.; WU, K.; WU, J.; BAI, J. Moderate-density molecular maps of *Eucalyptus urophylla* S. T. Blake and *E. tereticornis* Smith genomes based on RAPD markers. **Genetica**. v. 118, p. 59 - 67, 2003.

GAO, X.; HOU, Y.; EBINA, H.; LEVIN, H. L.; VOYTAS, D. F. Chromodomains direct integration of retrotransposons to heterochromatin. **Genome Research**. v. 18, p. 359 – 369, 2008.

GRANDBASTIEN, M.-A.; AUDEON, C.; BONNIVARD, E.; CASACUBERTA, J. M.; CHALHOUB, B.; COSTA, A.-P. P.; LE, Q. H.; MELAYAH, D.; PETIT, M.; PONCET, C.; TAM, S. M.; VAN SLUYS, M.-A.; MHIRI, C. Stress activation and genomic impact of *Tnt1* retrotransposons in Solanaceae. **Cytogenetics and Genome Research**. v. 110, p. 1 – 4, 2005.

GRATTAPAGLIA, D. e KIRST, M. *Eucalyptus* applied genomics: from gene sequences to breeding tools. **New Phytologist**. v. 179, p. 911 – 929, 2008.

GRATTAPAGLIA, D.; PLOMION, C.; KIRST, M.; SEDEROFF, R. R. Genomics of growth traits in forest trees. **Current Opinion of Plant Biology**. v. 12, p. 148 – 156, 2009.

GRATTAPAGLIA, D.; VAILLANCOURT, R. E. ; SHEPHERD, M.; THUMMA, B. R.; FOLEY, W.; KÜLHEIM, C.; POTTS, B. M.; MYBURG, A. A. Progress in Myrtaceae genetics and genomics: *Eucalyptus* as the pivotal genus. **Tree Genetics & Genomes**. v. 8. p. 463 – 508. 2012.

GRIFFITHS, A.; WESSLER, S.; LEWONTIN, R.; CARROLL, S. **Introdução à genética**, 9 ed. Editora Guanabara Koogan. 2008. 711p.

GORINSEK, B.; GUBENSEK, F.; KORDIS, D. Evolutionary Genomics of Chromoviruses in Eukaryotes. **Molecular Biology and Evolution**. v. 21, n. 5, p.781 – 798, 2004.

GOVAERTS, R.; SOBRAL, M.; ASHTON, P.; BARRIE, F.; HOLST, B. K.; LANDRUM, L. R.; MATSUMOTO, K.; MAZINE, F. F.; NIC LUGHADHA, E.; PROENÇA, C.; SOARES-SILVA, L. H.; WILSON, P. G.; LUCAS, E. J. World checklist of Myrtaceae. **Kew Publishing, Royal Botanic Gardens**, Kew. 2008.

HAMON, P.; DUROY, P.-O.; DUBREUIL-TRANCHANT, C.; COSTA, P. M. D'A.; DURET, C.; RAZAFINARIVO, N. J.; COUTURON, E.; HAMON, S.; KOCHKO, A.; PONCET, V.; GUYOT, R. Two novel *Ty1-Copia* retrotransposons isolated from coffee trees can effectively reveal evolutionary relationships in the *Coffea* genus (Rubiaceae). **Molecular Genetics Genomics**. v. 285, p. 447 – 460, 2011.

HANAHAN, D. Studies of transformation of *E. coli* with plasmids. **Journal Molecular Biology**. v. 156, p. 557 – 580, 1983.

HARFOUCHE, A.; MEILAN, R.; KIRST, M.; MORGANTE, M.; BOERJAN, W.; SABATTI, M.; MUGNOZZA, G. S. Accelerating the domestication of forest trees in a changing world. **Trends in Plant Science**. v. 17, n. 2, 2012.

HARPER, G.; HULL, R.; LOCKHART, B.; OLSZEWSKI, N. Viral sequences integrated into plant genomes. **Annual Review Phytopathology**. v.40, p. 119 – 36, 2002.

HAWKINS, J. S.; KIM, HR.; NASON, J. D.; WING, R. A.; WENDEL, J. F. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. **Genome Research**. v. 16, p. 1252 - 1261, 2006.

HE, P.; MA, Y.; DAÍ, H.; LI, L.; LIU, Y.; LI, H.; ZHAO, G.; ZHANG, Z.. Characterization of the Hormone and Stress-Induced Expression of *FaRE1* Retrotransposon Promoter in Strawberry. *Journal of Plant Biology*. v. 55, p.1–7, 2012.

HIROCHIKA, H.; SUGIMOTO, K.; OTSUKI, Y.; TSUGAWA, H.; KANDA, M. Retrotransposons of rice involved in mutations induced by tissue culture. **Proceeding of the National Academy of Sciences of the USA**. v. 93, p. 7783 – 7788, 1996.

HOLLISTER, J. D. e GAUT, B. S. Epigenetic silencing of transposable elements: A trade-off between expression reduced transposition and deleterious effects on neighboring gene. **Genome Research**. v. 19, p. 1419 – 1428, 2009.

HOSID, E.; BRODSKY, L.; KALENDAR, R.; RASKINA, O.; BELYAYEV, A. Diversity of Long Terminal Repeat Retrotransposon Genome Distribution in Natural Populations of the Wild Diploid Wheat *Aegilops speltoides*. **Genetics**. v. 190, p. 263 – 274, 2012.

HUANG, X.; LU, G.; ZHAO, Q.; LIU, X.; HAN, B. Genome-Wide Analysis of Transposon Insertion Polymorphisms Reveals Intraspecific Variation in Cultivated Rice. **Plant Physiology**. v. 148, p. 25 – 40, 2008.

ITO, H.; GAUBERT, H.; BUCHER, E.; MIROUZE, M.; VAILLANT, I.; PASZKOWSKI, J. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. **Nature**. v. 472, p. 115 – 119, 2011.

JÄÄSKELÄINEN, M.; CHANG, W.; MOISY, C.; SCHULMAN, A. H. Retrotransposon BARE displays strong tissue-specific differences in expression. **New Phytologist**. v. 200, p. 1000–1008, 2013.

JAWDAT, D.; AL-FAOURY, H.; AYYOUBI, Z.; AL-SAFADI, B. Molecular and ecological study of *Eryngium* species in Syria. **Biologia**. v. 65, n. 5, p. 796 - 804, 2010.

JERN, P.; SPERBER, O. O.; BLOMBERG, J. Use of Endogenous Retroviral Sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. **Retrovirology**. v. 2, n. 50, 2005.

JING, R.; KNOX, M. R.; LEE, J. M.; VERSHININ, A. V.; AMBROSE, M.; ELLIS, T. H. N.; FLAVELL, A. J. Insertional Polymorphism and Antiquity of *PDR1* Retrotransposon Insertions in *Pisum* Species. **Genetics**. v. 171, p. 741 - 752. 2005.

JING, R.; VERSHININ, A.; GRZEBYTA, J.; SHAW, P.; SMÝKAL, P.; MARSHALL, D.; AMBROSE, M. J.; ELLIS, T. H. N.; FLAVELL, A. J. The genetic diversity and evolution of field pea (*Pisum*) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. **BMC Evolutionary Biology**. v. 10, p. 44 – 64, 2010.

JING, R.; AMBROSE, M. A.; KNOX, M. R.; SMYKAL, P.; HYBL, P.; RAMOS, Á.; CAMINERO, C.; BURSTIN, J.; DUC, G.; VAN SOEST, L. J. M. W.; ŚWIĘCICKI, M.; PEREIRA, G.; VISHNYAKOVA, M.; DAVENPORT, G. F.; FLAVELL, A. J.; ELLIS, T. H. N. Genetic diversity in European *Pisum* germplasm collections. **Theoretical and Applied Genetics**. v. 125, p. 367 – 380, 2012.

JORDAN, E.; SAEDLER, H.; STARLINGER, P. Strong-Polar Mutations in the transferase gene of the Galactose Operon in *E. coli*. **Molecular Genetics**. v.100, p. 296 – 306, 1967.

JUKES, T. H. e CANTOR, C. R.. **Evolution of protein molecules**. In H. N. Munro, ed., Mammalian Protein Metabolism, Academic Press, New York. p. 21-132, 1969.

JUNGHANS, D. T.; ALFENAS, A. C.; BROMMONSCHENKEL, S. H.; ODA, S.; MELLO, E. J.; GRATTAPAGLIA, D. Resistance to rust (*Puccinia psidii* Winter) in Eucalyptus: mode of inheritance and mapping of a major gene with RAPD markers. **Theoretical and Applied Genetic**. v.108 (1), p. 175, 2003.

JURKA, J.; KAPITONOV, V. V.; PAVLICEK, A.; KLONOWSKI, P.; KOHANY, O.; WALICHIEWICZ, J. Repbase Update, a database of eukaryotic repetitive elements. **Cytogenetics and Genome Research**. v. 110, p. 462 – 467, 2005.

KALENDAR, R.; GROB, T.; REGINA, M.; SUONIEMI, A.; SCHULMAN, A.H. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. **Theoretical Applied Genetics** v. 98, p. 704 – 711, 1999.

KALENDAR, R. e SCHULMAN, A. H. IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. **Nature Protocols**. v.1, n. 5, p. 2478 – 2484, 2006.

KALENDAR, R.; FLAVELL, A. J.; ELLIS, T. H. N.; SJAKSTE, T.; MOISY, C.; SCHULMAN, A. H. Analysis of plant diversity with retrotransposon-based molecular markers. **Heredity**. v. 106, p. 520 – 530, 2011.

KALENDAR, R. The use of retrotransposon-based molecular markers to analyze genetic diversity. **Genetics and Breeding**. v. 48, p. 261 – 274, 2011.

KALEV, I. e HABECK, M. HHfrag: HMM-based fragment detection using HHpred. **Bioinformatics**. v. 27, p.3110–3116, 2011.

KASHKUSH, K.; FELDMAN, M.; LEVY, A. A. Gene Loss, Silencing and Activation in a Newly Synthesized Wheat Allotetraploid. **Genetics**. v. 160, p. 1651 – 1659, 2002.

KASHKUSH, K.; FELDMAN, M.; LEVY, A. A. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. **Nature Genetics**. v. 33, p. 102 – 106, 2003.

KASHKUSH, K. e KHASDAN, V. Large-Scale Survey of Cytosine Methylation of Retrotransposons and the Impact of Readout Transcription From Long Terminal Repeats on Expression of Adjacent Rice Genes. **Genetics**. v. 177, p. 1975 – 1985, 2007.

KAWAKAMI, T.; STRAKOSH, S. C.; ZHEN, Y.; UNGERER, M. C. Different scales of Ty1/*Copia*-like retrotransposon proliferation in the genomes of three diploid hybrid sunflower species. **Heredity**. v. 104, p. 341 – 350, 2010.

KENT, J. W. BLAT - The BLAST-Like Alignment Tool. **Genome Research**. v. 12, p. 656 – 664, 2002.

KHAN, A. M.; HATUN, S.; HOSSAIN, M. K.; RAHMAN, M. L.. Characterization of the Eucalyptus (*E. globulus*) leaves oil. **Journal of Bangladesh Chemical Society**. v. 25, n. 2, p. 97 - 100, 2012.

KIM, H.; TERAOKAMI, S.; NISHITANI, C.; KURITA, K.; KANAMORI, H.; KATAYOSE, Y.; SAWAMURA, Y.; SAITO, T.; YAMAMOTO, T. Development of cultivar-specific DNA markers based on retrotransposon-based insertional polymorphism in Japanese pear. **Breeding Science**. v. 62, p. 53 – 62, 2012.

KIMURA, M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. **Journal of Molecular Evolution**. v. 16, p. 111 - 120, 1980.

KRAITSHTEIN, Z.; YAAKOV, B.; KHASDAN, V.; KASHKUSH, K. Genetic and epigenetic dynamics of a retrotransposon after allopolyploidization of wheat. **Genetics**. v. 186, p. 801 – 812, 2010.

KOBAYASHI, S.; GOTO-YAMAMOTO, N.; HIROCHIKA, H. Retrotransposon-Induced Mutations in Grape Skin Color. **Science**. v. 304, p. 982, 2004.

KOKAEVA, Z. G.; ALESHIN, A. V.; BEREZOV, Y. I. Analysis of PDR1 Retrotransposon Insertions in the Pea (*Pisum Sativum* L.). **Moscow University Biological Sciences Bulletin**. v. 67, p. 48 – 51, 2012.

KOLANO, B.; BEDNARA, E.; WEISS-SCHNEEWEISS, H. Isolation and characterization of reverse transcriptase fragments of LTR retrotransposons from the genome of *Chenopodium quinoa* (Amaranthaceae). **Plant Cell Reports**. doi: 10.1007/s00299-013-1468-4. 2013.

KOONIN, E.V.; TATUSOV, R. L.; RUDD, K. E. Sequence similarity analysis of *Escherichia coli* proteins - Functional and evolutionary implications. **PNAS**. v. 92, p. 11921 – 11925, 1995.

KORIMBOCUS, J.; DANKS, C.; BOONHAM, N.; BARKER, I.; COATES, D. Improved detection of sugarcane leaf virus using a real-time fluorescent reverse-transcription polymerase chain reaction assay. **Phytopathology**. v. 90, p. 448 – 453, 2002.

KUMAR, A. e RZHETSKY, A. Evolutionary Relationships of Eukaryotic Kingdoms. **Journal of Molecular Evolution**. v. 42, p. 183 – 193, 1996.

KUMAR, A. e BENNETZEN, J. L. Plant Retrotransposons. **Annal Review of Genetics**. v. 33, p. 479 – 532, 1999.

KUMAR, A. e HIROCHIKA, H. Applications of retrotransposons genetic tools in plant biology. **Trends in Plant Science**. v. 6, p. 127 – 134, 2001.

KUMAR, P.; GUPTA, V.K.; MISRA, A. K.; MODI, D. R.; PANDEY, B. K. Potential of Molecular Markers in Plant Biotechnology. **Plant Omics Journal**. v. 2, n. 4, p. 141 – 162, 2009.

LADIGES, P. Y.; UDOVICIC, F.; NELSON, G. Australian biogeographical connections and the phylogeny of large genera in the plant family Myrtaceae. **Journal of Biogeography**. v. 30, p. 989 – 998, 2003.

LATEN, H. M. e GASTON, G. D. Plant Endogenous Retroviruses? A Case of Mysterious ORFs. In: Grandbastien, M.-A. e Casacuberta, J.M. (eds.). **Plant Transposable Elements**. 2012. p. 89 – 112.

LEITE, S. M. M.; BONINE, C. A.; MORI, E. S.; DO VALLE, C. F.; MARINO, C. L. Genetic Variability in a Breeding Population of *Eucalyptus urophylla* S.T. Blake. **Silvae Genetica**. v. 51, p. 5 – 6, 2002.

LIGHTBOURN, G. J.; JELESKO, J. G.; VEILLEUX, R. E. Retrotransposon-based markers from potato monoploids used in somatic hybridization. **Genome**. v. 50, p. 492 - 501. 2007.

LISCH, D. Epigenetic Regulation of Transposable Elements in Plants. **Annual Review of Plant Biology**. v. 60, p. 43 – 66, 2009.

LISCH, D. How important are transposons for plant evolution? **Genetics**. v. 3, p. 49 – 61, 2013.

LIVAK, K. J. e SCHMITTGEN, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta C_T}$ Method. **Methods**. v. 25, p. 402 – 408, 2001.

LLORENS, C.; FARES, M. A.; MOYA, A. Relationships of *gag-pol* diversity between Ty3/Gypsy and *Retroviridae* LTR retroelements and the three kings hypothesis. **BMC Evolutionary Biology**. v. 8, p. 276 – 295, 2008.

LLORENS, C.; FUTAMI, R.; RENAUD, G.; MOYA, A. Bioinformatic flowchart and database to investigate the origins and diversity of Clan AA peptidases. **Biology Direct**. v. 4, n. 3, 2009.

LOCKEY, C.; OTTO, E.; LONG, Z. Real-Time Fluorescence Detection of a Single DNA Molecule. **BioTechniques**. v. 24, p. 744 – 746, 1998.

LOPES, F. R.; JJINGO, D.; DA SILVA, C. R. M.; ANDRADE, A. C.; MARRACCINI, P.; TEIXEIRA, J. B.; CARAZZOLLE, M. F.; PEREIRA, G. A. G.; PEREIRA, L. F. P.; VANZELA, A. L. L.; WANG, L.; JORDAN, I. K.; CARARETO, C. M. A. Transcriptional

Activity, Chromosomal Distribution and Expression Effects of Transposable Elements in *Coffea* Genomes. **PLoS ONE**. v. 8(11), 175 - 191, 2013.

LOURENÇO, R. T. **Estrutura genômica de três megabases de DNA genômico (shotgun) de *Eucalyptus*: conteúdo nucleotídico, sequências repetitivas e genes**. 2004. 139f. Dissertação (Mestrado) Instituto de Biociências – Genética e Biologia Molecular, UNICAMP, Campinas. 2004.

LUO, S.; MACH, J.; ABRAMSON, B.; RAMIREZ, R.; SCHURR, R.; BARONE, P.; COPENHAVER, G.; FOLKERTS, O. The cotton centromere contains a Ty3-gypsy-like LTR retroelement. **PLoS One**. v. 7, e35261, 2012.

MA, J. e BENNETZEN, J. L. Rapid recent growth and divergence of Rice nuclear genomes. **PNAS**. v. 101, p. 12404–12410, 2004.

MANDOULAKANI, B. A.; PIRI, Y.; DARVISHZADEH, R.; BERNOOSI, I.; JAFARI, M. Retroelement Insertional Polymorphism and Genetic Diversity in *Medicago sativa* Populations Revealed by IRAP and REMAP Markers. **Plant Molecular Biology Reporter**. v. 30, n. 2, p. 286 – 296, 2012.

MANETTI, M. E.; ROSSI, M.; NAKABASHI, M.; GRANDBASTIEN, M.; VAN-SLUYS, M. -A. The *Tnt1* family member *Retrosol* copy number and structure disclose retrotransposon diversification in different *Solanum* species. **Molecular Genetics and Genomics**. v. 281, p. 261 - 271. 2009.

MARCHLER-BAUER, A.; LU, S.; ANDERSON, J. B.; CHITSAZ, F.; DERBYSHIRE, M. K.; DEWEESE-SCOTT, C.; FONG, J. H.; GEER, L. Y.; GEER, R. C.; GONZALES, N. R.; GWADZ, M.; HURWITZ, D. I.; JACKSON, J. D.; KE, Z.; LANCZYCKI, C. J.; LU, F.; MARCHLER, G. H.; MULLOKANDOV, M.; OMELCHENKO, M. V.; ROBERTSON, C. L.; SONG, J. S.; THANKI, N. T.; YAMASHITA, R. A.; ZHANG, D.; ZHANG, N.; ZHENG, C.; BRYANT, S. H. CDD: a Conserved Domain Database for the functional annotation of proteins. **Nucleic Acids Research**. v. 39, (1), p. 225 – 229.

MARCON, H.; DOMINGUES, D.; MARINO, C. Identification of potential transcriptionally active *Copia* LTR retrotransposons in *Eucalyptus*. **BMC Proceedings**. v. 5, n. 7, p. 164, 2011.

MARTIN-SANZ, A.; CAMINERO, C.; JING, R.; FLAVELL, A. J.; DE LA VEJA, M. P. Genetic diversity among Spanish pea (*Pisum sativum* L.) landraces, pea cultivars and the World *Pisum* sp. core collection assessed by retrotransposon-based insertion polymorphisms (RBIPs). **Spanish Journal of Agricultural Research**. v. 9, p. 166 – 178, 2011.

MARTINI, A. J. **O Plantador de *Eucalyptus*: a questão da preservação florestal no Brasil e o regaste documental do legado de Edmundo Navarro de Andrade**. 2004. 332f. Dissertação (Mestrado). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2004.

MCCARTHY, E. M. e MCDONALD, J. F. LTR_STRUC: a novel search and identification program for LTR retrotransposons. **Bioinformatics**. v. 19, p. 362 – 367, 2003.

MCCLINTOCK, B. The Origin and Behavior of Mutable Loci in Maize. **Proceeding of the National Academy of Sciences of the USA**. v. 36, n. 60, p. 344 – 355, 1950.

MERCHANT, A.; TAUSZ, M.; ARNDT, S. K.; ADAMS, M. A. Cyclitols and carbohydrates in leaves and roots of 13 *Eucalyptus* species suggest contrasting physiological responses to water deficit. **Plant. Cell & Environment**. v. 29, p. 2017 - 2029, 2006.

MIZRACHI, E.; HEFER, C. A.; RANIK, M.; JOUBERT, F.; MYBURG, A. A. *De novo* assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. **BMC Genomics**. v. 11, n. 681, 2010.

MICHALAK, P. Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. **Heredity**. v. 102, p. 45 – 50, 2009.

MOISY, C.; GARRISON, K. E.; MEREDITH, C. P.; PELS, F. Characterization of ten novel Ty1/*Copia*-like retrotransposon families of the grapevine genome. **BMC Genomics**. v. 9, n. 469, 2008.

MYBURG, A. A.; GRIFFIN, A. R.; SEDEROFF, R. R.; WHETTEN, R. W. Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F₁ hybrid based on a double pseudo-backcross mapping approach. **Theoretical and Applied Genetics**. v. 107, p. 1028 – 1042, 2003.

MYBURG, A.; GRATTAPAGLIA, D.; TUSKAN, G.; JENKINS, J.; SCHMUTZ, J.; MIZRACHI, E.; HEFER, C.; PAPPAS, G.; STERCK, L.; PEER, Y. V.; HAYES, R.; ROKHSAR, D. The *Eucalyptus grandis* Genome Project: Genome and transcriptome resources for comparative analysis of woody plant biology. **BMC Proceedings**. v.5, n. 7, p. 20, 2011.

NAIR, A. S.; TEO, C. H.; SCHWARZACHER, T.; HARRISON, P. H. Genome classification of banana cultivars from South India using IRAP markers. **Euphytica**. v. 144, p. 285 – 290, 2005.

NASCIMENTO, L. C.; LEPIKSON, J. N.; SALAZA, M. M.; CAMARGO, E. L. O.; MARQUES, W. L.; GONÇALVES, D. C.; VIDAL, R. O.; PEREIRA, G. A. G.; CARAZZOLLE, M. F. An integrated database of *Eucalyptus* spp. genome Project. **BMC Proceedings**. v. 5, n.7, p.170, 2011.

NEUMANN, P.; NAVRÁTILOVÁ, A.; KOBLÍŽKOVÁ, A.; KEJNOVSKÝ, E.; HŘIBOVÁ, E.; HOBZA, R.; WIDMER, R.; DOLEŽEL, J.; MACAS, J. Plant centromeric retrotransposons: a structural and cytogenetic perspective. **Mobile DNA**. v. 2, 2011.

NEWASKAR, G. S.; CHIMOTE, V. P.; MEHETRE, S. S.; JADHAV, A. S. Interspecific hybridization in *Gossypium* L.: characterization of progenies with different ploidy-confirmed multigenomic backgrounds. **Plant Breeding**. v. 132, p. 211 – 216, 2013.

NOVAES, E.; DROST, D. R.; FARMERIE, W. G.; PAPPAS JR, G. J.; GRATTAPAGLIA, D.; SEDEROFF, R. R.; KIRST, M. High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. **BMC Genomics**. v. 9, n. 312, 2008.

NOVIKOVA, O.; MAYOROV, V.; SMYSHLYAEV, G.; FURSOV, M.; ADKISON, L.; PISARENKO, O.; BLINOV, A. Novel clades of chromodomain-containing *Gypsy* LTR retrotransposons from mosses (Bryophyta). **The Plant Journal**. v. 56, p. 562 – 574, 2008.

NOVIKOVA, O. Chromodomains and LTR retrotransposons in plants. **Communicative & Integrative Biology**. v. 2, n.2, p.158 - 162, 2009.

OCHOGAVÍA, A. C.; SEIJO, J. G.; GONZÁLEZ, A. M.; PODIO, M.; DUARTE SILVEIRA, E.; MACHADO LACERDA, A. L.; TAVARES DE CAMPOS CARNEIRO, V.; ORTIZ, J. P.; PESSINO, S. C. Characterization of retrotransposon sequences expressed in inflorescences of apomictic and sexual *Paspalum notatum* plants. **Sex Plant Reprod**. v. 24, p. 231 – 246, 2011.

OHNO, S. So much „junk“ DNA in our genome. **Evolution of genetic systems. Brookhaven symposia in biology**. v. 23, 1972.

ORGEL, L. E. e CRICK, F. H. Selfish DNA: the ultimate parasite. **Nature**. v. 284, p. 604 – 607, 1980.

OUJJEH, B. e BENTOUATI, A. Chromosome numbers of the 59 species of *Eucalyptus* L'Herit. (Myrtaceae). **Caryologia**. v. 59 (3). p. 207 – 212, 2006.

PAGNOTTA, M. A.; MONDINI, L.; PORCEDDU, E. Quantification and organization of *WIS2-1A* and *BARE-1* retrotransposons in different genomes of *Triticum* and *Aegilops* species. **Molecular Genetics and Genomics**. v. 282, p. 245 – 255, 2009.

PAHALAWATTA, V.; DRUFFEL, K.; PAPPU, H. A new and distinct species in the genus *Caulimovirus* exists as an endogenous plant pararetroviral sequence in its host, *Dahlia variabilis*. **Virology**. v. 376, p. 253 – 257, 2008.

PALHARES, A.C.; RODRIGUES-MORAIS, T.B.; SLUYS, M.A.; DOMINGUES, D.S.; MACCHERONI, W.; JORDÃO, H.; SOUZA, A.P.; MARCONI, T.G.; MOLLINARI, M.; GAZAFFI, R.; GARCIA, A.A.F.; VIEIRA, M.L.C. A novel linkage map of sugarcane with evidence for clustering of retrotransposon-based markers. **BMC Genetics**. v. 13, p. 51, 2012.

PANDOTRA, P.; HUSAIN, M. K.; RAM, G.; GUPTA, S.; GUPTA, A. P. Retrotransposon based genetic status of North-West Himalayan *Zingiber officinale* revealed high heterogeneity. **Journal of Plant Biochemistry and Biotechnology**. doi 10.1007/s13562-013-0196-8. 2013.

PARK, M.; JO, S.; KWON, J.-K.; PARK, J.; AHN, J. H.; KIM, S.; LEE, Y.-H.; YANG, T.-J.; HUR, C.-G.; KANG, B.-C.; KIM, B.-D.; CHOI, D. Comparative analysis of pepper and tomato reveals euchromatin expansion of pepper genome caused by differential accumulation of Ty3/*Gypsy*-like elements. **BMC Genomics**. v. 12, n. 85, 2011.

PEREDO, E. L.; ARROYO-GARCIA, R.; REVILLA, M. A. Epigenetic changes detected in micropropagated hop plants. **Journal of Plant Physiology**. v. 166, p. 1101 – 1111, 2009.

PETIT, M.; GUIDAT, C.; DANIEL, J.; DENIS, E.; MONTORIOL, E.; BUI, Q. T.; LIM, K. Y.; KOVARIK, A.; LEITCH, A. R.; GRANDBASTIEN, M.-A.; MHIRI, C. Mobilization of retrotransposons in synthetic allotetraploid tobacco. **New Phytologist**. v.186, p. 135 – 147, 2010.

PICAULT, N.; CHAPARRO, C.; PIEGU, B.; STENGER, W.; FORMEY, D.; LLAURO, C.; DESCOMBIN, J.; SABOT, F.; LASSERRE, E.; MEYNARD, D.; GUIDERDONI, E.;

PIEDNOËL, M.; CARRETE-VEJA, G.; RENNERS, S. Characterization of the LTR retrotransposon repertoire of a plant clade of six diploid and one tetraploid species. **The Plant Journal**. doi: 10.1111/tpj.12233. 2013.

POCZAI, P.; VARGA, I.; LAOS, M.; CSEH, A.; BELL, N.; VALKONEN, J. P. T.; HYVÖNEN, J. Advances in plant gene-targeted and functional markers: a review. **Plant Methods**. v. 9, n. 6, 2013.

PUNTA, M.; COGGILL, P. C.; EBERHARDT, R. Y.; MISTRY, J.; TATE, J.; BOURSNEILL, C.; PANG, N.; FORSLUND, K.; CERIC, G.; CLEMENTS, J.; HEGER, A.; HOLM, L.; SONNHAMMER, E. L. L.; EDDY, S. R.; BATEMAN, A.; FINN, R. D. The Pfam protein families database. **Nucleic Acids Research**. v. 40, p.290 – 301, 2012.

RIESEBERG, L. H.; RAYMOND, O.; ROSENTHAL, D. M.; LAI, Z.; LIVINGSTONE, K.; NAKAZATO, T.; DURPHY, J. L.; SCHWARZBACH, A. E.; DONOVAN, L. A.; LEXER, C. Major Ecological Transitions in Wild Sunflowers Facilitated by Hybridization. **Science**. v. 301, p. 1211 – 1218, 2003.

ROCHA, M. E. N. e SANTOS, C. L. O uso comercial e popular do *Eucalyptus globulus* labill – Myrtaceae. **Saúde e Ambiente em Revista**. v. 2, n. 2, p. 23 – 34, 2008.

ROCHETA, M.; CARVALHO, L.; VIEGAS, W.; MORAIS-CECÍLIO, L. *Corky*, a Gypsy-like retrotransposon is differentially transcribed in *Quercus suber* tissues. **BMC Research Notes**. v. 5, n. 432, 2012.

RODRIGUES, J.; MEIER, D.; FAIX, O.; PEREIRA, H. Determination of tree to tree variation in syringyl/guaiacyl ratio of *Eucalyptus globulus* wood lignin by analytical pyrolysis. **Jounal Analitical and Applied Pyrolysis**. v. 48, p. 121 – 128, 1999.

RODRIGUES, M. R.; BRAVO, J. P.; SASSAKI, F. T.; SEVERINO, F. E.; MAIA, I. G. The tonoplast intrinsic aquaporin (TIP) subfamily of *Eucalyptus grandis*: Characterization of EgTIP2, a root-specific and osmoticstress-responsive gene. **Plant Science**. v. 213, p. 106 – 113, 2013.

ROHLF, F. J. NTSYS-pc: numerical taxonomy and multivariate analysis system. **Exeter Software, Setauket, NY**. 2000

ROSSI, M.; ARAUJO, P. G.; SLUYS, M. -A. V. Survey of transposable elements in sugarcane expressed sequence tags (ESTs). **Genetic Molecular Biology**. v. 24 (1-4). p. 141 – 146, 2001.

RUIJTER, J. M.; RAMAKERS, C.; HOOGAARS, W. M. H.; KARLEN, Y.; BAKKER, O.; VAN DEN HOFF, M. J. B.; MOORMAN, A. F. M.. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. **Nucleic Acids Research**.v. 37, p. 1 – 12, 2009.

SABOT, F. e SCHULMAN, A. H. Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. **Heredity**. v. 97, p. 381 – 388, 2006.

SAITOU, N. e NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular Biology Evolution**. v. 4, p.406 – 425, 1987.

SALAZAR, M. M.; NASCIMENTO, L. C.; CAMARGO, E. L. O.; GONÇALVES, D. C.; LEPIKSON NETO, J.; MARQUES, W. L.; TEIXEIRA, P. J. P. L.; MIECZKOWSKI, P.; MONDEGO, J. M. C.; CARAZZOLLE, M. F.; DECKMANN, A. C.; PEREIRA, G. A. G. Xylem transcription profiles indicate potential metabolic responses for economically relevant characteristics of *Eucalyptus* species. **BMC Genomics**. v. 14, n. 201, 2013.

SANMIGUEL, P.; GAUT, B. S.; TIKHONOV, A.; NAKAJIMA, Y.; BENNETZEN, J. L. The paleontology of intergene retrotransposons of maize. **Nature Genetics**. v. 20, p. 43 – 45. 1998.

SANMIGUEL, P. e BENNETZEN, J. L. Evidence that a Recent Increase in Maize Genome size was caused by the Massive amplification of Intergene Retrotransposons. **Annals of Botany**. v. 82A, p. 37 – 44, 1998.

SANTANA, M. F.; ARAÚJO, E. F.; SOUZA, J. T.; MIZUBUTI, E. S. G.; QUEIROZ, M. V. Development of molecular markers based on retrotransposons for the analysis of genetic variability in *Moniliophthora perniciosa*. **European Journal Plant Pathology**. v. 134, p. 497 – 507, 2012.

SARTORETTO, L. M.; SALDANHA, C. W.; CORDER, M. P. M. Transformação genética: estratégias e aplicações para o melhoramento genético de espécies florestais. **Ciência Rural**. v. 38, p. 861 - 871, 2008.

SCHULMAN, A. H.; FLAVELL, A. J.; ELLIS, T. H. The application of LTR retrotransposons as molecular markers in plants. **Methods in Molecular Biology**. v. 260, p. 145 – 173, 2004.

SCHULMAN, A. H. Molecular markers to assess genetic diversity. **Euphytica**. v.158, p. 313 – 321, 2007.

SCHULMAN, A. H.; FLAVELL, A. J.; PAUX, E.; ELLIS, T. H. N. The Application of LTR Retrotransposons as Molecular Markers in Plants. In: Bigot, Y. (ed.). **Mobile Genetic Elements**. Humana Press. 2012. v. 859, n.7, p. 115 – 154.

SETTA, N.; METCALFE, C. J.; CRUZ, G. M. Q.; OCHOA, E. A.; VAN SLUYS, M.-A. Noise or Symphony: Comparative Evolutionary Analysis of Sugarcane Transposable Elements with Other Grasses. In: Grandbastien, M.-A. e Casacuberta, J.M. (eds.). **Plant Transposable Elements**. Topics in Current Genetics. v. 24, p. 169 - 192, 2012.

SHAPIRO, J. A. Mobile DNA and evolution in the 21st century. **Mobile DNA**. v. 1, p. 1 – 14, 2010.

SHIMAZAKI, M.; FUJITA, K.; KOBAYASHI, H.; SUZUKI, S. Pink-Colored Grape Berry Is the Result of Short Insertion in Intron of Color Regulatory Gene. **PLoS ONE**. v. 6, n. 6, 2011.

SILVA, F.A.S. e AZEVEDO, C.A.V. Principal components analysis in the software assistat-statistical attendance. In: World Congress on Computers in Agriculture, 7. American Society of Agricultural and Biological Engineers, Reno-NV-USA. 2009.

SMÝKAL, P. Development of an efficient retrotransposon-based fingerprinting method for rapid pea variety identification. **Journal of Applied Genetics**. v. 47, n. 3, p. 221 – 230, 2006.

SOARES, E. 2010. Eucalipto: novas técnicas, grandes lucros. Disponível em: <http://painelflorestal.com.br/noticia-7543-calipto+novas+tecnicas+grandes+lucros.htm>. Acesso em 10 maio de 2010.

SOLEIMANI, V.D.; BAUM, B.R.; JOHNSON, D.A. Quantification of the retrotransposon *BARE-1* reveals the dynamic nature of the barley genome. **Genome**. v. 49, p. 389 – 396, 2006.

SQUIRES, J.; GILLESPIE, T.; SCHOELZ, J. E.; PALUKAITIS, P. Excision and Episomal Replication of Cauliflower mosaic virus Integrated into a Plant Genome. **Plant Physiology**. v. 155, p. 1908 – 1919, 2011.

STATON, E. S.; UNGERER, M. C.; MOORE, R. C. The Genomic Organization Of *Ty3* / *Gypsy* – Like Retrotransposons in *Helianthus* (Asteraceae) Homoploid Hybrid Species. **American Journal of Botany**. v. 96, n. 9, p. 1646 – 1655, 2009.

STEANE, D. A.; WEST, A. K.; POTTS, B. M.; OVENDEN, J. R.; REID, J. B. Restriction Fragment Length Polymorphisms in Chloroplast DNA from Six Species of *Eucalyptus*. **Australian Journal Botany**. v. 39, p. 399 – 414, 1991.

STEANE, D. A.; NICOLLE, D.; SANSALONI, C. P.; PETROLI, C. D.; CARLING, J.; KILIAN, A.; MYBURG, A. A.; GRATTAPAGLIA, D.; VAILLANCOURT, R. E. Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (Myrtaceae) using high-throughput, genome-wide genotyping. **Molecular Phylogenetics and Evolution**. v. 59, p. 206 - 224, 2011.

SMÝKAL, P. Development of an efficient retrotransposon-based fingerprinting method for rapid pea variety identification. **Journal of Applied Genetics**. v. 47, n. 3, p. 221 – 230, 2006.

STEINBAUEROVÁ, V.; NEUMANN, P.; NOVÁK, P.; MACAS, J. A widespread occurrence of extra open reading frames in plant *Ty3/Gypsy* retrotransposons. **Genetica**. v. 139, p. 1543 – 1555, 2011.

SUGIMOTO, K.; TAKEDA, S.; HIROCHIKA, H. MYB-Related Transcription Factor NtMYB2 Induced by Wounding and Elicitors is a Regulator of the Tobacco Retrotransposon *Tto1* and Defense-Related Genes. **The Plant Cell**. v. 12, p. 2511 – 2527, 2000.

SUONIEMI, A.; TANSKANEM, J.; SCHULMAN, A. H. *Gypsy*-like retrotransposons are widespread in the plant kingdom. **The Plant Journal**. v. 13, n.5, p. 699 – 705, 1998.

TAKEDA, S., SUGIMOTO, K., OTSUKI, H.; HIROCHIKA, H. A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon *Tto1* is involved in responsiveness to

tissue culture, wounding, methyl jasmonate and fungal elicitors. **Plant Journal**.18, 383 - 393, 1999.

TAMURA, K.; STECHER, G.; PETERSON, D.; FILIPSKI, A.; KUMAR, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. **Molecular Biology and Evolution**. v.30, p. 2725 – 2729, 2013.

TANHUANPAA, P.; KALENDAR, R.; SCHULMAN, A. H.; KIVIHARJU, E. A major gene for grain cadmium in oat (*Avena sativa* L.). **Genome**. v. 50, p. 588 – 594, 2007.

TENAILLON, M. I.; HOLLISTER, J. D.; GAUT, B. S. A triptych of the evolution of plant transposable elements. **Trends in Plant Science**. v. 15, p. 471 – 478, 2010.

TEO, C. H.; TAN, S. H.; HO, C. L.; FARIDAH, Q. Z.; OTHMAN, Y. R.; HESLOP-HARRISONS, J. S.; KALENDAR, R.; Schulman, A. H. Genome Constitution and Classification Using Retrotransposon-Based Markers in the Orphan Crop Banana. **Journal of Plant Biology**. v.48, n. 1, p. 96 – 105, 2005.

UNGERER, M. C.; STRAKOSH, S. C.; STIMPSON, K. M. Proliferation of *Ty3/Gypsy*-like retrotransposons in hybrid sunflower taxa inferred from phylogenetic data. **BMC Biology**. v. 7, n. 40, 2009.

UNGERER, M. K. e KAWAKAMI, T. Transcriptional Dynamics of LTR Retrotransposons in Early Generation and Ancient Sunflower Hybrids. **Genome Biology and Evolution**. v. 5, n. 2, p. 329 – 337, 2013.

VICIENT, C. M.; JÄÄSKELÄINEN, M. J.; KALENDAR, R.; SCHULMAN, A. H. Active Retrotransposons Are a Common Feature of Grass Genomes. **Plant Physiology**. v. 125 (3). p. 1283 – 1292, 2001.

VICIENT, C. M. Transcriptional activity of transposable elements in maize. **BMC Genomics**. v.11, p. 601, 2010.

VILLAR, E.; KLOPP, C.; NOIROT, C.; NOVAES, E.; KIRST, M.; PLOMION, C.; GION, J. M. RNA-Seq reveals genotype-specific molecular responses to water deficit in *Eucalyptus*. **BMC Genomics**, v. 12, n. 538, 2011.

VITTE, C.; ISHII, T.; LAMY, F.; BRAR, D.; PANAUD, O. Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). **Molecular Genetics Genomics**. v. 272, p. 504 – 511, 2004.

VOYTAS, D. F.; KONIECZNY, A.; CUMMINGS, M. P.; AUSUBEL, F. M. The Structure, Distribution and Evolution of the *Tal* Retrotransposable Element family of *Arabidopsis thaliana*. **Genetics**. v. 126, p. 713 – 721, 1990.

VUKICH, M.; GIORDANI, T.; NATALI, L.; CAVALLINI, A. *Copia* and *Gypsy* retrotransposons activity in sunflowers (*Helianthus annuus* L.). **BMC Plant Biology**. v. 9, p. 150 – 162, 2009.

WANG, H. e LIU, J. S. LTR retrotransposon landscape in *Medicago truncatula*: more rapid removal than in Rice. **BMC Genomics**. v. 9, n. 382, 2008.

WANG, H.-Y.; TIAN, Q.; MA, Y.-Q.; WU, Y.; MIAO, G.-J.; MA, Y.; CAO, D.-H; WANG, X. -L.; LIN, C.; PANG, J.; LIU, B. Transpositional reactivation of two LTR retrotransposons in rice-*Zizania* recombinant inbred lines (RILs). **Hereditas**. v. 147, p. 264 – 277, 2010.

WAWRZYNSKI, A.; ASHFIELD, T.; CHEN, N. W. G.; MAMMADOV, J.; NGUYEN, A.; PODICHETI, R.; CANNON, S. B.; THAREAU, S.; AMELINE-TORREGROSA, C.; CANNON, E.; CHACKO, B.; COULOUX, A.; DALWANI, A.; DENNY, R.; DESHPANDE, S.; EGAN, A. N.; GLOVER, N.; HOWELL, S.; ILUT, D.; LAI, H.; DEL CAMPO, S. M.; METCALF, M.; O'BLENESS, M.; PFEIL, R. E.; RATNAPARKHE, M. B.; SAMAIN, S.; SANDERS, I.; SÉGURENS, B.; SÉVIGNAC, M.; SHERMAN-BROYLES, S.; TUCKER, D. M.; YI, M.; DOYLE, J. J.; GEFFROY, V.; ROE, M. V.; MAROOF, M. A. S.; YOUNG, N. A.; INNES, R. A. Replication of Nonautonomous Retroelements in Soybean Appears to Be Both Recent and Common. **Plant Physiology**. v. 148, p. 1760 - 1771, 2008.

WHITNEY, K.D. e GARLAND, T. Jr. Did Genetic Drift Drive Increases in Genome Complexity? **PLoS Genetics**. v. 6, n. 8, 2010.

WICKER, T. e KELLER, B. Genome-wide comparative analysis of *copia* retrotransposons in Triticeae, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual *copia* families. **Genome Research**. v.17, n. 7, p. 1072 – 1081, 2007.

WICKER, T.; SABOT, F.; HUA-VAN, A.; BENNETZEN, J.L.; CAPY, P.; CHALHOUB, B.; FLAVELL, A.; LEROY, P.; MORGANTE, M.; PANAUD, O.; PAUX, E.; SANMIGUEL, P.; SCHULMAN, A. H. A unified classification system for eukaryotic transposable elements. **Nature Reviews Genetics**. v. 8, n. 12, p. 973 – 982, 2007.

WILHELM, M. e WILHELM, F.-X. Reverse transcription of retroviruses and LTR retrotransposons. **Cellular and Molecular Life Sciences**. v. 58, p. 1246 – 1262, 2001.

WRIGHT, D. A. e VOYTAS, D. F. *Athila4* of *Arabidopsis* and *Calypso* of Soybean Define a Lineage of Endogenous Plant Retroviruses. **Genome Research**. v. 12, p. 122 – 131, 2002.

XU, Z. e WANG, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. **Nucleic Acids Research**. v. 35, p. 265 – 268, 2007.

YAAKOV, B. e KASHKUSH, K. Mobilization of Stowaway-like mites in newly formed allohexaploid wheat species. **Plant Molecular and Biology**. v. 80, p. 419 – 427, 2012.

YAAKOV, B.; BEN-DAVID, S.; KASHKUSH, K. Genome-Wide Analysis of Stowaway-Like *MITEs* in Wheat Reveals High Sequence Conservation, Gene Association, and Genomic Diversification. **Plant Physiology**. v. 161, p. 486 – 496, 2013.

YAAKOV, B.; MEYER, K.; BEN-DAVID, S.; KASHKUSH, K. Copy number variation of transposable elements in Triticum–*Aegilops* genus suggests evolutionary and revolutionary dynamics following allopolyploidization. **Plant Cell Reports**. doi:10.1007/s00299-013-1472-8. 2013.

YIN, H.; LIU, J.; XU, Y.; LIU, X.; ZHANG, S.; MA, J.; DU, J. *TARE1*, a Mutated *Copia*-Like LTR Retrotransposon Followed by Recent Massive Amplification in Tomato. **PLoS ONE**. v. 8, n. 7, 2013.

ZHAO, M. e MA, J. Co-evolution of plant LTR-retrotransposons and their host genomes. **Protein Cell**. v. 4, p. 493 – 501, 2013.

ZHOU, J.; GONG, H.; YANG, T. J.; MENG, J. Retrotransposons - a Major Driving Force in Plant Genome Evolution and a Useful Tool for Genome Analysis. **Journal Crop Science Biotechnology**. v. 12, n. 1, p. 1 – 8, 2009.