Identification of potential transcriptionally active Copia LTR retrotransposons in *Eucalyptus*

Helena Marcon1,2*, Douglas Domingues1*, Celso Marino1*

*From* IUFRO Tree Biotechnology Conference 2011: From Genomes to Integration and Delivery
Arraial d’Ajuda, Bahia, Brazil. 26 June - 2 July 2011

**Background**

Long Terminal Repeat retrotransposons (LTR-RTs) represent the most abundant genomic component in all plant genomes thus far investigated. They are transposable elements that replicate through a “copy/paste” mechanism that relies on reverse transcription and integration of a RNA intermediate. Plant LTR-RTs can be divided in two major superfamilies: *Copia* and *Gypsy* [1]. LTR-RTs have impact on genome size variation, as well as in the expression of adjacent genes in their host genomes, providing a “genomic plasticity” [2]. Their transcription was believed to be extremely repressed in plants. However, despite their potential mutagenic and deleterious effects, LTR-RTs were proven to be transcriptionally active in several plant species [3].

*Eucalyptus* is one of the most commercially important forest genus in the world, due to their superior growth, broad adaptability and multipurpose wood properties. Most molecular studies in *Eucalyptus* are focused on cellulose production and wood development, and there are few works on genome composition, structure and evolution. *Pinus* and *Populus*, the tree genera with most available genomic resources, have several works analyzing their repertoire of LTR-RTs [i.e 4, 5], but only one study characterized LTR-RTs in *Eucalyptus* [6], with no detailed manual checking or phylogenetic analysis. Here, we used FOREST database as a starting point to identify transcriptionally active *Copia* LTR-RTs in *Eucalyptus*, that were further analyzed regarding their *in silico* expression, evolutionary diversity, and distribution in public genomic databases.

**Methods**

A previous survey with 88 *Copia* LTR-RTs from diverse plants defined six major common evolutionary *Copia*-lineages [7]. The 22 *Arabidopsis thaliana* families analyzed in that study were used as queries to the identify *Eucalyptus* EST sequences related to *Copia* elements in FOREST database [8], by tBLASTx (e-value >1e-50). Sequences were then analyzed in RepBase [9] to confirm their similarity to *Copia* LTR-RTs. *Eucalyptus* ESTs with >200bp of copia-like retrotransposon fragments were used to identify complete copies in *Eucalyptusgrandis* genome v.1.0 in a BLASTn search (identity >80%; in a region >250bp). We picked up 10000bp surrounding the aligned region, that were analyzed using LTR-Finder [10] and LTR_STRUC [11]. Full-length LTR-RTs were then used as queries in GenBank to retrieve related *Eucalyptus* EST sequences (>200bp; >80% identity). Phylogenetic analyses using the reverse transcriptase of these elements (alignment in MUSCLE, Maximum Likelihood method, bootstrap 1000 replicates) were done using MEGA 5.01 [12].

**Results**

Stem, calli and seedlings were the cDNA libraries from FOREST database with most EST sequences, in this *Copia* LTR-RT search. We identified 20 consensus sequences (total: 36 ESTs) from 3 tissues, roots, leaves and flower-buds. We also identified 29 ESTs in GenBank from xylem, root apex and cold-stressed plants (Table 1). Using EST data, we identified six full-length retrotransposons families that had different copy number in the *Eucalyptus* genome, estimated by BLAST searches (cutoff 1e-50). Copy number ranged from 24 to 262 (Table 1). Phylogenetic analyses showed that they are members of the *Ale*, *Angela*, *GMR* and *Ivana* evolutionary lineages (figure 1). *Ale* was the evolutionary...
lineage encompassing families with highest and lowest copy number (Table 1).

**Conclusion**

In summary, the present data demonstrate the potential impact of future studies about functional and genomic analysis of LTR-RTs in *Eucalyptus*. This is the first characterization of full-length Copia LTR-RTs families in *Eucalyptus* genome with potential transcriptional activity, giving insights about phylogenetic diversity and copy number variation of retrotransposons in this tree.

**Acknowledgements**

HSM was supported by a fellowship from Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and CNPq.

**Author details**

1 Departamento de Genética, UNESP, Botucatu, São Paulo, Brazil. 2 IAPAR, Londrina, Paraná, Brazil.

**Published:** 13 September 2011

**References**

1. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Challoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH: A unified classification system for eukaryotic transposable elements. *Nat Rev* 2007, 8:973-982.
2. Du J, Tian Z, Hans CS, Latten HM, Cannon SB, Jackson SA, Shoemaker RC, Ma J: Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. *Plant J* 2010, 63:584-598.
3. Vicent CM, Jaaskelainen MJ, Kalendar R, Schulman AH: Active retrotransposons are a common feature of grass genomes. *Plant Phys* 2001, 125:1285-1292.
4. L’Homme Y, Séguin A, Tremblay FM: Different classes of retrotransposons in coniferous spruce species. *Genome* 2000, 43(6):1084-1089.
5. Zhou F, Xu Y: RepPop: a database for repetitive elements in *Populus trichocarpa*. *BMC Genomes* 2009, 10:14-23.
6. Bacc Júnior M, Soares RRS, Tajara E, Ambar G, Fischer CN, Guilherme IR, Costa EP, Miranda VFO: Identification and frequency of transposable elements in *Eucalyptus*. *Genet Mol Biol* 2005, 28(3):634-639.
7. Wicker T, Keller B: Genome-wide comparative analysis of retrotransposons in *Triticeae*, *rice*, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families. *Genome Research* 2007, 17:1072-1081.
8. Vicentini R, Sastaki FT, Gimenes MA, Maia IG, Menossi M: In silico evaluation of the *Eucalyptus* transcriptome. *Gen Mol Biol* 2005, 28:487-495.
9. Jurka J, Kapitonov VV, Pavlicek A, Kluczewski P, Kohany O, Walichiewicz J: Repbase Update, a database of eukaryotic repetitive elements, *CytoGeneticcs and Genome Research* 2005, 110:462-467.
10. Xu Z, Wang H: *LTR_FINDER*: an efficient tool for the prediction of full-length LTR retrotransposons. *Nuc Acids Res* 2007, 35:265-268.

---

**Table 1** Overall features of LTR-RTs analyzed.

| Family       | Lineage | Genomic copy number | FOREST cDNA libraries | GenBank cDNA libraries |
|--------------|---------|----------------------|-----------------------|------------------------|
| RTE_copia_Eu_1 | Ale     | 28                   | seedlings             | xylem                  |
| RTE_copia_Eu_2 | Ale     | 262                  | roots, leaves         | xylem                  |
| RTE_copia_Eu_3 | Ale     | 24                   | root                  | xylem, cold-stressed   |
| RTE_copia_Eu_4 | Angela  | 243                  | seedlings, calli      | xylem                  |
| RTE_copia_Eu_5 | Ivana   | 54                   | leaves, root, calli, wood | xylem                  |
| RTE_copia_Eu_6 | GMR     | 63                   | leaves, seedlings     | xylem                  |

**Figure 1** Phylogenetic tree of Copia LTR-RTs elements. Based on Du and collaborators (2010).
11. McCarthy EM, McDonald JF: LTR STRUC: a novel search and identification program for LTR retrotransposons. Bioinformatics 2003, 19:362-367.
12. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 2011, (submitted).

doi:10.1186/1753-6561-5-S7-P164
Cite this article as: Marcon et al.: Identification of potential transcriptionally active Copia LTR retrotransposons in Eucalyptus. BMC Proceedings 2011 5(Suppl 7):P164.