

Article

Copaifera langsdorffii novel putative lncRNA: interspecies conservation analysis in adaptive response to different biomes

Monica F. Danilevicz¹, Kanhu C. Moharana², Thiago M. Venancio², Luciana O. Franco³, Sérgio R.S. Cardoso³, Mônica Cardoso³, Flávia Thiebaut¹, Adriana S. Hemerly¹, Francisco Prosdocimi⁴, Paulo C.G. Ferreira^{1,*}

¹ Laboratório de Biologia Molecular de Plantas, Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro -Brasil; monica.danilevicz@gmail.com (M.F.D.), flaviabqi@gmail.com (F.T.)

² Universidade Estadual do Norte Fluminense Darcy Ribeiro, RJ – Brasil ; thiago.venancio@gmail.com

³ Instituto de Pesquisa do Jardim Botânico, Rio de Janeiro, RJ – Brasil; lfranco@jbrj.gov.br

⁴ Laboratório de Genômica e Biodiversidade; Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro -Brasil

* Correspondence: paulof@bioqmed.ufrj.br ; Tel.: +5521988842585

Abstract: lncRNA are involved in multiple regulatory pathways, its versatile mode of action has disclosed a new layer in gene regulation. They are reportedly modulated during plant development, with specific tissue functions and in response to stresses. In this study, we analyzed lncRNA from leave samples collected from the legume *Copaifera langsdorffii* (copaiba) from two divergent ecosystems: Cerrado (CER) and Atlantic Rain Forest (ARF). We identified 8020 novel lncRNAs, from which 2893 transcripts were regulated above 2-fold and 566 above 5-folds in either condition. This putative lncRNA set was compared with seven *Fabaceae* genomes, of which 1747 and 1879 transcripts (from ARF and CER, respectively) aligned to at least two genomes. Further, 2194 copaiba lncRNAs were successfully mapped to at least one of six *Fabaceae* transcriptomes. The secondary structures of the lncRNAs that were conserved and differentially expressed between the populations were predicted using *in silico* methods. Our results indicate the potential involvement of lncRNAs in the adaptation of *C. langsdorffii* to two different biomes.

Keywords: novel lncRNA; lncRNA conservation; *Copaifera*; epigenetics; adaptive response;

1. Introduction

Copaifera is a genus of native trees from Latin America tropical regions and Western Africa. The species *C. langsdorffii*, *C. reticulata*, *C. cearensis*, *C. multijuga* among others are popularly known as “Copaiba” [69]. The *Copaifera* sp. oil is extracted through V-shaped cut in the stem bark and has been used by indigenous populations from the Brazilian Amazon as a powerful antimicrobial, anti-inflammatory and “overall healing purposes” [69]. Many of this alleged features have been investigated by pharmacological studies, confirming its anti-inflammatory capacity [7, 20, 70] and proving it to be an efficient alternative to treat dental infections [61], gastrointestinal disorders [50], endometriosis [59], skin ulcers [23, 41, 49], and to be applied to skin scaffold implants increasing tissue angiogenesis [43]. The Copaiba oil resin has been shown to exert larvicidal activity against *Aedes aegypti* [51, 58], and antibacterial activity *in vitro* [71]. The leaf extracts also presented leishmanicidal and antimalarial activities [36], also being effective as a biopesticide against lepidoptera [53].

The amount of Copaiba oil production is influenced by the climate and soil conditions, as more oil is produced at locations with clay soil, during the raining season [2]. In Brazil, *C. langsdorffii* is a

widely occurring species, included in a broad range of ecosystems like Cerrado, Atlantic Rain Forest and Caatinga, which are very distinct biomes [15], requiring diverse adaptive mechanisms. Such plasticity to adapt to different biomes is a complex regulation, involves several genetic, evolutionary and epigenetic fine tuning, which may also include long non-coding RNAs (lncRNAs).

lncRNAs are considered to be RNA transcripts, defined as longer than 200bp, with no apparent protein coding capacity [30, 31]. They hold many resemblances with mRNA, such as similar epigenetic marks to promote expression, binding sites for RNA polymerase II at their genome loci. They are often polyadenylated at 3' end and receive 5' CAPs, especially when acting outside the nucleus [30, 31]. Because they are (i) less conserved than mRNA at the sequence level across different species and (ii) regularly transcribed at low levels, lncRNAs were once considered transcriptional noise. lncRNAs tends to form secondary and tertiary structures, in which the molecule conformation is crucial to regulate their targets, leading to their functional domains and genomic positions to be more conserved than their sequence [31].

lncRNAs interspecies conservation may be influenced by the regulatory mechanism they play, whereas their sequence conservation depends on whether (a) the lncRNA molecule acts as a regulator or (b) the simply transcription of the lncRNA regulates its target gene [30]. Most functional lncRNAs undergo post-transcriptional processing and retain higher conservation of splice sites. This indicates that they most likely function in the mature form [24, 68]. In vertebrates, it is argued that transposable elements and bidirectional transcription may play an important role in the evolution and rapid turnover of lncRNAs [30].

There are several types of lncRNAs classified according to their genomic positions as: sense, natural antisense, bidirectional, intronic and intergenic; and they may also act through *cis* or *trans* regulation [5, 10, 55]. lncRNAs exhibit relatively low expression patterns compared to mRNAs, showing a specific profile depending on the (i) tissue or cell type observed [10], (ii) developmental stage and (iii) environmental stress response [9, 14].

A considerable amount of lncRNAs may act as chromatin regulators [68, 52]. For example, APOLO, a lncRNA responsive to auxin, interacts with the chromatin, leading to a loop formation encompassing the PID gene (key regulator of polar auxin transport) and regulating its expression [4]. HOTAIR, ANRIL and KCNQ1OT1 also are known to bind to more than one histone-modifying complex, acting as regulators [34]. COLDAIR and COLDWRAP are lncRNA described to act regulating the FLOWERING LOCUS C (FLC), associate to Polycomb protein complex to stably repress FLC during the vernalization process, prompting adaptive fitness and development in *Arabidopsis thaliana* [33]. In rice, it was shown that photoperiodic-sensitive male sterility (PSMS) is carried out by a lncRNA called LDMAR, in which a SNP led to change in secondary structure and subsequently repression of its expression at long day conditions [19]. ENOD40 and ASCO are examples of conserved lncRNA among legumes, which are involved in the re-localization of Nuclear Speckles RNA Binding Protein (NSRs) in *Medicago truncatula* and *A. thaliana*, albeit ASCO interaction with AtNSRs induces alternative splicing activity in auxin signaling [6, 13]. Several other lncRNA have their functions experimentally tested, such as HDI promoting photomorphogenesis in red light [28] and NERDL association to wood formation in *P. tormentosa* [56].

In response to inorganic phosphorus starvation, the lncRNA *cis*-NAT PHO1;2 acts regulating Pi homeostasis as a translational enhancer of OsPHO1;2 increasing its uptake when overexpressed in a mutated rice lineage [27]. Evidence of lncRNA involved in stress response and adaptation was also observed in nitrogen deficient *Populus* [16], maize under drought stress [86] and *A. thaliana* submitted to salt stress [62]. Moreover, lncRNAs have also been found to act as sRNA precursors [9, 16, 40, 74, 82] or acting as bait to sRNA regulation by target mimicry [21, 57, 9]; some lncRNA are target mimics to the plant miRNA response to viral infection [74]. Collectively, lncRNA have been implicated in various cell and molecular processes, including post-transcriptional regulation, post-translational regulation of protein activity and protein re-localization, organization of protein complexes, cell-cell signaling and intrinsically connected to adaptive fitness and overall plant processes [5, 8, 55].

In this study, we identified 8,020 novel lncRNAs through bioinformatics analysis of which approximately 565 were shown to be up regulated above 5 times in CER or ARF populations, which might be related to their adaptation to such diverse environments. lncRNAs tend to have their primary sequence evolving under a relaxed constraint, being less likely to present high interspecies sequence similarity. Nevertheless, we were able to identify one single transcript that matched a *Glycine max* lncRNA from CANTATA database [63]. The high sequence similarity between this transcript from soybean and Copaiba indicates that it is potentially functional. Considering also that it was regulated under the conditions tested, it may be involved in the adaptive response. In addition, we were able to map 156 lncRNAs to seven *Fabaceae* genomes, of which ten transcripts exhibited more than 5-fold up regulation among the conditions. The emerging epigenetic studies involving lncRNAs function and conservation have shown their involvement to several types of biotic and abiotic stresses. Thus, the conservation of lncRNAs among *Fabaceae* species validates their transcription. The differential expression observed for hundreds of lncRNAs suggests that they take part in regulatory pathways that lead to adaptive responses in Copaiba.

2. Results

2.1. Identification of novel and differentially expressed lncRNA

In order to identify lncRNAs and elucidate their adaptive roles, we sequenced the transcriptomes of *C. langsdorffii* leaves obtained from trees growing at two different ecosystems, the (i) Atlantic Rain Forest (ARF, humid condition) and the (ii) Cerrado (CER, dry forest) in Brazil during the drought season for both locations. A total of 138,175 and 199,556 transcripts were assembled from ARF and CER, respectively. On these transcripts, a series of filters were applied. Initially we removed sequences shorter than 200bp and those containing ORFs encoding peptides of at least 100 aa. Next, we ran CPC (v0.9-r2) [35] and PLEK (v1.2) [37] to assess the protein coding potential of these transcripts. Transcripts classified as non-coding by both software were compared to establish one-to-one correspondence between ARF and CER transcripts. This correspondence was defined using bi-directional BLASTN (v2.2.31+) [12] and MCL (implemented in scps, v0.9.8[45]). Reads were remapped against the assembled transcriptomes with Bowtie2 [64] and transcriptional levels estimated with cufflinks v2.2.1 [64]. Transcripts with at least 1 RPKM (Reads Per Kilobase of transcript per Million mapped reads) were kept for downstream analyses (8,020 transcripts). The majority of these (2,327 transcripts) presented over two fold regulation, while 566 transcripts were regulated above 5 fold (Figure 1, supplementary table 1).

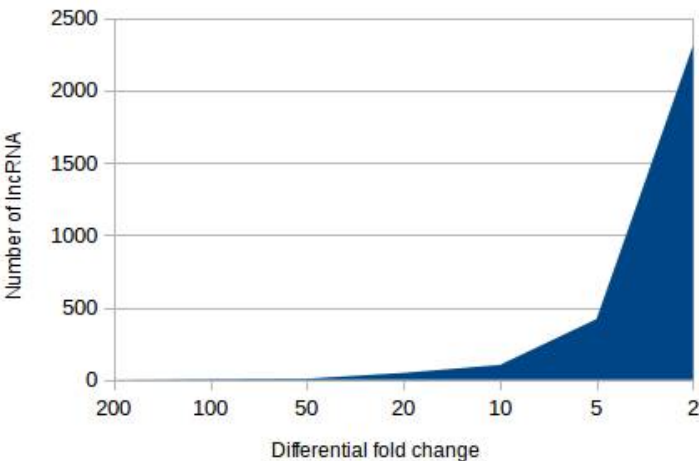


Figure 1: Putative lncRNA identified in ARF and CER samples and its fold change regulation in comparison to each other: There were 2893 differentially regulated transcripts identified from a total of 8020 copaiba lncRNAs. The majority of the transcripts are 2 to 5 times differently expressed on either sample, yet there are 581 transcripts regulated above 5 fold in either sample.

Table 1: Stem-loop secondary structure of highly regulated conserved copaiba lncRNA: The selected RNA have their MFE estimated by ViennaRNA package. The secondary structure is deemed to be stable when the MFE ≤ -80 kcal, thus from the 12 transcripts.

ARF ID	CER ID	ARF RPKM	CER RPKM	ARF/CER	CER/ARF	MFE (kcal/mol)
JCF44_0000048876	JCF45_0000090780	1269.38	4.83663	262.4513349	0.00381023	-209.9
JCF44_0000039348	JCF45_0000129049	854.553	11.2997	75.62616707	0.013222936	-220.5
JCF44_0000061237	JCF45_0000001210	149.628	3.75544	39.84300109	0.025098511	-189.2
JCF44_0000046349	JCF45_0000133784	286.172	9.83557	29.09561927	0.034369435	-193
JCF44_0000021616	JCF45_0000120089	80.4121	3.17854	25.29843891	0.039528131	-125.5
JCF44_0000041906	JCF45_0000057106	90.1082	3.75251	24.01278078	0.04164449	-52.5
JCF44_0000027393	JCF45_0000069991	55.4647	3.02796	18.3175141	0.054592561	-70.6
JCF44_0000021638	JCF45_0000000795	9.75664	108.186	0.090183942	11.08844848	-106.1
JCF44_0000094004	JCF45_0000091462	11.7198	131.109	0.089389744	11.18696565	-151.9
JCF44_0000075087	JCF45_0000057481	12.3338	138.6	0.088988456	11.23741264	-118.9
JCF44_0000042254	JCF45_0000109819	2.39024	28.4969	0.083877194	11.92219191	-162.3
JCF44_0000040403	JCF45_0000101583	80.5029	1132.28	0.07109805	14.06508337	-87.9

2.2. Interspecies lncRNA conservation analysis

2.2.1. Positional conservation and genome alignment analysis

Since there is no reference genome for *C. langsdorffii*, we performed a similarity search using Bowtie2 (v2.3.4.1) [64] against the genomes of the following related species: (i) *Vicia faba*, (ii) *Glycine max*, (iii) *Medicago truncatula*, (iv) *Phaseolus vulgaris*, (v) *Lotus japonica*, (vi) *Vigna unguiculata* and (vii) *Cicer reticulatum*. Genome sequences were downloaded from NCBI Genomes (<ftp://ftp.ncbi.nlm.nih.gov/genomes/>). From the alignments, we found 1,747 and 1,879 transcripts respectively from ARF and CER samples to match at least one of genomes used. Using DrawVenn (<http://bioinformatics.psb.ugent.be/webtools/Venn>) application we were able to identify 156 transcripts that aligned to all seven *Fabaceae* genomes in both ARF and CER samples (Figure 2).

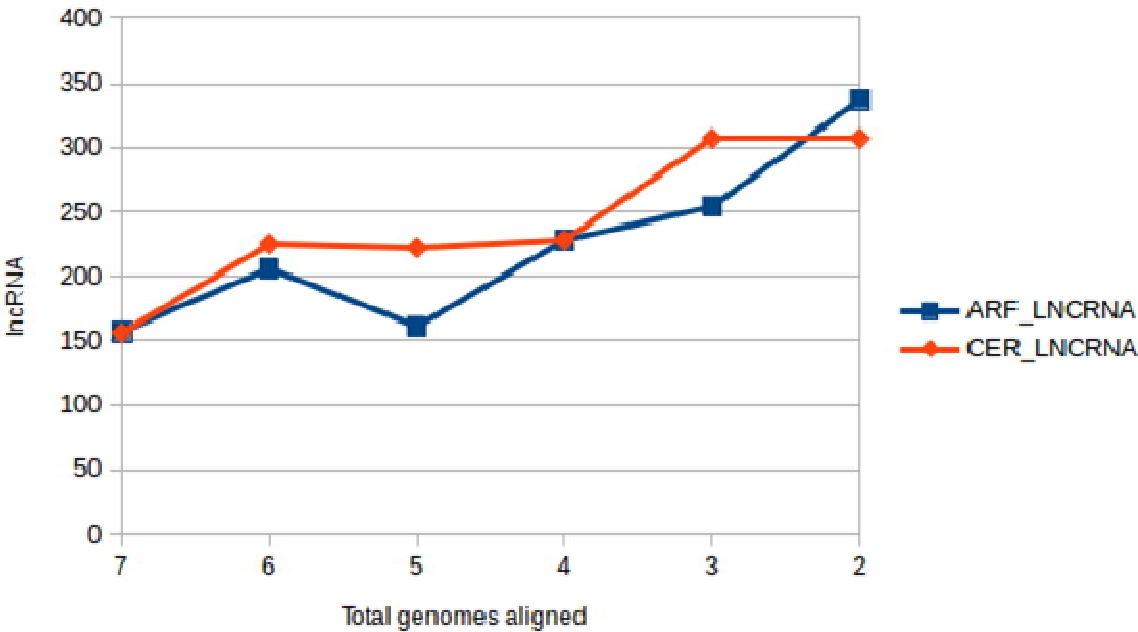


Figure 2: lncRNA interspecies conservation and genome alignment analysis: the graph shows the total number of genomes versus the total number of transcripts aligned to *Fabaceae* genomes (*V. faba*, *G. max*, *M. truncatula*, *P. vulgaris*, *L. japonica*, *V. unguiculata* and *C. reticulatum*). The blue line

represents the lncRNA gather from the ARF samples and the red line represents the lncRNA gather from the CER sample, their alignment profile is very similar as expected, around 1800 transcripts aligned to at least one of the genomes used.

From the subset of 156 copaiba lncRNAs aligned to all seven *Fabaceae* genomes, 45 transcripts presented above 2-fold differential regulation (Figure 3). To understand if the copaiba lncRNA transcripts aligned to all seven *Fabaceae* genomes hold positional conservation, we selected 10 transcripts exhibiting at least 5-fold up-regulation to retrieve the information regarding their genomic locations. For this analysis, we used the reference genomes of *P. vulgaris*, *G. max* and *M. truncatula*, as they are better assembled and annotated. However, the genes closely located to the lncRNA loci are described only as “hypothetical protein-coding”, or “plant-like protein”. Further, none of the 6,133 lncRNA annotated in the *G. max* genome were located close to copaiba putative lncRNA alignment locations. The *M. truncatula* or *P. vulgaris* annotation files had no information on lncRNAs.

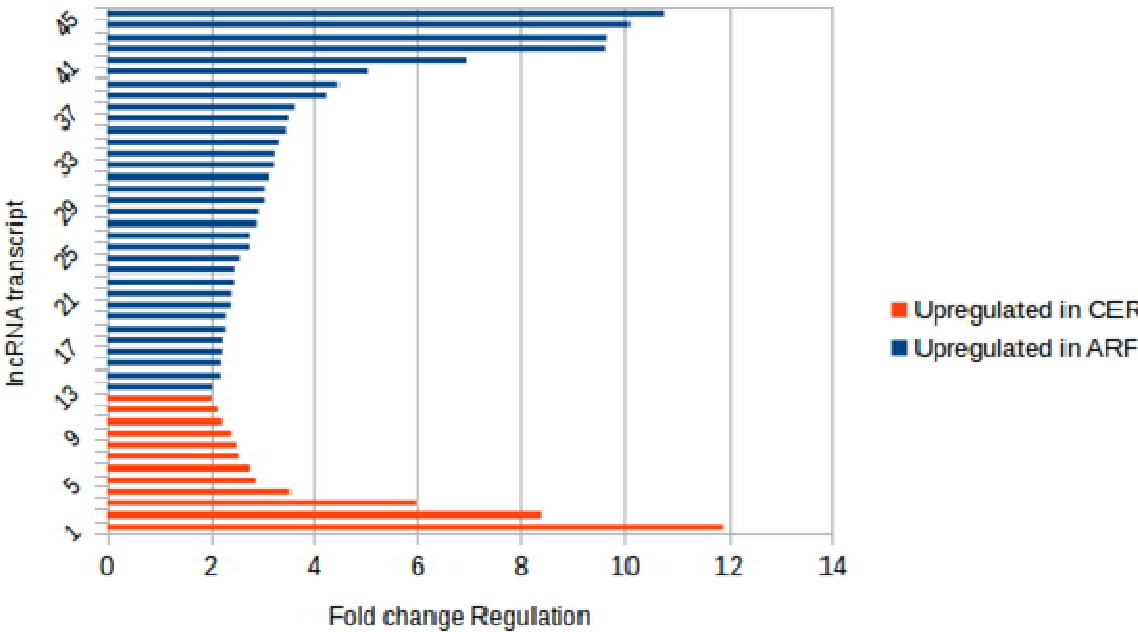


Figure 3: Subset of transcripts which aligned to all Fabaceae genomes presenting differential expression: From 156 transcripts which aligned to all genomes analyzed, there were 45 copaiba lncRNAs upregulated in either condition represented in this graph. Each transcript is represented by a single bar, in red are indicated the lncRNAs upregulated in CER samples in relation to ARF. In blue are indicated the lncRNAs upregulated in ARF samples in relation to CER. In the x axis is indicated the fold change regulation of the transcripts.

2.2.2. Identification of putative *C. langsdorffii* lncRNA in EST sequences of other Fabaceae species.

We downloaded EST and cDNA libraries from Phytozome and NCBI databases from *C. reticulatum*, *P. vulgaris*, *M. truncatula*, *G. max*, *V. unguiculata* and *L. japonicus*, to identify whether the copaiba lncRNA transcripts that aligned to the genomes could also be found in available transcriptome libraries, which comprise both mRNA and poly-A ncRNA. lncRNAs have a specific expression profile, considered to be lower than mRNAs. The likelihood of expression is highly influenced by tissue and specific condition [10]; therefore the identification of copaiba lncRNA in *Fabaceae* transcriptome is susceptible to be underestimated. This is due to the fact that transcripts specific profile of expression and relatively low amounts may not be observed even if present in the reference transcriptomes.

We used BLASTN (v2.2.31+)[12] to search putative copaiba lncRNAs against downloaded transcript sequences. Two identity thresholds, 50% and 90%, were used in this analysis, along with

e-value and coverage threshold of 10^{-15} and 50%, respectively. We found that 27% (2,194 lncRNA) aligned to at least one species' transcriptome (using 50% identity filter on BLASTN) (Figure 4). While only 3.3% (264 lncRNA) aligned with 90% identity to at least one *Fabaceae* expressed transcripts library and a single transcript (JCF44_0000056614) aligned to five species. It is remarkable to notice that 227 transcripts aligned to *G. max* with more than 90% identity, which might due to *G. max* to be more intensively studied species than the others (Figure 5).

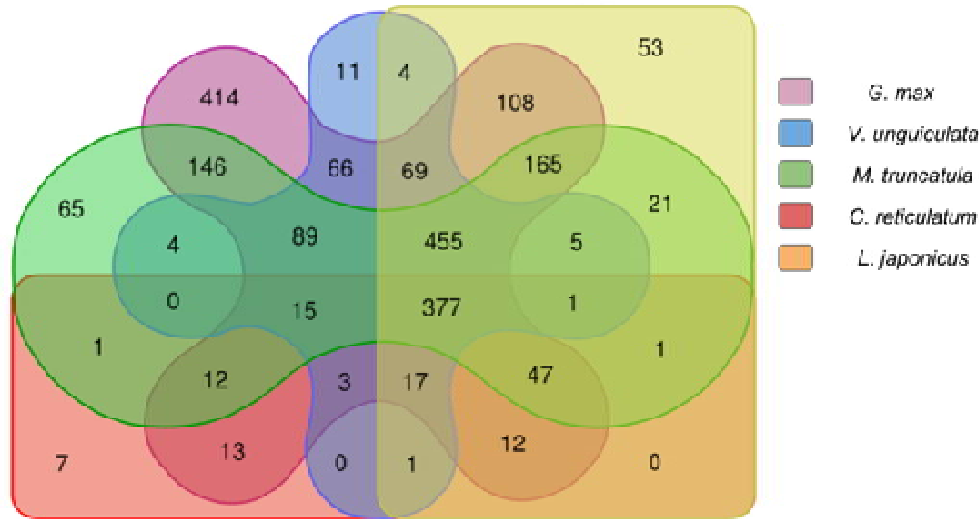


Figure 4: Count of copaiba lncRNA aligned to each Fabaceae transcriptome with Blastn (50% identity): In this diagram is shown the amount of copaiba lncRNA aligned to the transcriptome of each species, segregated by a color pattern indicated on the legend. In the colored overlapped area are the transcripts which were aligned to more than one species transcriptome, and its respective amount. There are only five species shown in this diagram, for illustration purposes we left out *P. vulgaris*.

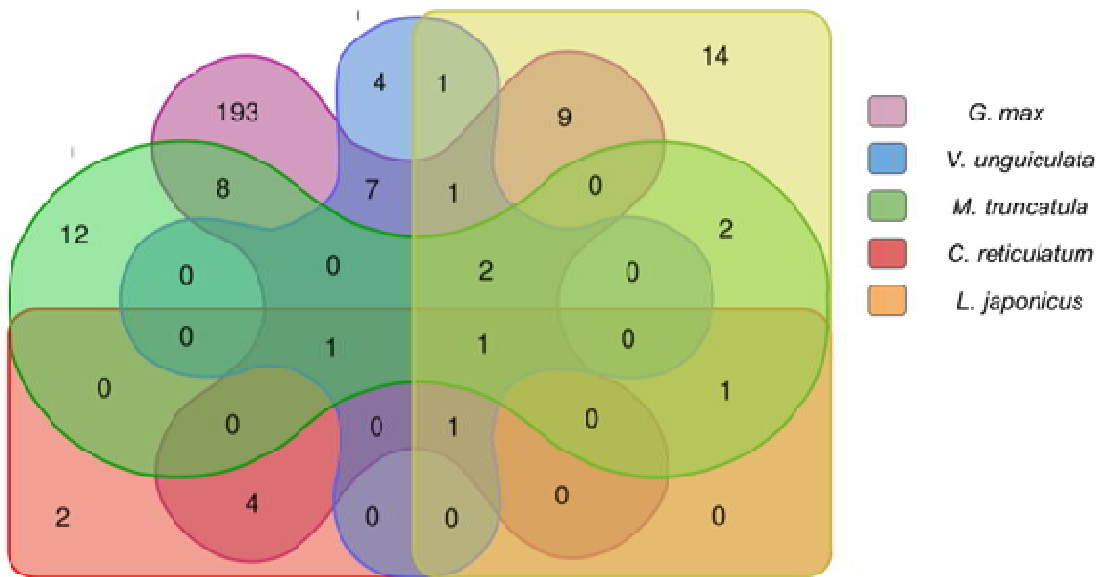


Figure 5: Count of copaiba lncRNA aligned to each Fabaceae transcriptome with Blastn (90% identity): In this diagram is shown the amount of copaiba lncRNA aligned to the transcriptome of each species, segregated by a color pattern indicated on the legend. In the colored overlapped area

are the transcripts which were aligned to more than one species transcriptome, and its respective amount. There are only five species shown in this diagram, for illustration purposes we left out *P. vulgaris*.

2.2.3. Comparison analysis of conserved putative lncRNA

To assess whether the copaiba lncRNAs aligned to *Fabaceae* transcribed libraries were also aligned to *Fabaceae* genomes, we used the Bowtie2 [64] aforementioned result, aiming to determine the amount of transcripts that are present in the genome and also are actively transcribed in the cDNA, EST libraries used (Figure 6, supplementary table 2). Thus, we identified 1,141 transcripts that aligned to both genome and transcriptome comparisons of at least one species. From this total, there were 36 copaiba lncRNA transcripts which aligned to both genome and transcriptome of six *Fabaceae* species (Figure 7).

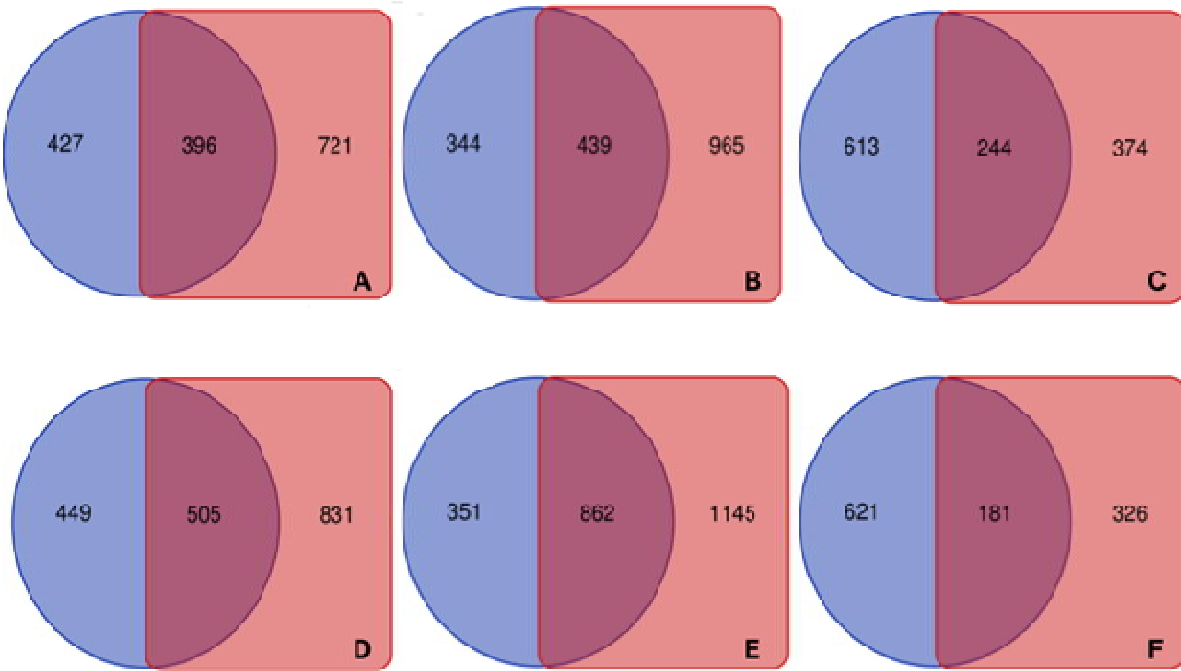


Figure 6: Blastn comparison of copaiba lncRNA aligned to each species genome and cDNA: In the venn diagram is compared the amount of copaiba transcripts aligned to the genome (blue circle), to the transcriptome (red square) our to both (overlapped area). **A** is the *V. unguiculata* comparison, **B** is *M. truncatula*, **C** is *P. vulgaris*, **D** is *L. japonicus*, **E** is *G. max* and **F** is *C. reticulatum*. It is possible to observe that *G. max*, *L. japonicus* and *M. truncatula* presented a higher number of overall aligned transcripts and also overlapped ones.

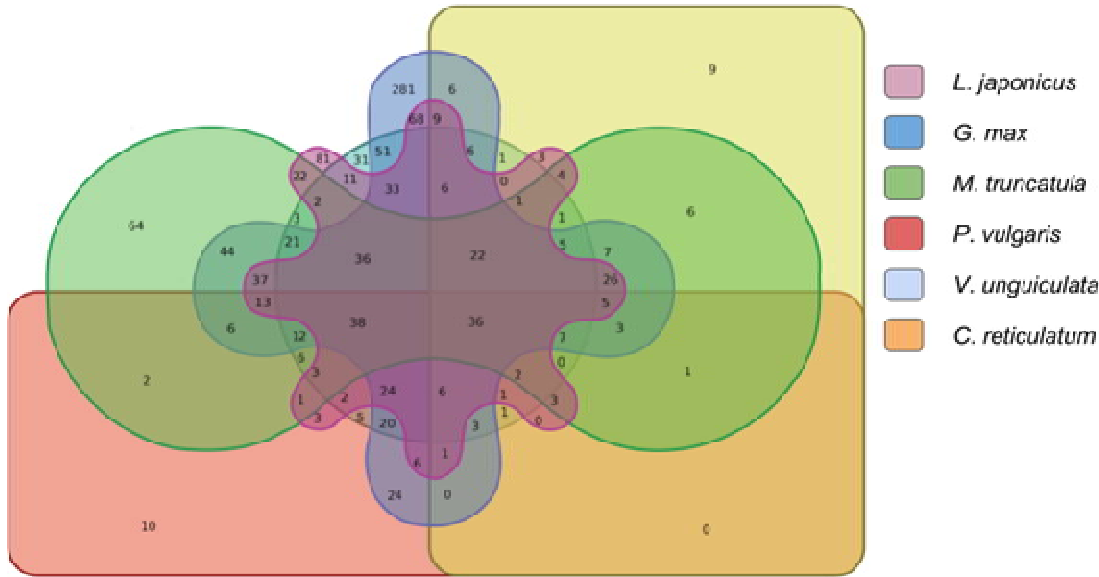


Figure 7: Total transcripts aligned to Fabaceae genomes and transcriptome: This diagram shows the amount of copaiba lncRNA aligned to the both genome and transcriptome of each species segregated by a color pattern indicated on the legend. In the colored overlapped area are the transcripts which were aligned to more than one species genome and transcriptome, and its respective amount.

2.2.4. Expression analysis of the conserved putative lncRNA among Fabaceae.

Most of the lncRNA transcripts that aligned to both genome and transcriptomes of *Fabaceae* species displayed similar RPKM values for ARF and CER. However, several lncRNA (254 out of 1,141) were regulated at either condition above 2-fold, indicating that they might be involved with the adaptation to different environments. (Figure 8). From the total transcripts which aligned to both genome and transcriptome of *Fabaceae* species, it is possible to identify the formation of two clear groups of putative lncRNA formed in each condition. The values and identification of the putative lncRNA are available at supplementary table 3.

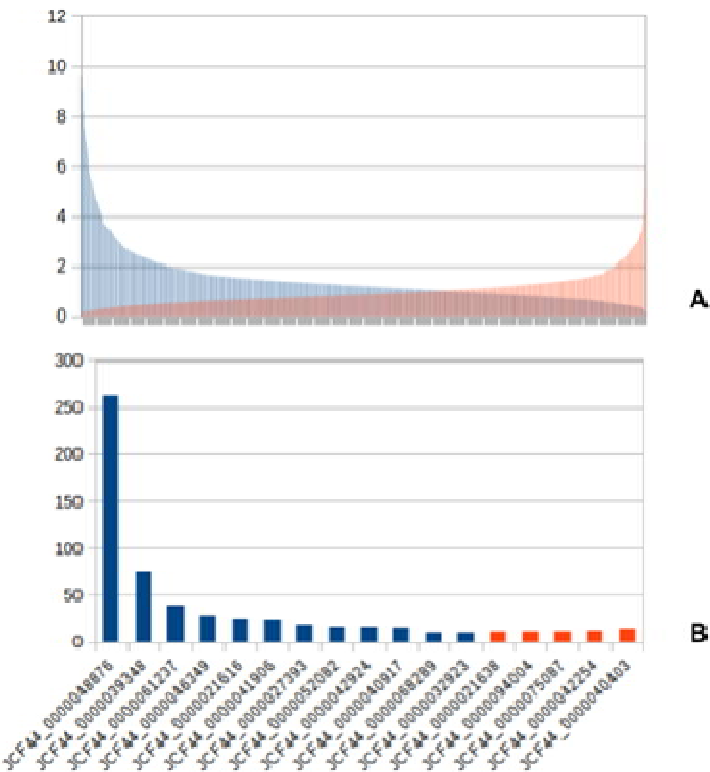


Figure 8: Fold change comparison of copaiba lncRNA conserved in Fabaceae species: From 1141 lncRNA which aligned to multiple *Fabaceae* genome and transcriptome, there was 254 lncRNA regulated above 2 fold in either ARF or CER samples. **A** displays the regulated transcripts between 2 and 10 fold, the blue bars are up regulated in ARF samples, and red bars are up regulated in CER samples. Each bar corresponds to a single transcript. **B** displays the regulated transcripts above 10 fold, the blue bars are up regulated in ARF samples, and red bars are up regulated in CER samples. Each bar corresponds to a single transcript.

2.2.5. Identification of Known lncRNA

In order to investigate conserved lncRNAs in closely related species, we used the putative transcripts obtained and compared with BLASTN against CANTATA and GREnc lncRNA databases for *S. bicolor*, *G. max*, *M. truncatula* and *P. vulgaris*, which summed up approximately 32,000 sequences. This comparison found a single transcript (JCF45_0000011974/ JCF44_0000015840) that matched a *G. max* lncRNA from CANTATA database (CNT2032069) with 90% identity and e-value of $4e^{-97}$. This single transcript is 6.5 fold up-regulated in ARF samples. This transcript was also conserved in *G. max*, *L. japonicus* and *P. vulgaris* genome analysis.

The high sequence identity of the JCF45_0000011974/JCF44_0000015840 transcript with other legume genomes and with a *G. max* lncRNA transcript indicates that it plays important biological roles, which could be related to the adaptation to different niches, as it is differentially regulated between ARF and CER populations.

2.3. Stem-loop secondary structure of regulated putative lncRNA.

Recent studies have suggested that the secondary and even tertiary structure of lncRNAs are conserved and critical for the transcript to be functionally active [42, 44]. In vertebrates it was observed that the conservation of secondary structure of the lncRNA MEG3 regulates a tumor suppression function, regardless of its primary sequence [85]. The secondary structure of lncRNA is regarded as one of the multidimensional conservation pressure that long no-coding transcripts can

suffer [32], and as a result of their sequence length, most RNA transcripts are prone to form secondary structure [66]. However ncRNA present some distinguishable features, they show higher thermostability than coding transcripts, (i) their temperature melting (T_m) is significantly higher and (ii) with greater negative free energy values (Minimum Free Energy - MFE) [32, 42]. Another study observed that functional transcripts tend to present higher *in silico* second structure stability (with greater negative MFE), suggesting a link between secondary structure stability and functionality [53, 73].

There are lncRNA in which the secondary structure dictates their functionality, in humans the lncRNA MEG3 acts as a tumor suppressor based rather in the structure than primary sequence conservation [42], also the steroid receptor RNA activator (SRA), one of the few lncRNA that has its secondary structure experimentally defined, is reported to interact with many proteins and be related to breast cancer development. Although SRA primary sequence is mutated, the secondary structure, and the nucleotides involved in the stabilization of the structure are highly conserved, suggesting their direct involvement in the lncRNA functionality [47]. Experimental analysis comparing the folding energy of lncRNAs and mRNAs were capable to differentiate lncRNA based on their higher MFE [32].

Therefore, we performed folding analysis using ViennaRNA (v2.4.8) [39]. In which the stability of secondary structure of RNA can be inferred by MFE, regarding any values below -80 kcal/mol to be structurally stable [47]. We selected ten differentially regulated lncRNAs (table 1), from the set of conserved lncRNA in both *Fabaceae* genome and transcriptome analysis, the majority of them presented MFE below -100kcal/mol. Once these structures were shown to be stable, it is expected that they may have a functional role in which a secondary structure formation is likely to be meaningful. At the supplementary figure 1, is possible to observe the expected second structure of the given lncRNAs analyzed.

3. Discussion

Forest trees are unique group to study adaptability traits, based on their life span and endurance to biotic and abiotic stress [79]. Thus, in the present study we identified 8,020 putative lncRNAs, some of which regulated above 2- and 5-fold in either copaiba population originated from different biomes. Also a thorough analysis of copaiba coding transcripts is being conducted by our group (Franco et al., in preparation), leading to a deeper understanding about their adaptability to these different environments at mRNA and epigenetics level, which is critical to copaiba management and conservation [1]. Similarly, Xu et al. [83] performed a study with *Miscanthus lutarioriparius* populations from two different environments, and observed an expression profile on which the lncRNA presented higher fold change expression than usually observed with mRNA, suggesting them to be more sensitive and responsive to environmental changes than coding transcripts.

Positional conservation in the genome analysis of lncRNA loci neighboring specific orthologous genes plays an interesting role, particularly when their primary and secondary structures are not completely conserved through species. For example, in humans, a small segment of AIRN lncRNA overlaps with the IGF2R promoter (insulin-like growth factor 2 receptor) and is sufficient to cause IGF2R silencing. This phenomenon was also identified in other lncRNAs [25, 44]. Through previous interspecies comparison [66], it was observed that the positional conservation may act by regulating the target gene solely by the transcription of the lncRNA.

In the present copaiba analyses, a total of 22% of the lncRNAs aligned to at least one of the genomes analyzed. The observed amount lncRNA conservation is coherent with the literature, since the conservation of lncRNA among the same family is much smaller than that of protein-coding genes [29, 65]. A conservation analysis between human and other placental mammals estimates only 44% of the lncRNA is conserved, while also their promoter seem to be under conservative pressure [17]. In plants, between *Z. mays* and *S. bicolor* there are approximately 25% lncRNA conserved, and *Z. mays* compared to *A. thaliana* presented only 2% conservation [37]. Thus *C. langsdorffii*'s 22% conservation along the *Fabaceae* family lies in the expected conservation threshold.

Within these conserved putative lncRNAs, some were regulated in either condition and is possible to observe that the majority of them are upregulated in ARF plants, with a high fold change. The high conservation and regulation among these few transcripts suggests that they are under evolutionary constraints, possibly involved in the plant regulatory machinery [44].

Often the lncRNA primary sequence can be degenerated while the position is maintained and the act of transcription itself is enough to function as epigenetic regulators to closely related genes [29, 42, 44, 65, 66]. A study performed comparing putative lncRNA among the *Brassicaceae*, *Aethionemeae* and *Cleomaceae* families uncovered several transcripts that were thought to be lineage-specific, instead they were in fact positionally conserved, although sequence divergent [44].

While comparing the copaiba lncRNA to several *Fabaceae* genomes has many advantages, such as to understand their position, classify the lncRNA as sense, natural antisense, bidirectional, intronic and intergenic, and possibly infer its function according to gene proximity [44]. Mapping the lncRNA transcripts to other species cDNA and EST sequences is a way to confirm that it is actively transcribed. Some previous studies that performed this analysis using vertebrates' lncRNA noticed that some transcripts may hold sequence similarities to genomic untranscribed regions [25]. In our study, a higher percentage of lncRNA was paired with EST and cDNA libraries than to the genomic alignment analysis; it might be due to lncRNA post transcriptional processing which prevents the lncRNA to be correctly mapped to the genome. The transcripts alignment analysis suggests more orthologous lncRNAs are being actively transcribed throughout the *Fabaceae* family than was estimated by genomic analysis.

Regarding that lncRNAs frequently present low expression profile and tend to be expressed in a tissue or condition specific manner [29, 47, 65], the transcripts that didn't aligned to any of the expressed *Fabaceae* libraries analyzed may still be conserved, although not expressed in the particular condition in which the library samples were taken from. It can be illustrated by the fact that, when comparing copaiba lncRNAs to two databases of putative plants lncRNAs, there was a single transcript that aligned to a known *G. max* lncRNA, and mapped to three *Fabaceae* genomes but didn't aligned to *G. max* EST and cDNA libraries. Hence, it is essential to understand if the comparison analysis is insufficient to state whether the transcript is expressed or not in a given species.

Moreover, we observed that 14.2% of copaiba lncRNA overlapped the genomic and transcription comparative analysis, while the majority presented similar expression levels in both ARF and CER samples, there were 254 transcripts regulated in either sample above 2-fold. In figure 8, we can detect a tendency to form two distinct expression profiles among the samples, in which 17 lncRNA were up regulated above 10-fold. Considering the samples are from the same type of tissue, and similar developmental stage, it is reasonable to infer that if the lncRNA identified is functional, the RPKM differential expression between the samples is associated to the plants response to environmental stimuli. It should be mentioned also that, although there are no replicates in the experiment, RNA samples were prepared from a pool of 10 plants from each local.

The *in silico* folding prediction of 12 most regulated lncRNAs which were conserved at genome and transcriptome level, most of the transcripts were regarded to have a stable secondary structure. Thus, a further investigation of these transcripts regarding their functionality is needed, since their structural stability and post-transcriptional processing indicates they are likely to regulate their target expression as molecule, not solely by being transcribed. Thus, a model of their putative second structure and possible targets identification are the natural research directions to further understand the role of lncRNAs in the adaptive response of copaiba and possibly other closely related woody plants.

5. Materials and Methods

5.1. Plant Material Collection

The *Copaifera langsdorffii* leaf samples were collected from two different biomes, ten individual samples were taken from Atlantic Rainforest, at Área de Proteção Ambiental da Bacia do Rio São

João – Mico Leão, Silva Jardim, RJ, Brazil. And another ten individual samples were collected from Cerrado ecosystem at Estação Ecológica do Jardim Botânico de Brasília – EEJBB, Distrito Federal, Brazil, in August, during the conspicuous annual dry season. All biological material harvested was placed in RNA later-like buffer, kept at -80°C until extraction.

5.2. RNA extraction to Sequencing

RNA extraction was performed individually from samples collected from Atlantic Rainforest and Cerrado. After, they were pooled together for the library construction, and named JCF-44 and JCF-45, respectively. It was then sent to Fasteris Life Sciences SA (Plan-les-Ouates, Switzerland) for the multiplex sequencing performed on Illumina HiSeq 2000I using the single-end 76 cycle protocol.

5.3. De novo transcriptome assembly and gene expression profiling

Sequencing reads were checked using FastQC and searched against NCBI nr database using BLASTN (megablast, e-value 1e-5; alignment length ≥ 90 and identity $\geq 80\%$)[3]. Reads aligning to non-plant organisms were removed. Raw reads were pre-processed using Trimmomatic (v0.36) [11] for trimming adapters, trailing 15 bases and tailing 5 bases for each read and other quality cleaning. *De novo* transcriptome assembly was performed using Trinity (v2.3.2; default settings)[72]. Fasta headers were renamed according to the sample ids (e.g. JCF44_xxx and JCF45_xxx). One-to-one correspondence between the transcriptomes was detected by bi-directional BLASTN (v2.2.31+)[3]. Stringent criteria were used for this analysis and only top hits with e-value 1e-10, query coverage of $\geq 50\%$, identity $\geq 90\%$, bit-score ≥ 50 were considered. Filtered reads were aligned to one-to-one transcripts using Bowtie2 (v. 2.2.9)[64] and RPKM values estimated using Cufflinks (v2.2.1)[64].

5.4. lncRNA identification

The one-to-one transcripts were screened for lncRNA on several parameters. Transcripts longer than 200bp with maximum ORF size of less than 100 amino acids by Getorf (EMBOSS:6.6.0.0) were used to predict putative lncRNA. These transcripts were analyzed using two software CPC (v0.9-r2) [35] and PLEK (v1.2) [37]. Only transcripts classified as non-coding in both CPC and PLEK analysis were considered as putative lncRNA. Further we kept only those putative lncRNAs with RPKM ≥ 1 .

5.5. Interspecies lncRNA conservation analysis

To identify other lncRNA which holds sequence conservation among *Fabaceae* species, we used BlastN 2.7.1+ (evalue 10^{-15} , identity 90% -qcov 50%) [12] against the downloaded GRENC (<http://grenc.sciencedesigners.com/>) and CANTATA database (<http://cantata.amu.edu.pl/>), both databases present putative lncRNA obtained through their own bioinformatics pipeline, which is similar to the one we used to filter our own. Using DrawVenn application (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) we identified the ones mapped to more than one genome. The EST and cDNA from *Fabaceae* species was obtained from NCBI (<ftp://ftp.ncbi.nlm.nih.gov/genomes/>) and Phytozome databases (<https://phytozome.jgi.doe.gov/>), we identified similar transcripts using BlastN 2.7.1+ (evalue 10^{-15} , identity 50% and 90%, query cov 50%) [12], followed by DrawVenn analysis (<http://bioinformatics.psb.ugent.be/webtools/Venn/>).

5.6. Second structure modeling

The second structure modeling of lncRNA was made using Vienna RNAfold (v2.4.8) (Lorenz, 2011) [39] package, at 25° C, default parameters.

Supplementary Materials: Table 1: Presents the novel putative copaiba lncRNA identified to be regulated above 2-fold in the comparison of CER and ARF samples. In this table are the identification of the transcripts in each sample, separated by columns and its respective RPKM expression. **Table 2:** Presents the novel copaiba lncRNA which were expressed in both genome and transcriptome of other *Fabaceae* species at the conservation analysis. The table presents their identification in each sample, and the respective RPKM

expression. **Table 3:** Presents novel copaiba lncRNA that were expressed in both genome and transcriptome of other *Fabaceae* species at the conservation analysis, and were regulated above 2-fold between CER and ARF samples, indicating they are potentially involved in the plant adaptive response to the different biomes. This table presents the identification in each sample, and the respective RPKM expression. **Figure 1:** Presents the secondary structures predicted by Vienna RNAfold for the 12 most highly regulated copaiba lncRNA conserved in both genome and transcriptome analysis. The image displays the MFE secondary structure with the base-pair probability in color code (blue indicates zero and red indicates maximum probability).

Author Contributions: Conceptualization, PF, MC, AH, SC., F.T. and LF.; Methodology, LF, SC, MD, TV and KM; Software, TV, KM, FP.; Validation, MD, LF, and SC Formal Analysis, MD.; Investigation, MD.; Resources, SC, LF.; Data Curation, MD.; Writing-Original Draft Preparation, MD.; Writing-Review & Editing, MD., FP, KM, TV; Visualization, MD.; Supervision, PF.; Project Administration, PF.; Funding Acquisition, PF

Funding: This research was funded by FAPERJ grant number E12/200.134/2014

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Aitken SN., Yeaman S., Holliday JA., Wang., T Curtis-McLane S. Adaptation, migration or extirpation: Climate change outcomes for tree populations. *Evolutionary Applications* **2008** 1 : 95 – 111.
2. Alencar, J. D. C. Estudos Silviculturais De Uma População Natural De Copaifera Multijuga Hayne-Leguminosae, Na Amazônia Central. 2 - Produção De Óleo-Resina. *Acta Amazonica* **1982**, 12 (1), 75–89.
3. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* , 25, 3389–3402.
4. Ariel, F.; Jegu, T.; Latrasse, D.; Romero-Barrios, N.; Christ, A.; Benhamed, M.; Crespi, M. Noncoding Transcription by Alternative RNA Polymerases Dynamically Regulates an Auxin-Driven Chromatin Loop. *Molecular Cell* **2014**, 55 (3), 383–396.
5. Ariel, F.; Romero-Barrios, N.; Jégu, T.; Benhamed, M.; Crespi, M. Battles and Hijacks: Noncoding Transcription in Plants. *Trends in Plant Science* **2015**, 20 (6), 362–371.
6. Bardou F., et al., Long noncoding RNA modulates alternative splicing regulators in Arabidopsis, *Dev. Cell* 30 (2) **2014** 166–176.
7. Basile, A.; Sertié, J.; Freitas, P.; Zanini, A. Anti-Inflammatory Activity of Oleoresin from Brazilian Copaifera. *Journal of Ethnopharmacology* **1988**, 22 (1), 101–109.
8. Bazin, J.; Bailey-Serres, J. Emerging Roles of Long Non-Coding RNA in Root Developmental Plasticity and Regulation of Phosphate Homeostasis. *Frontiers in Plant Science* **2015**, 6.
9. Ben Amor, B., Wirth, S., Merchan, F., Laporte, P., d'Aubenton-Carafa, Y., Hirsch, J., et al. Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. *Genome Res.* **2009** 19, 57–69.
10. Bhatia, G.; Goyal, N.; Sharma, S.; Upadhyay, S.; Singh, K. Present Scenario of Long Non-Coding RNAs in Plants. *Non-Coding RNA* **2017**, 3 (2), 16.
11. Bolger, A. M., Lohse, M., & Usadel, B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* **2014**.
12. Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., & Madden T.L. "BLAST+: architecture and applications." *BMC Bioinformatics* **2008** 10:421.
13. Campalans A., KondorosiA., Crespi M., Enod40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*, *Plant Cell* **2004** 16 1047–105
14. Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., et al. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* **2012** 491, 454–457

15. Carvalho, M. C.; Gomide, L. R.; Santos, R. M. D.; Scolforo, J. R. S.; Carvalho, L. M. T. D.; Mello, J. M. D. Modeling Ecological Niche Of Tree Species In Brazilian Tropical Area. *Cerne* **2017**, 23 (2), 229–240.
16. Chen, M.; Wang, C.; Bao, H.; Chen, H.; Wang, Y. Genome-Wide Identification and Characterization of Novel LncRNAs in Populus under Nitrogen Deficiency. *Molecular Genetics and Genomics* **2016**, 291 (4), 1663–1680.
17. Derrien T et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* **2012**, 22:1775–1789
18. Diederichs, S. The Four Dimensions of Noncoding RNA Conservation. *Trends in Genetics* **2014**, 30 (4), 121–123.
19. Ding, J.; Lu, Q.; Ouyang, Y.; Mao, H.; Zhang, P.; Yao, J.; Xu, C.; Li, X.; Xiao, J.; Zhang, Q. A Long Noncoding RNA Regulates Photoperiod-Sensitive Male Sterility, an Essential Component of Hybrid Rice. *Proceedings of the National Academy of Sciences* **2012**, 109 (7), 2654–2659.
20. Falcão, H. D. S.; Lima, I. O.; Santos, V. L. D.; Dantas, H. D. F.; Diniz, M. D. F.; Barbosa-Filho, J. M.; Batista, L. M. Review of the Plants with Anti-Inflammatory Activity Studied in Brazil. *Revista Brasileira de Farmacognosia* **2005**, 15 (4), 381–391.
21. Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., et al.. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* **2007** 39, 1033–1037
22. Gai, Y.-P.; Yuan, S.-S.; Zhao, Y.-N.; Zhao, H.-N.; Zhang, H.-L.; Ji, X.-L. A Novel LncRNA, MuLnc1, Associated With Environmental Stress in Mulberry (*Morus Multicaulis*). *Frontiers in Plant Science* **2018**, 9.
23. Gushiken, L. F. S.; Hussni, C. A.; Bastos, J. K.; Rozza, A. L.; Beserra, F. P.; Vieira, A. J.; Padovani, C. R.; Lemos, M.; Junior, M. P.; Silva, J. J. M. D.; Nóbrega, R. H.; Martinez, E. R. M.; Pellizzon, C. H. Skin Wound Healing Potential and Mechanisms of the Hydroalcoholic Extract of Leaves and Oleoresin of *Copaifera Langsdorffii* Desf. Kuntze in Rats. *Evidence-Based Complementary and Alternative Medicine* **2017**, 1–16.
24. Guttman, M.; Donaghey, J.; Carey, B. W.; Garber, M.; Grenier, J. K.; Munson, G.; Young, G.; Lucas, A. B.; Ach, R.; Bruhn, L.; Yang, X.; Amit, I.; Meissner, A.; Regev, A.; Rinn, J. L.; Root, D. E.; Lander, E. S. LincRNAs Act in the Circuitry Controlling Pluripotency and Differentiation. *Nature* **2011**, 477 (7364), 295–300.
25. Hezroni, H.; Koppstein, D.; Schwartz, M. G.; Avrutin, A.; Bartel, D. P.; Ulitsky, I. Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Reports* **2015**, 11 (7), 1110–1122.
26. Hezroni, H.; Perry, R. B.-T.; Meir, Z.; Housman, G.; Lubelsky, Y.; Ulitsky, I. A Subset of Conserved Mammalian Long Non-Coding RNAs Are Fossils of Ancestral Protein-Coding Genes. *Genome Biology* **2017**, 18 (1).
27. Jabnounge, M., Secco, D., Lecampion, C., Robaglia, C., Shu, Q., and Poirier, Y. A rice cis-natural antisense RNA acts as a translational enhancer for its cognate mRNA and contributes to phosphate homeostasis and plant fitness. *Plant Cell* **2013** 25, 4166–4182
28. Jiao Y., O.S. Lau, X.W. Deng, Light-regulated transcriptional networks in higher plants, *Nat. Rev. Genet.* **2007** 8 (3) 217–230.
29. Johnsson, P.; Lipovich, L.; Grandér, D.; Morris, K. V. Evolutionary Conservation of Long Non-Coding RNAs; Sequence, Structure, Function. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2014**, 1840 (3), 1063–1071.
30. Kapusta, A., and Feschotte, C. Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. *Trends Genet.* **2014** 30:439–452
31. Kashi, K.; Henderson, L.; Bonetti, A.; Carninci, P. Discovery and Functional Analysis of LncRNAs: Methodologies to Investigate an Uncharacterized Transcriptome. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **2016**, 1859 (1), 3–15.
32. Kertesz, M. et al. Genome-wide measurement of RNA secondary structure in yeast. *Nature* **2010** 467, 103–107.
33. Kim, D.-H.; Sung, S. Vernalization-Triggered Intragenic Chromatin Loop Formation by Long Noncoding RNAs. *Developmental Cell* **2017**, 40 (3).

34. Kitagawa, M.; Kotake, Y.; Ohhata, T. Long Non-Coding RNAs Involved in Cancer Development and Cell Fate Determination. *Current Drug Targets* **2012**, *13* (13), 1616–1621.
35. Kong, L.; Zhang, Y.; Ye, Z.-Q.; Liu, X.-Q.; Zhao, S.-Q.; Wei, L.; Gao, G. CPC: Assess the Protein-Coding Potential of Transcripts Using Sequence Features and Support Vector Machine. *Nucleic Acids Research* **2007**, *35* (suppl_2).
36. Leandro, L. M.; Vargas, F. D. S.; Barbosa, P. C. S.; Neves, J. K. O.; Silva, J. A. D.; Veiga-Junior, V. F. D. Chemistry and Biological Activities of Terpenoids from Copaiba (*Copaifera* Spp.) Oleoresins. *Molecules* **2012**, *17* (4), 3866–3889.
37. Li L, Eichten SR, Shimizu R, Petsch K, Yeh C-T, Wu W, et al. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol.* **2014**;15:R40.
38. Liu X,Hao L, Li D, Zhu L, Hu S,Long Non-coding RNAs and Their Biological Roles in Plants, *Genomics, Proteomics & Bioinformatics*, **2015** 137-147, ISSN 1672-0229
39. Lorenz, R. and Bernhart, S.H. and Höner zu Siederdissen, C. and Tafer, H. and Flamm, C. and Stadler, P.F. and Hofacker, I.L. "ViennaRNA Package 2.0", *Algorithms for Molecular Biology* **2011**, 6:1 page(s): 26,.
40. Ma X, Shao C, Jin Y, Wang H, Meng Y. Long non-coding RNAs: a novel endogenous source for the generation of Dicer-like 1-dependent small RNAs in *Arabidopsis thaliana*. *RNA Biol* **2014**;11:373–90.
41. Masson, D. D. S. Atividades Cicatrizante e Antimicrobiana Do Óleo-Resina De Copaiba (*Copaifera Langsdorffii*) Em Úlceras Cutâneas. *Rev. Bras. Pl. Med.*, **2013**, v.15, n.4, supl.I, p.664-66.
42. Mercer, T. R.; Mattick, J. S. Structure and Function of Long Noncoding RNAs in Epigenetic Regulation. *Nature Structural & Molecular Biology* **2013**, *20* (3), 300–307.
43. Millás ALGM., Siveira JVW., Barbosa R., Puzzi M., Koh IHJ., Bittencort E., Implantes de scaffolds de plga incorporados com óleo-resina do gênero copaifera. 14º Congresso da Sociedade Latino Americana de Biomateriais, Órgãos Artificiais e Engenharia de Tecidos **2017** - SLABO.
44. Mohammadin et al. *BMC Plant Biology* **2015** 15:217
45. Nepusz, T., Sasidharan, R. & Paccanaro, A. SCPS: a fast implementation of a spectral method for detecting protein families on a genome-wide scale. *BMC Bioinformatics* **11**, 120 (2010).
46. Nitsche A, Stadler PF. Evolutionary clues in lncRNAs. *Wiley Interdiscip Rev RNA* **2017** 8(1). doi:10.1002/wrna.1376
47. Novikova IV, Hennelly SP, Sanbonmatsu KY, Rna K. Do lncRNAs have secondary and tertiary structure? Sizing up long non-coding RNAs. *BioArchitecture.* **2012**;2:189–99.
48. Novikova, I.V., Hennelly, S.P., and Sanbonmatsu, K.Y.. Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res* **2012**. *40*, 5034–5051
49. Paiva, L.; Rao, V.; Gramosa, N.; Silveira, E. Gastroprotective Effect of *Copaifera Langsdorffii* Oleo-Resin on Experimental Gastric Ulcer Models in Rats. *Journal of Ethnopharmacology* **1998**, *62* (1), 73–78.
50. Paiva, L.; Gurgel, L.; Silva, R.; Tomé, A.; Gramosa, N.; Silveira, E.; Santos, F.; Rao, V. Anti-Inflammatory Effect of Kaurenoic Acid, a Diterpene from *Copaifera Langsdorffii* on Acetic Acid-Induced Colitis in Rats. *Vascular Pharmacology* **2002**, *39* (6), 303–307.
51. Prophiro, J. S.; Silva, M. A. N. D.; Kanis, L. A.; Silva, B. M. D.; Duque-Luna, J. E.; Silva, O. S. D. Evaluation of Time Toxicity, Residual Effect, and Growth-Inhibiting Property of *Carapa Guianensis* and *Copaifera* Sp. in *Aedes Aegypti*. *Parasitology Research* **2011**, *110* (2), 713–719.
52. Rinn, J. L., and Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **2012** *81*, 145–166. biochem 051410–092902
53. Sâmia, R. R.; Oliveira, R. L. D.; Moscardini, V. F.; Carvalho, G. A. Effects of Aqueous Extracts of *Copaifera Langsdorffii* (*Fabaceae*) on the Growth and Reproduction of *Spodoptera Frugiperda* (J. E. Smith) (Lepidoptera: Noctuidae). *Neotropical Entomology* **2016**, *45* (5), 580–587.
54. Steffens, S.; Thiel, H.-J.; Behrens, S.-E. The RNA-Dependent RNA Polymerases of Different Members of the Family Flaviviridae Exhibit Similar Properties in Vitro. *Journal of General Virology* **1999**, *80* (10), 2583–2590
55. Shafiq, S.; Li, J.; Sun, Q. Functions of Plants Long Non-Coding RNAs. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **2016**, *1859* (1), 155–162.
56. Shi, W.; Quan, M.; Du, Q.; Zhang, D. The Interactions between the Long Non-Coding RNA NERDL and Its Target Gene Affect Wood Formation in *Populus Tomentosa*. *Frontiers in Plant Science* **2017**, *8*.

57. Shin, H.; Shin, H.-S.; Chen, R.; and Harrison, M. J.. Loss of At4 function impacts phosphate distribution between the roots and the shoots during phosphate starvation. *Plant J* **2006**, *45*, 712–726.
58. Silva, H. H. G. D.; Geris, R.; Filho, E. R.; Rocha, C.; Silva, I. G. D. Larvicidal Activity of Oil-Resin Fractions from the Brazilian Medicinal Plant *Copaifera Reticulata* Ducke (Leguminosae-Caesalpinoideae) against *Aedes Aegypti* (Diptera, Culicidae). *Revista da Sociedade Brasileira de Medicina Tropical* **2007**, *40* (3), 264–267.
59. Silva, J. H. D.; Borges, V. R. D. A.; Pereira, L. D. C. B.; Ferrari, R.; Mattos, R. M. D.; Barros, E. G. D. O.; Palmero, C. Y.; Fernandes, P. D.; Carvalho, P. R. D.; Sousa, V. P. D.; Cabral, L. M.; Nasciutti, L. E. The Oil-Resin of the Tropical Rainforest Tree *Copaifera Langsdorffii* reduces Cell Viability, Changes Cell Morphology and Induces Cell Death in Human Endometriotic Stromal Cultures. *Journal of Pharmacy and Pharmacology* **2015**, *67* (12), 1744–1755.
60. Somarowthu et al., HOTAIR Forms an Intricate and Modular Secondary Structure, *Molecular Cell* **2015**, /10.1016/
61. Souza, A. B.; Martins, C. H. G.; Souza, M. G. M.; Furtado, N. A. J. C.; Heleno, V. C. G.; Sousa, J. P. B. D.; Rocha, E. M. P.; Bastos, J. K.; Cunha, W. R.; Veneziani, R. C. S.; Ambrósio, S. R. Antimicrobial Activity of Terpenoids from *Copaifera Langsdorffii* Desf. against Cariogenic Bacteria. *Phytotherapy Research* **2010**.
62. Sunkar, R. Novel and Stress-Regulated MicroRNAs and Other Small RNAs from Arabidopsis. *The Plant Cell Online* **2004**, *16* (8), 2001–2019.
63. Szczesniak MW, Rosikiewicz W, Makalowska I. CANTATADB: a Collection of Plant Long Non-coding RNAs. *Plant Cell Physiol.* 2015;
64. Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D. R.; Pimentel, H.; Salzberg, S. L.; Rinn, J. L.; Pachter, L. Differential Gene and Transcript Expression Analysis of RNA-Seq Experiments with TopHat and Cufflinks. *Nature Protocols* **2012**, *7* (3), 562–578.
65. Ulitsky I, Bartel DP. LincRNAs: genomics, evolution, and mechanisms. *Cell* **2013**, *154*:26–46
66. Ulitsky, I. Evolution to the Rescue: Using Comparative Genomics to Understand Long Non-Coding RNAs. *Nature Reviews Genetics* **2016**, *17* (10), 601–614.
67. Van Bel, M.; Diels, T.; Vancaester, E.; Kreft, L.; Botzki, A.; Van de Peer, Y.; Coppens, F.; Vandepoele, K. PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics *Nucleic Acids Res* **2017**.
68. Vance, K. W.; Ponting, C. P. Transcriptional Regulatory Functions of Nuclear Long Noncoding RNAs. *Trends in Genetics* **2014**, *30* (8), 348–355.
69. Veiga-Junior VF, Pinto AC 2002. The *Copaifera* L. genus. *Quim Nova* 25: 273-286.
70. Veiga, V.; Rosas, E.; Carvalho, M.; Henriques, M.; Pinto, A. C. Chemical Composition and Anti-Inflammatory Activity of Copaiba Oils from *Copaifera Cearensis* Huber Ex Ducke, *Copaifera Reticulata* Ducke and *Copaifera Multijuga* Hayne—A Comparative Study. *Journal of Ethnopharmacology* **2007**, *112* (2), 248–254.
71. Veneziani, R. S.; Abrão, F.; Costa, L. D. A.; Ambrósio, S.; Bastos, J.; Martins, C. Antibacterial Evaluation of *Copaifera Langsdorffii* Oleoresin and Its Isolated Compounds against Multiresistant Bacteria. *Planta Medica* **2014**, *80* (10).
72. Wagner, M.; Fulton, B.; Henschel, R. Performance Optimization for the Trinity RNA-Seq Assembler. *Tools for High Performance Computing 2015* **2016**, 29–40.
73. Wan, Y.; Qu, K.; Ouyang, Z.; Kertesz, M.; Li, J.; Tibshirani, R.; Makino, D.L.; Nutter, R.C.; Segal, E.; and Chang, H.Y. Genome-wide measurement of RNA folding energies. *Mol. Cell* **48**, **2012** 169–181.
74. Wang, J.; Yu, W.; Yang, Y.; Li, X.; Chen, T.; Liu, T.; Ma, N.; Yang, X.; Liu, R.; Zhang, B. Genome-Wide Analysis of Tomato Long Non-Coding RNAs and Identification as Endogenous Target Mimic for MicroRNA in Response to TYLCV Infection. *Scientific Reports* **2015**, *5* (1).
75. Wang, Z.; Schwacke, R.; Kunze, R. DNA Damage-Induced Transcription of Transposable Elements and Long Non-Coding RNAs in Arabidopsis Is Rare and ATM-Dependent. *Molecular Plant* **2016**, *9* (8), 1142–1155.
76. Wang, T.; Zhao, M.; Zhang, X.; Liu, M.; Yang, C.; Chen, Y.; Chen, R.; Wen, J.; Mysore, K. S.; Zhang, W.-H. Novel Phosphate Deficiency-Responsive Long Non-Coding RNAs in the Legume Model Plant *Medicago truncatula*. *Journal of Experimental Botany* **2017**, *68* (21-22), 5937–5948.

77. Wang J, Meng X, Dobrovolskaya OB, Orlov YL, Chen M, Non-coding RNAs and their roles in stress response in plants. *Genomics Proteomics Bioinformatics* **2017** 15:301–312

78. Wang, L.; Xia, X.; Jiang, H.; Lu, Z.; Cui, J.; Cao, F.; Jin, B. Genome-Wide Identification and Characterization of Novel LncRNAs in Ginkgo Biloba. *Trees* **2018**.

79. Wang, Li, et al. "Genome-Wide Identification and Characterization of Novel LncRNAs in Ginkgo Biloba." *Trees*, **2018**, doi:10.1007/s00468-018-1724-x

80. Wilding, L. P.; Smeck, N. E.; Hall, G. F. *Pedogenesis and Soil Taxonomy*; Elsevier, **1983**.

81. Wu, H.-J., Wang, Z.-M., Wang, M., and Wang, X.-J. Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants. *Plant Physiol* **2013**. 161, 1875–1884. doi: 10.1104

82. Xin, M.; Wang, Y.; Yao, Y.; Song, N.; Hu, Z.; Qin, D.; Xie, C.; Peng, H.; Ni, Z.; Sun, Q. Identification and Characterization of Wheat Long Non-Protein Coding RNAs Responsive to Powdery Mildew Infection and Heat Stress by Using Microarray Analysis and SBS Sequencing. *BMC Plant Biology* **2011**, 11 (1), 61.

83. Xu, Q.; Song, Z.; Zhu, C.; Tao, C.; Kang, L.; Liu, W.; He, F.; Yan, J.; Sang, T. Systematic Comparison of LncRNAs with Protein Coding MRNAs in Population Expression and Their Response to Environmental Change. *BMC Plant Biology* **2017**, 17 (1).

84. Xuan H, Zhang L, Liu X, Han G, Li J, Li X, et al. PLNlncRbase: a resource for experimentally identified lncRNAs in plants. *Gene* **2015**;573:328–32

85. Zhang, X. et al. Maternally expressed gene 3 (MEG3) noncoding ribonucleic acid: isoform structure, expression, and functions. *Endocrinology* **2010**, 151, 939–947.

86. Zhang, W.; Han, Z.; Guo, Q.; Liu, Y.; Zheng, Y.; Wu, F.; Jin, W. Identification of Maize Long Non-Coding RNAs Responsive to Drought Stress. *PLoS ONE* **2014**, 9 (6).