RESEARCH

Genome-wide identification and characterization of *Fusarium circinatum*-responsive IncRNAs in *Pinus radiata*

Cristina Zamora-Ballesteros¹²*, Jorge Martín-García¹², Aroa Suárez-Vega³ and Julio Javier Diez¹²

**Abstract**

**Background:** One of the most promising strategies of Pine Pitch Canker (PPC) management is the use of reproductive plant material resistant to the disease. Understanding the complexity of plant transcriptome that underlies the defence to the causal agent *Fusarium circinatum*, would greatly facilitate the development of an accurate breeding program. Long non-coding RNAs (lncRNAs) are emerging as important transcriptional regulators under biotic stresses in plants. However, to date, characterization of lncRNAs in conifer trees has not been reported. In this study, transcriptomic identification of lncRNAs was carried out using strand-specific paired-end RNA sequencing, from *Pinus radiata* samples inoculated with *F. circinatum* at an early stage of infection.

**Results:** Overall, 13,312 lncRNAs were predicted through a bioinformatics approach, including long intergenic non-coding RNAs (92.3%), antisense lncRNAs (3.3%) and intronic lncRNAs (2.9%). Compared with protein-coding RNAs, pine lncRNAs are shorter, have lower expression, lower GC content and harbour fewer and shorter exons. A total of 164 differentially expressed (DE) lncRNAs were identified in response to *F. circinatum* infection in the inoculated versus mock-inoculated *P. radiata* seedlings. The predicted cis-regulated target genes of these pathogen-responsive lncRNAs were related to defence mechanisms such as kinase activity, phytohormone regulation, and cell wall reinforcement. Co-expression network analysis of DE lncRNAs, DE protein-coding RNAs and lncRNA target genes also indicated a potential network regulating pectinesterase activity and cell wall remodelling.

**Conclusions:** This study presents the first comprehensive genome-wide analysis of *P. radiata* lncRNAs and provides the basis for future functional characterizations of lncRNAs in relation to pine defence responses against *F. circinatum*.

**Keywords:** *Pinus radiata*, *Fusarium circinatum*, lncRNA, Transcriptomics network, Conifer defence, RNA-Seq.

**Background**

The study of the different types and functions of non-coding RNAs (ncRNAs) has recently gained prominence [1]. During the last decade, a new class of ncRNA, long non-coding RNA (lncRNA), has emerged as another eukaryotic transcript class where all transcripts greater than 200 nt in length that lack coding potential are included [2].

Similar to protein-coding genes, lncRNAs are transcribed by RNA polymerase II, capped, polyadenylated and usually spliced [3]. Accumulating evidence supports that lncRNAs participate in many cellular processes by regulating gene expression in a *cis-*regulatory manner, influencing genes around their transcription site, or leaving their transcription sites to exert their function elsewhere as a *trans-*acting transcript [4]. Sense and anti-sense, intergenic as well as intronic (located into an intron) are the main groups for classifying the lncRNAs according to their orientation with respect to the nearest protein-coding gene in the genome and genomic location [5]. Known
mechanism of action including molecular signalling, decoys (binding to regulatory elements such as miRNAs blocking their molecular interaction), guides (directing specific RNA-protein complexes to specific targets) and scaffolds as central platforms for regulation, are associated to the majority of IncRNAs [6].

The growing number of studies focusing on the interference of plant IncRNAs in different biological processes, including fertility, photomorphogenesis, wood formation, and biotic and abiotic stress, has demonstrated their important regulatory role in the transcription system [7–9]. Some of these IncRNAs have been experimentally validated, most of them being from model plants. For example in Arabidopsis, two IncRNAs, COOLAIR and COLDAIR, have been shown to be crucial in the regulation of cold stress response [10, 11]. Likewise, DRIR IncRNA regulates the expression of a series of genes involved in drought stress-responsive [12]. The regulatory role of the IncRNA IPS1 has also been reported blocking the miRNA mir399 that suppress the expression of the gene responsible for the phosphate uptake [13]. In Populus tomentosa, the interaction of the NERD gene and its regulatory IncRNA NERDL, is involved in the wood formation processes [14]. Moreover, some IncRNAs associated with biotic stress have been characterized in plants. These included IncRNAs that regulate positively the expression of defence-related PR genes such as ELENA1, identified in Arabidopsis as a factor enhancing resistance against the pathogen Pseudomonas syringae, and IncRNA39026 that increases resistance against Phytophthora infectans in tomato [15, 16]. The biosynthesis or signalling of plant hormones have been altered by IncRNAs as well. In cotton plants, the silencing of two IncRNAs (GhIncNAT-ANX2 and GhIncNAT-RLP7) led to increased resistance to Verticillium dahliae and Botrytis cinerea, possibly due to the transcriptional induction of two lipoxygenases involved in the jasmonic acid defence signalling pathway [17]. In addition, overexpression of IncRNA ALEX1 in rice increased jasmonic acid levels enhancing resistance to the bacteria Xanthomonas oryzae pv. oryzae [18].

Next Generation Sequencing (NGS) technologies and computational methods have enabled a deeper study of the transcriptomic data and have been widely applied for the identification and characterization of plant IncRNAs [19]. Recently, a number of IncRNAs involved in plant-pathogen interactions has been computationally predicted in non-model plants. In Brassica napus, 931 IncRNAs were identified in response to Sclerotinia sclerotiorum infection, one of them (TCONS_00000966) as antisense regulator of genes involved in plant defence [20]. Li et al. [21] discovered Musa acuminata IncRNAs related to resistance against Fusarium oxysporum f. sp. cubense infection. Particularly, IncRNAs involved in the expression of pathogenesis-related proteins and peroxidases were mainly induced in the resistant cultivar, whereas IncRNAs related to auxin and salicylic acid signal transductions could predominantly be induced in the susceptible cultivar. In the Paulownia witches’ broom disease interaction, nine IncRNAs were predicted to target twelve genes based on a co-expression network model in the tree [22]. In kiwifruit leaves infected by P. syringae, a weighted gene co-expression network analysis revealed a number of IncRNAs closely related to plant immune response and signal transduction [23]. Likewise, Feng et al. [24] identified 14,525 IncRNAs related to the walnut anthracnose resistance. This analysis showed that the target genes of the up-regulated IncRNAs were enriched in immune-related processes during the infection of the causal agent Colletotrichum gloeosporioides. These studies highlight the important role of IncRNAs in plant defence, thus further research is needed to decipher their function and interference in the transcriptomic system.

Fusarium circinatum is an invasive pathogen that causes Pine Pitch Canker (PPC). This disease affects conifers, resulting in a serious economic and ecological impact on nurseries and pine stands [25]. Since the first report in 1945 in North America, the presence of F. circinatum has been notified in 14 countries of America as well as Asia, Africa and Europe [26]. The long-distance dispersion as a result of globalization of plant trade and movement of contaminated soil and seed, represents the main pathway for new introductions of the pathogen into disease-free regions [27]. The establishment of the disease in field is of great concern since no feasible measures are available to control or eradicate F. circinatum [28]. Thus, the development of resistant genotypes through breeding and/or genetic engineering may be one of the most efficient PPC management strategies in the long-term [28, 29]. In this context, several transcriptome analyses with the aim of unravelling molecular defence responses have provided detailed insights about the molecular mechanisms underlying disease progression in the Pinus-F. circinatum pathosystem. These studies have examined the response of hosts through a different degree of susceptibility, from highly susceptible (Pinus radiata, Pinus patula) to moderate (Pinus pinaster) and highly resistant (Pinus tecunumanii, Pinus pinea) [30–35]. However, the role of IncRNAs in the regulation of defence network in conifers has not been studied yet. In the present study, a strand-specific RNA-Seq has been conducted in order to characterize IncRNAs present in high susceptible P. radiata and elucidate how IncRNA
expression profiles change in response to *F. circinatum* infection.

**Results**

**Disease monitoring**

The survival analysis revealed clear significant differences between the inoculation and control conditions ($\chi^2 = 116, p < 0.001$). At 10 days post inoculation (dpi) all seedlings inoculated with *F. circinatum* showed symptoms of PPC (resin and/or necrosis at the stem and wilting) and started to die at 33 dpi (Fig. 1). A seedling was considered dead when the pathogen had caused a girdling lesion and more than 75% of the needles were necrotic. By the end of the experiment, 92.2% of the inoculated seedlings had died. No mortality was recorded for control seedlings.

**Deep sequencing and transcripts assembly**

High-throughput strand-specific RNA-Seq of nine libraries constructed from stem tissue of *P. radiata* inoculated with *F. circinatum* and mock-inoculated were analysed. Raw data of the experiment have been deposited at the NCBI under the SRA numbers SRR15100123-31 (BioProject PRJNA742852). Almost 590 million 150-base pair-end reads on polyadenylated (polyA) selected RNAs were generated by the Illumina platform. RNA-Seq reached average depths of ca. 65.5 million reads (55 to 84 million reads; Table S1). After adapter and low-quality nucleotides trimming, an average of 78% of paired reads and 11% of mates from broken pairs were retained. Approximately 74.21% and 70.33% of reads from inoculated and mock-inoculated libraries successfully mapped to the *Pinus taeda* reference genome, respectively (Table S1). Considering the infected samples, the average of 2.63% reads mapped to the *F. circinatum* genome confirmed the presence of the pathogen.

Nine high-depth transcriptomes were generated. Six of them were reconstructed from *P. radiata* inoculated with *F. circinatum*, and the other three were generated from the mock-inoculated seedlings. After merging all of them, the unique transcriptome assembled were composed of 87,427 loci and 127,677 transcripts, with 43.1% GC content. A total of 51,212 (40.11%) transcripts were shared with the reference annotation file (Pita_v2.01.gtf) and discarded for lncRNA detection analysis since these transcripts were known as protein-coding RNAs. The remaining 76,465 transcripts were further categorized into different class codes according to its relationship with its closest reference transcript (Table 1).

**Genome-wide identification and characterization of pine lncRNAs**

The 76,465 total unknown transcripts were subjected to several sequential filter steps to obtain the lncRNA transcripts (Fig. 2). A total of 13,312 lncRNAs (length $\geq$ 200 nt, open reading frame coverage < 50%, and potential coding score < 0.5) and 47,473 potential new isoforms were obtained at the end of the pipeline. Using the FEELnc classifier module, the class distributions of the pine lncRNAs was performed according to their location relative to the nearest protein-coding gene based

---

**Fig. 1** Survival probability plot determined using the Kaplan-Meier estimate for *P. radiata* seedlings inoculated with *F. circinatum*
Table 1  Number of unknown pine transcripts associated to the potential class codes indicating unannotated transcripts according to GffCompare software classification

| Class code | After assembly<sup>a</sup> | LncRNAs predicted<sup>b</sup> | Description<sup>c</sup> |
|------------|-----------------------------|----------------------------|-------------------------|
|            | Transcript no. | %     | Transcript no. | %     |                                      |
| x          | 902             | 0.71  | 446           | 3.35 | Overlapping an exon of an annotated gene at the opposite strand |
| i          | 1,178           | 0.92  | 383           | 2.88 | Fully contained in a known intron |
| y          | 500             | 0.39  | 189           | 1.42 | Contains a reference gene within its intron |
| p          | 516             | 0.4   | 0             | 0    | Adjacent to the 5' end of an annotated gene at the same strand |
| u          | 45,705          | 35.8  | 12,280        | 92.32| Intergenic region |

<sup>a</sup> Number of transcripts belonging to each class code from the non-redundant set of transcripts generated after transcript assembly

<sup>b</sup> Number of transcripts belonging to each class code from the total number of predicted IncRNAs

<sup>c</sup> Brief explanation of the class codes

Fig. 2  The pipeline of IncRNA identification and functional prediction

- Strand-specific, 150 bp paired-end RNA-Seq RAW DATA
  - Preprocessing of reads
    - FASTQC + Trimmomatic
  - Mapping and transcript assembly
    - HISAT2 + StringTie

  **127,077 transcripts**

- Transcription class code
  - (TCC) designation
    - GffCompare

- Transcript expression
  - StringTie -e

  **3,750 IncRNAs annotated with adjacent genes**

  **14 IncRNAs annotated with co-expression genes**

  Gene ontology and KEGG pathways enrichment analysis of IncRNA target genes
  - GOSeq/CEMiTool

- Potential cis-acting IncRNAs

- Potential trans-acting IncRNAs

- Location comparison with protein-coding genes

- Co-expression network analysis
  - CEMiTTool

- Filtering based on machine learning: Coding Potential Score < 0.5
  - FEELnc_codpot

- Identity IncRNAs and their partner genes
  - FEELncClassifier

  **13,312 putative IncRNAs**

- Filtering out
  - Annotated transcripts
    - EnTAP
  - Transcript length > 200 bp
    - Exon number > 2
    - FEELnc_filter

  **76,465 transcripts**

- Filtering out:
  - FPKM < 0.1
  - TCC "="
on the reconstructed transcriptome. The majority of the IncRNAs were long intergenic non-coding RNAs (lincRNAs) with 12,291 (92.3%) transcripts, followed by long non-coding natural antisense transcripts (IncNATs) with 445 (3.3%) transcripts and 383 (2.9%) intronic transcripts. In addition, 25 IncRNAs were also identified as known miRNA precursors belonging to 10 miRNA families being the most represented MIR160, MIR159 and MIR1314. The Rfam and miRBase analyses also allowed the identification of 174 transcripts that were found to be distributed among 32 conserved RNA families including ribosomal RNA (rRNAs), transfer RNA (tRNAs), histones and several small nucleolar RNAs (snoRNAs; Table S2-S3).

The average length of protein-coding transcripts (1,200 bp) was higher than that of lincRNAs (750 bp), IncNATs (452 bp) and intronic IncRNAs (565 bp). However, while most of IncNATs and intronic IncRNAs showed short lengths (300 bp), lincRNAs and protein-coding transcripts exhibited a similar trend of length distribution (Fig. 3A). Overall, the size distribution of the IncRNAs ranged from 200 to 7,393 bp, with the majority of these transcripts ranging from 200 to 400 bp. Differences in the analysis of the exon number were also found. While the IncRNAs showed an average exon number of 2.5, the protein-coding transcripts had 4.1 exons (Fig. 3B). This analysis also revealed that two-exon transcripts were the most represented in this study. The highest ratio of two-exon transcripts was found in IncNATs (77.3%) and intronic IncRNAs (75.7%), followed by lincRNAs (66.9%). In the group of protein-coding transcripts, the ratio of two-exon transcripts was not so high (32%). Regarding the exon length, similarly to the transcript length, the exons belonging to the IncNAT and intronic IncRNA transcripts showed shorter lengths (100-300 bp) than those belonging to protein-coding transcripts (Fig. 3C). Once again, the distribution of the exon lengths from

![Fig. 3](https://example.com/fig3.jpg)  
**Fig. 3** Characterization of IncRNA transcripts showed differences with the characteristics of protein-coding transcripts in *P. radiata*. **A** Transcript size distribution for lincRNAs, IncNATs, intronic IncRNAs and protein-coding RNAs. **B** Number of exons per transcript for lincRNAs, IncNATs, intronic IncRNAs and protein-coding RNAs. **C** Exon size distributions for lincRNAs, IncNATs, intronic IncRNAs and protein-coding RNAs. **D** FPKM distribution of lincRNAs and protein-coding RNAs. **E** GC content of lincRNAs and protein-coding RNAs.
the lincRNA transcripts was similar to that of protein-coding transcripts.

The average expression levels of lncRNAs in terms of fragment per kilobase of exon per million mapped reads (FPKM) was lower (3.3) than those of protein-coding transcripts (5.6; Fig. 3D). In addition, the GC content in lncRNAs (41%) was slightly lower than that in protein-coding transcripts (44.8%), showing the intronic lncRNA transcripts the lowest percentage (Fig. 3E).

All the lncRNA transcripts were aligned against the known lncRNAs of 10 different plant species from the CANTATA database: Chenopodium quinoa, Brassica napus, Malus domestica, Zea mays, Arabidopsis thaliana, Oryza rufipogon, Vitis vinifera, Populus trichocarpa, Prunus persica and Ananas comosus. Likewise, known lncRNAs of all plant species present in the GreeNc database, except those species already examined with the CANTATA database, were confronted with the lncRNAs of P. radiata. A number of 1,131 (8.6%) lncRNAs were conserved across the ten species of CANTATA (Table S4). In addition, a total of 1,421 (10.8%) lncRNA transcripts, corresponding to known lncRNA genes from the GreeNc database (Table S5), were obtained. Therefore, 2,552 (19.3%) lncRNAs showed homology with known lncRNAs from other plant species. The highest homology ratio (number of hits of pine lncRNAs with those of each plant species to the total number of lncRNAs of each plant species) was observed with the woody plant P. trichocarpa (5.03%; Figure S1).

**Differential expression analysis in response to F. circinatum infection and prediction of candidate target genes**

The expression changes of lncRNAs between the P. radiata seedlings inoculated with F. circinatum and controls were analysed. The principal component analyses (PCA) allowed to identify two sample outliers among the pathogen-inoculated condition that were discarded for the differential expression analysis (Figure S2). A total of 164 lncRNA transcripts were identified as differentially expressed (p-value < 0.05, log2 (|Fold-change|) ≥ 1) under the pathogen infection, 146 of which were up-regulated and 18 down-regulated (Table S6-S7). Among the differentially expressed (DE) lncRNAs, 157 were lincRNA transcripts and the remainder were two intronic lncRNAs, one lncNATs, and four lncRNA transcripts containing a coding-protein in its intron. DE lncRNAs were clustered in a heat map in order to visualize the expression pattern of both conditions of the analysis (Fig. 4). On the other hand,
2,369 protein-coding RNA were up-regulated and 189 down-regulated by the pathogen infection (Table S8-S9).

**Analysis of IncRNAs cis-interacting genes**

To predict the role of cis-acting IncRNAs of *P. radiata* in response to *F. circinatum*, the protein-coding transcripts located within a 10 kb window upstream and 100 kb downstream were investigated. A total of 4,268 IncRNA–mRNA interaction pairs were recorded by the FEELnc classifier module (Table S10). However, one IncRNA could have more than one target gene, and a target gene could be the target of one or more IncRNAs. In fact, a number of 2,760 candidate cis target genes were observed for 3,750 IncRNAs, of which 3,342 had a single candidate target gene and 408 IncRNAs had multiple interactions. The maximum number of target genes for a single IncRNA was five, which was reached by seven IncRNAs (Table S11). Moreover, the 73% of the 2,760 candidate target genes were targeted by one IncRNA, while one candidate target gene could be targeted by up to 30 different IncRNAs.

In total, 39 candidate target genes were predicted for the 37 DE IncRNAs (Table 2). The function prediction of these DE IncRNAs was based on the functional annotation of their nearby target genes. Among these targeted genes, there were genes encoding for receptor-like protein kinases (RLKs), enzymes associated to the cell-wall reinforcement and lignification (pectin methylesterases inhibitor, uclacyanin and 4-coumarate-CoA ligase), and enzymes involved in the attenuation of oxidative stress (glutathione S-transferase). One RLK that was predicted to be targeted by the up-regulated IncRNAPiRa.29753.1 was, in turn, induced by the pathogen infection. Two pectin methylesterases (PME) were predicted to be regulated by IncRNAPiRa.23041.2 and IncRNAPiRa.22160.1 transcribed in the same orientation in a downstream location. One of the targeted PME was DE by the pathogen infection, whereas the other PME did not. Moreover, the coding region for 4-coumarate-CoA ligase 3 (4CL3) targeted by IncRNAPiRa.33098.2 was also present among the differentially expressed genes (DEGs) of the coding RNAs analysis. One gene harbours the DNA-binding motif MYB, a transcription factor with a role in plant stress tolerance, was potentially regulated by a IncNAT (IncRNA-PiRa.31525.1). The IncRNAPiRa.85000.6 IncRNA, which was predicted to target an ethylene receptor 2 (*ETR2*) gene involved in the ethylene signal transduction pathway, was transcribed in the same strand and orientation than its RNA partner from an upstream location. In addition, two genes encoding for photoassimilate-responsive protein 1 (*PAR1*) were predicted to be targeted by IncRNAPiRa.61651.3 and IncRNAPiRa.33277.3, the latter being DE between conditions.

The pine IncRNA IncRNAPiRa.79902.12 was predicted to target two genes encoding for the pyruvate decarboxylase 1 (*PDC1*) enzyme, which both were up-regulated by the pathogen infection. Furthermore, one gene that participates in chromatin modifications (chromatin remodelling 24) and three genes that contain canonical RNA-binding domains (pentatricopeptide repeat-containing protein, ribosomal RNA methyltransferase FtsJ domain containing protein, CCCH-type ZnF protein) were predicted to be targeted in an antisense manner by IncRNAs. None of the latter three genes were differentially expressed in the coding RNAs expression analysis.

The enrichment analysis of Gene Ontology (GO) terms and KEGG pathways of the nearby protein-coding RNAs revealed potential functions in which DE IncRNAs could be involved (Fig. 5). The three target genes regulating the down-regulated IncRNAs were not associated to any GO term neither KEGG pathway, thus the analysis showed results only for the up-regulated IncRNAs (Table S12). Biological and metabolic processes were the most representative GO terms for the biological process category, followed by macromolecule metabolic process and response to stimulus and stress in this dataset. Several GO terms associated with low-oxygen conditions including response to hypoxia and response to decreased oxygen levels were enriched. In addition, catabolism and metabolism of allantoin were also enriched. Genes involved in cell periphery and cell wall were represented for cellular components. For molecular functions, the pine IncRNAs were enriched for GO terms such as catalytic activity, binding and hydrolase activity. The KEGG pathways enriched in the target genes of the up-regulated IncRNAs were ‘glycolysis/gluconeogenesis’ and ‘microbial metabolism in diverse environments’ (Table S13).

**Co-expression gene modules associated with *P. radiata* defence response**

A dendrogram, in which the samples were clustered according to their condition using the CEMiTool package, was generated (Fig. 6A). The modular expression analysis revealed genes that may act together or are similarly regulated during the defence responses to *F. circinatum* infection. The dissimilarity threshold of 0.8 was used as a cut-off on hierarchical clustering, which identified two co-expression modules (Fig. 6B and C). The largest module contained 320 co-expressed transcripts (M1): 307 DEGs, 13 DE IncRNAs, and three targeted genes (*PDC1*, PME and RLK; Table S14). Transcripts in M1 were enriched mainly for biological processes related to the pectinesterase activity and cell wall remodeling among others (Fig. 6D; Table S15). Indeed, three DEGs encoding for pectin methylesterase 17 were identified as gene hubs in this module (Table 3). The
### Table 2  Candidate target genes predicted to interact with DE IncRNA transcripts

| LncRNA               | Log₂FC | Targeted gene | Log₂FC | Direction and type | Location | Description of targeted gene                                                                 |
|----------------------|--------|---------------|--------|--------------------|----------|-----------------------------------------------------------------------------------------------|
| lncRNAPiRa.44237.18 | 9.76   | PITA_00496    |        | Convergent, downstream | Pentatricopeptide repeat-containing protein A4g13650 |
| lncRNAPiRa.64325.1  | 3.15   | PITA_01014    |        | Containing, exonic   | Transcript with domain: DUF4228                   |
| lncRNAPiRa.32343.2  | 11.3   | PITA_13284    |        | Convergent, downstream | CYCD2                                             |
| lncRNAPiRa.35491.1  | 9.18   | PITA_33574    |        | Convergent, downstream | Leaf rust 10 disease-resistance locus receptor-like protein kinase-like 1.2 isoform X1         |
| lncRNAPiRa.42942.2  | 8.54   | PITA_15284    |        | Divergent, upstream  | Ribosomal RNA methyltransferase FtsJ domain-containing protein                                |
| lncRNAPiRa.22160.1  | 9.5    | PITA_12411    | 6.58   | Same strand, downstream | Pectin methyltransferase 17                        |
| lncRNAPiRa.31525.1  | 9.63   | PITA_31792    |        | Convergent, downstream | Transcript with domain: Myb_DNA-binding           |
| lncRNAPiRa.79902.12 | 7.12   | PITA_05666    | 10.1   | Same strand, upstream | PDC1                                               |
| lncRNAPiRa.70333.4  | 7.9    | PITA_01539    | 11.9   | Same strand, downstream | PDC1                                               |
| lncRNAPiRa.51697.3  | 3.35   | PITA_34628    |        | Containing, intronic  | Uclacyanin 1                                      |
| lncRNAPiRa.61651.3  | 8.53   | PITA_42898    |        | Convergent, downstream | PAR1                                               |
| lncRNAPiRa.33277.3  | 3.35   | PITA_08467    | 3.54   | Convergent, upstream  | PAR1                                               |
| lncRNAPiRa.45077.2  | 8.14   | PITA_26106    |        | Convergent, downstream | Purple acid phosphatase                           |
| lncRNAPiRa.23041.2  | 9.27   | PITA_28262    |        | Same strand, downstream | Pectin methyltransferase 17                        |
| lncRNAPiRa.85490.1  | 6.99   | PITA_28228    |        | Divergent, upstream  | unknown [Picea sitchensis]                        |
| lncRNAPiRa.47042.1  | 7.53   | PITA_13092    |        | Same strand, upstream | Transcript with domain: PP2C Non-symbiotic hemoglobin 1 (HB)                                |
| lncRNAPiRa.19024.1  | 5.58   | PITA_42377    | 5.17   | Same strand, downstream | Transcript with domain: Peptidase S9 Peptidease class-I                                      |
| lncRNAPiRa.25700.7  | 3.79   | PITA_23327    |        | Containing, intronic  | Peptidase S9                                      |
| lncRNAPiRa.25968.1  | 2.91   | PITA_42840    | 4.85   | Same strand, downstream | Peptidase class-USP                               |
| lncRNAPiRa.29628.1  | 6.8    | PITA_10474    |        | Same strand, downstream | Transcript with domain: Glyc-olytic-Fructose-bisphosphate aldolase class-I                   |
| lncRNAPiRa.80857.1  | 6.78   | PITA_28959    |        | Nesting, intronic    | ALN                                                |
| lncRNAPiRa.29753.1  | 7.2    | PITA_38537    | 6.4    | Same strand, downstream | leaf rust 10 disease-resistance locus receptor-like protein kinase-like protein 2.4          |
| lncRNAPiRa.33098.2  | 6.8    | PITA_43179    | 5.01   | Same strand, downstream | 4-coumarate-CoA ligase, partial (4CL3)            |
| lncRNAPiRa.80336.1  | 4.89   | PITA_17252    |        | Convergent, downstream | Protein chromatin remodeling 24                  |
| lncRNAPiRa.61651.4  | 5.82   | PITA_42898    |        | Convergent, downstream | unknown [Picea sitchensis]                        |
| lncRNAPiRa.64704.5  | 6.91   | PITA_22879    |        | Same strand, downstream | Transcript with domain: GSTL1                     |
| lncRNAPiRa.75647.1  | 5.93   | PITA_04032    |        | Divergent, upstream  | Transcript with domain: RRM_1                     |
| lncRNAPiRa.85000.6  | 9.91   | PITA_16990    |        | Same strand, upstream | Ethylene receptor 2 (ETR2)                       |
| lncRNAPiRa.33190.1  | 7.85   | PITA_44567    |        | Same strand, downstream | Transcript with domain: EamA                      |
| lncRNAPiRa.31184.1  | 2.25   | PITA_16807    |        | Divergent, downstream | Transcript with domain: LEA_3                     |
second module (M2) consisted of 30 DEGs and one DE lncRNA (Table S14), however, no significant GO terms were identified. The top gene hubs of both modules are shown in Table 3.

**Discussion**

Over the past decade, the complexity of eukaryote genome expression has become apparent mainly due to the development of next-generation sequencing...
technologies. Particularly, the sequencing of RNA (RNA-Seq) has revealed an important part of non-coding transcriptome that should not be ignored. Indeed, a large number of studies have recently reported lncRNAs to be essential in the regulation of a wide range of biological and molecular processes by activating their nearby protein-coding genes using a cis-mediated mechanism or distant genes in a trans-acting manner [36]. Stress conditions lead to transcriptomic reprogramming where lncRNAs also play a key role. In plants, numerous lncRNAs under biotic stress have been identified to date, although further studies for non-model plants are still required. In the last years, the transcriptomic responses of conifers to fungal infections have been increasingly studied. In particular, several transcriptomic studies have demonstrated that the F. circumatum infection causes substantial changes in the pine gene expression [30–35]. However, to our knowledge, no reports investigating the long non-coding RNAs of conifer trees in response to fungal attacks have been published so far. The results reported here, therefore, provide a first insight into the regulatory mechanisms of lncRNAs involved in defence reactions against F. circumatum of a highly susceptible species such as P. radiata at an early stage of infection.

The percentage of P. radiata reads mapped the P. taeda reference genome conforms to acceptable mapping ratios [37], which can be expected as genome-wide comparisons between both species have revealed a significant collinearity between their genomes using restriction fragment length polymorphisms (RFLPs) and microsatellite markers [38]. Although P. taeda and P. radiata belong to different subsections (Australes and Oocarpace, respectively), they are in the same subgenus (Pinus) and section (Pinus). Therefore, in absence of conifer genome sequence of the species of interest, the use of those of taxonomically related species in order to perform an unannotated transcript identification pipeline can be employed to assist and improve the transcript assembly process when using short reads [39].

The combination of the strand-specific RNA-Seq approach and high coverage sequencing (up to 84 million reads per sample) allowed the identification of lncRNAs that are commonly expressed at low levels and lncNATs that would otherwise have been difficult to find [40]. Moreover, the visual analysis of the PCA identified two sample outliers among the pathogen-inoculated condition (Figure S2), which may be due to technical failures during the multi-step process of RNA-Seq experiment (mRNA isolation, reverse transcription, library construction and sequencing). Since these errors, known as batch effects, lead to decreased statistical power [41], outliers were discarded for the differential expression analysis. Overall, a total of 13,312 lncRNAs were identified from the P. radiata transcriptome, of which 164 were F. circinatum-responsive lncRNAs comprised mainly by intergenic lncRNAs (Table S6-S7). This is consistent with previous analyses where the number of lncRNAs in response to biotic stress was comparable. In Paulownia tomentosa, two similar studies found 112 and 110 lncRNAs to be involved in phytoplasma infection [22, 42]. Similarly, among 94 and 302 lncRNAs were identified in susceptible and resistant M. acuminata roots in response to F. oxysporum f. sp. cubense, with the highest value in the resistant roots after 51 h post-inoculation [21]. The number of S. scleroti-, rum-responsive lncRNAs was slightly higher in B. napus with 662 at 24 h decreasing until 308 at 48 h [20]. In addition, intergenic lncRNAs were also the most abundant responsive transcripts in all these studies. Therefore, the pattern appears to follow the same trend in conifer trees.

In general, lncRNAs demonstrate low and tissue-specific expression patterns and lack of conservation [3, 43, 44]. Indeed, lncRNAs of P. radiata showed lower expression than the protein-coding RNAs, and only 19.3% of them were conserved among 46 different plant species. However, the low level of transcriptome conservation in P. radiata to angiosperms has also been shown in xylem tissues (15-32%; E-value ≤ 10^{-5}), compared with the highly conserved xylem transcriptome within conifers (78-82%; E-value ≤ 10^{-5}) [45]. Thus, it may not be a characteristic of conifer lncRNAs. The genomic features of the lncRNA transcripts of P. radiata were consistent with those previously characterized in other organisms [46]. As expected, the lncRNAs were shorter in terms of overall length and contained lower number of exons (Fig. 3 A and B). The analysis showed a bias toward two-exon lncRNAs, which is explained by the retention of only monoexonic transcripts with antisense localization in the identification process of lncRNAs. The length of the exons was also shorter in lncNATs and intronic lncRNAs when comparing with

(See figure on next page.)

**Fig. 6** Two co-expression modules were identified among DE IncRNAs, DEGs and targeted genes using CEMiTool package. (A) Dendrogram of samples clustered according to their condition. (B) Gene set enrichment analysis (GSEA)-based identification of two gene co-expression modules. Red coloring denotes a positive NES score, while blue coloring denotes a negative NES score. (C) Expression profiles for both expression modules (M1, M2). Each line represents a transcript and its change in expression across conditions. (D) Barplot for top GO terms enriched in M1 module. x-axis and colour transparency display - log_{10} of the Benjamini-Hochberg (BH)-adjusted p-value. Dashed vertical line indicates BH-adjusted p-value threshold of 0.05.
Fig. 6 (See legend on previous page.)
protein-coding RNAs, however, the distribution of the length of exons belonging to lncRNAs was closer to that of the protein-coding transcripts (Fig. 3C). In this regard, some exceptions have been found in other plants such as cotton (Gossypium arboretum) and chickpea (Cicer aritinum) where the exon length of the lncRNAs was even longer than protein-coding RNAs [47]. The GC content of the assembled transcripts of P. radiata (43.1%) was similar to that of the transcriptome of other Pinus species such as P. tecunumanii (44%) [32]. Separately, the GC content in pine lncRNAs (41%) was lower than in protein-coding RNAs (44.8%), which had been reported before as a common feature of lncRNAs due to different evolutionary pressures in ORFs [48].

The role of lncRNAs in the positive or negative regulation of gene expression is well known [3]. One of the conserved mechanisms of action of the lncRNAs is their function as decoys by sequestering RNA-binding proteins (RBP), miRNAs or chromatin-modifying complexes [6]. Thus, the lncRNA ultimately inhibits its particular function. Several DE lncRNAs of P. radiata inoculated by F. circinatum seem to fit into this functional mechanism. Four antisense lncRNAs were predicted to target genes encoding RBPs including pentatricopeptide repeat-containing protein (PPR2), ribosomal RNA methyltransferase FtsJ domain containing protein, CCCH-type zinc finger protein and RNA recognition motif (RRM) containing protein. Moreover, another antisense DE lncRNA was predicted to target a chromatin-remodelling gene (Table 2). Therefore, the reprogramming exerted by the infection of F. circinatum on pine transcription affects not only the protein-coding genes, but also the non-coding part of the genome.

The induction of plant defences is a complex biological process that causes a dramatic transcriptomic reprogramming throughout the genome [49]. Previous studies have shown that a vast number of genes are either up- or down-regulated in response to F. circinatum infection [31, 33–35]. Several functional groups of genes have repeatedly been identified as induced upon the pathogen infection. These groups include signal perception and transduction, biosynthesis of defence hormone and secondary metabolites, and cell wall reinforcement and lignification. Some of the GO terms enriched by the potential target genes of the IncRNAs identified in this study were related to these functional groups including biological processes such as cell wall modification and signalling of the abscisic acid (ABA), ethylene (ET) and cytokinin hormones (Table S12). These results suggest for the first time that the IncRNAs may play a key role in the process of pine defence to F. circinatum as previously reported in other pathosystems [8, 50]. Indeed, the enrichment of various GO terms related to the ABA signalling suggests an involvement of the pine IncRNAs in this pathway. In turn, ABA accumulation has been previously associated with increased PPC susceptibility [51–54], therefore, a deeper investigation of these IncRNAs in the pine is needed in order to better understand the complex regulation of ABA responses.

Plant signalling molecules such as protein kinases, reactive oxygen species (ROS) and hormones are critical in mounting an appropriate defence response [55]. Genes with kinase activity have a role in signal transduction triggering the downstream signalling. Two genes with predicted functions in receptor-like kinase were cis-regulated by IncRNAs, being one of them DE by the pathogen infection (Table 2). The other one was potentially regulated by a IncNAT. Positive cis-regulatory feature of NATs by mediating histone modifications at the locus has been previously reported [44]. This behaviour has been also seen in LAIR, a rice IncNAT that up-regulates the expression of its neighbour leucine-rich repeat receptor kinase [56]. Despite that a large number of genes [43] encoding glutathione S-transferases (GSTs) were up-regulated under the pathogen infection (Table S8), the GST predicted to be regulated by the downstream IncRNAPiRa.64704.1 was not among the DEGs. Joshi et al. [20] also identified one IncRNA of B. napus located in the upstream of a gene encoding for a GST in response to S. sclerotiorum infection. GST genes are highly induced under biotic stress due to their role in the attenuation of oxidative stress and the participation in hormone transport [57]. In addition, a transcript predicted to encode a non-symbiotic hemoglobin 1, which is involved in ROS and NO scavenging [58], was DE in the analysis and predicted to be targeted by IncRNAPiRa.19024.1 (Table 2).

These findings seem to indicate that IncRNAs could be also involved in the cell detoxification after an oxidative burst provoked by a fungal infection.

### Table 3 Potential gene hubs of each co-expression gene module

| Transcript | Description |
|------------|-------------|
| **Hub genes - M1** |
| PITA.22172.1 | Pectin methylesterase 17 |
| PITA.22173.1 | Pectin methylesterase 17 |
| PITA.04671 | Pectin methylesterase 17 |
| PITA.84236.10 | Alcohol dehydrogenase, partial (ADH1) |
| PITA.08271 | Early nodulin-93-like |
| **Hub genes - M2** |
| PITA.37728.4 | 2-methylene-furan-3-one reductase |
| PITA.32547.3 | unknown |
| PITA.69828.1 | hypothetical protein |
| PITA.75382.2 | Glutathione S-transferase, partial (GST) |
| PITA.87100.2 | Phophytinase, chloroplastic-like |
Phytohormones trigger an effective defence response against biotic stress [59]. Several studies have pointed to IncRNAs as participants in the complex network of hormone regulation. In *M. acuminata* infected by *F. oxysporum* f. sp. *cubense*, IncRNAs were found to be predominantly associated with auxin and salicylic acid signal transduction in susceptible cultivars, whereas all phytohormones were potentially regulated by IncRNAs in resistant cultivars [21]. Genes related to the salicylic acid-mediated defence process were co-expressed with IncRNAs in kiwifruit plant challenged with the bacteria *P. syringae* [23]. Likewise, IncRNAs of resistant walnuts to *C. gloeosporioides* were predicted to trans-regulate genes involved in defence pathways of the jasmonic acid and auxins [24]. A previous transcriptome analysis of *P. radiata* showed the induction of abscisic acid signalling under the infection of *F. circinatum* [31]. A type 2 C protein phosphatase (PP2C) family gene, which negatively regulates abscisic acid responses [60, 61], could be regulated by IncRNA*PiRa.47042.1* located upstream in the same strand despite not belonging to the DEGs (Table 2). The implication of this IncRNA in the abscisic acid signalling regulation would need further investigation.

The phytohormone ethylene represents one of the core components of the plant immune system [62]. When ethylene binds with its ETRs activates the transcriptional cascade of ethylene-regulated genes [63]. Seedlings of *P. tecunumanii*, *P. patula*, *P. pinea* and *P. radiata* inoculated with *F. circinatum* have demonstrated to induce ethylene biosynthesis and signalling genes [31, 33, 35]; however, only *ETR2* has been found to be induced in the moderate resistant specie *P. pinaster* at 5 and 10 dpi [34]. Under stress conditions, when the concentration of ethylene is high, the transcription of *ETR2* contributes to the stabilization of ethylene levels by attenuating its signalling output and restore the ability to respond to subsequent ethylene signal [64]. In the present study, *ETR2* has not been DE in *P. radiata* but was presumably influenced by IncRNA*PiRa.85000.6*, which has been DE by *F. circinatum* (Table 2). Therefore, we can hypothesize that the ethylene response seems to be fine-tuned in *P. pinaster*, which does not occur in *P. radiata*, possibly due to the influence of this IncRNA located upstream of its transcription. It would be worthwhile to further investigate the regulatory function of this IncRNA as it could be a key factor in overcoming the PPC disease.

The potential function of IncRNAs in wood formation has been previously observed in different plant species. In a study of cotton IncRNAs, these were enriched for lignin catabolic processes and their role in lignin biosynthesis by regulating the expression of LAC4 was suggested [65]. In *Populus*, 16 genes targeted by IncRNAs were involved in wood formation processes, including lignin biosynthesis [9], and 13 targeted genes were associated to cellulose and pectin synthesis [66]. In addition, the IncRNA NERDL regulates the Needed for rdr2-independent DNA methylation (NERD) gene, which is also involved in the wood formation in *Populus* [67]. The enzyme that catalyse the hemicellulose xyloglucan was predicted to be targeted by an IncRNA of *Paulownia tomentosa* and had a role in the hyperplasia caused by a phytoplasma infection [22]. Cell wall reinforcement and lignification are the most common induced defences against pathogens, for that, the cell wall suffers a remodelling process that has been documented in the *P. radiata-F. circinatum* pathosystem [31, 35]. The demethylesterification of pectin, controlled by PMEs, is considered to affect the porosity of the cell wall and, thus, exposes the plant to an easier degradation by pathogen enzymes [68]. However, PME activity has been also associated with the activation of plant immunity and resistance against pathogens [69]. In a recent study, in contrast to *P. radiata*, the resistant species *P. pinea* infected by *F. circinatum* showed a high induction of pectin methyltransferase inhibitor (PMEI) genes and an inhibition of PMEs [35]. In this study, two IncRNAs were predicted to target two PMEs, one of them was up-regulated by the pathogen infection, which could suggest a positive regulation from the IncRNA activity (Table 2). In addition, the co-expression analysis of *F. circinatum* responsive IncRNAs and mRNAs indicated a clear enrichment for PME activity (Fig. 6D). The transcriptional regulation of these enzymes could be related to the susceptibility of *P. radiata* and would be worth further investigation. Another gene containing a cellulase domain was also up-regulated in the expression analysis of protein-coding RNAs and predicted to be regulated by an induced IncRNA (Table 2). Moreover, the analysis identified a potential IncRNA cis-regulating positively a gene encoding for 4CL3 (Table 2), one of the key enzymes of the phenylpropanoid pathway. In plants, this pathway leads to the production of secondary metabolites and cell wall lignification, both associated to plant defence. The transcriptional regulation of the 4CL gene by IncRNAs has been also reported in *P. tomentosa*, that together with the targeted gene encoding the caffeoyl-CoA 3-O-methyltransferase (CCOMT) enzyme by another IncRNA, highlighted the potential role of these molecules in lignin formation in wood with different properties [9]. These findings provide increasing evidence for the involvement of IncRNAs in cell wall remodelling and lignification process.

Although the role of the hypoxia in the plant-pathogen interaction has not yet been determined, hypoxia-responsive genes have been reported to be induced in some plants during pathogen infections [70]. Indeed, the analysis of DEGs showed that a large number of
genes encoding for PDC1 and alcohol dehydrogenase 1 (ADH1), which are required in the fermentative pathway under low-oxygen conditions, were highly induced by *F. circinatum* infection (>10 log2[fold change]; Table S8). Among them, two PDC1 were potentially targeted by two pine IncRNAs (Table 2). This together with the functional analysis results of the IncRNAs where several enriched GO terms were associated to hypoxia suggests a role of pine IncRNAs in an insufficient oxygen situation.

**Conclusions**
In summary, the computational analysis allowed to identify 13,312 IncRNAs in *P. radiata*. Compared to the protein-coding RNAs, the IncRNAs were shorter, with fewer exons and showed lower expression levels. In total 164 IncRNAs were reported as responsive to *F. circinatum* infection. GO enrichment of genes that either overlap with or are neighbours of these pathogen-responsive IncRNAs suggested involvement of important defence processes including signal transduction and cell wall reinforcement. These results present a comprehensive map of IncRNAs in *P. radiata* under *F. circinatum* infection and provide a starting point to understand their regulatory mechanisms and functions in conifer defence. In turn, a thorough understanding of the mechanism of gene regulation will contribute to the improvement of breeding programs for resistant pine commercialization, one of the most promising approaches for PPC management.

**Methods**

**Inoculum preparation and inoculation trial**
The *F. circinatum* isolate 072 obtained from an infected *P. radiata* tree in the North of Spain (Cantabria, Spain) was used. The isolate was cultured in Petri dishes containing PDA medium (Scharlab S.L., Spain) for a week at 25 °C. Then, to stimulate the sporulation of the fungus, four mycelial agar plugs were subcultured in an Erlenmeyer flask with 100 mL of PDB medium (Scharlab S.L., Spain) and incubated in an orbital shaker at 150 rpm during 48 h at 25°C. Afterwards, the conidial suspension was adjusted with a haemocytometer at 10⁶ conidia mL⁻¹ for the inoculation.

Six-month-old seedlings from *P. radiata* seeds originating from the same provenance (Galicia, Spain), which had been assessed and provided by the Consellería do Medio Rural (Xunta de Galicia, Spain), were used for the inoculation trial. The plants, with an approximate stem diameter of 2.5 ± 0.5 cm, were inoculated on the stem by making a wound with a sterile scalpel and pipetting 10 µL of conidial suspension [71]. The same process was applied for the control seedlings that were mock-inoculated with sterilized distilled water. The inoculated wound was immediately sealed with Parafilm® to prevent drying. Sixty seedlings were inoculated for each treatment (inoculation with pathogen and mock-inoculation). Plants were placed in a growth chamber at 21.5 °C with a 14-h photoperiod and kept for 67 days during which symptoms were monitored according to the scale of symptoms (slightly modified) described by Correll et al. [72], where 0 = healthy plant, 1 = resin and/or necrosis at the point of inoculation and healthy foliage, 2 = resin and/or necrosis beyond the point of inoculation, 3 = accentuated wilting and appreciable dieback, 4 = dead plant. Mortality rates were daily recorded.

The survival analysis based on the non-parametric estimator Kaplan-Meier [73] was performed with the “Survival” package [74] to test the mortality of the plants. Survival curves were created with the “Survfit” function and the differences between the curves were tested with the “Survdiff” function. All analyses were performed using R software environment [75].

**RNA extraction and paired-end strand-specific sequencing**
A piece of the stem from the upper part of the inoculation point (ca. 1 cm length) was sampled at four dpi for the transcriptomic analysis. The timing of sampling was chosen since it could be an adequate representation of the initial phase of the infection process, which has been previously established as a seven-day period [76]. The harvested tissues were immediately frozen in liquid nitrogen and ground to a fine powder using a mortar and pestle. RNA extractions were performed using the Spectrum™ Plant Total RNA Kit (Sigma Aldrich, USA) following the manufacturer’s protocols including the optional on-column DNase 1 digestion (DNASE10-1SET, Sigma-Aldrich, St. Louis, MO, USA). After RNA extraction, samples were transferred to RNase- and DNase-free tubes (Axygen®, USA) and stored at -80 °C. The concentration and purity of the RNA extracted were measured using the Multiiskan GO Spectrophotometer (A₂₆₀/A₂₃₀ ≥ 1.8, A₂₆₀/A₁₂₀ ≥ 1.8 and concentration > 50 ng/μL; Thermo Fisher Scientific, Waltham, MA, USA). RNA integrity was checked by agarose gel electrophoresis (1% TAE).

Six biological replicates of inoculated and three of mock-inoculated treatment randomly selected from the inoculation trial were sent to Macrogen Co. (Seoul, South Korea) for sequencing. Sequenced samples showed a RNA integrity number (RIN) ≥ 7 measured by an Agilent 2100 Bioanalyzer. The strand-specific RNA-Seq libraries were constructed using the Illumina TruSeq Stranded mRNA protocol with polyadenylated
mRNAs and lncRNAs enrichment and an insert size of 300 bp (150 × 2 paired-end reads). Sequencing was performed on the Illumina NovaSeq 6000 Sequencing System (Illumina Inc., USA).

Genome mapping and reference-based transcriptome assembly

All sequenced libraries were assessed for quality control using FastQC v.0.11.9 [77] and trimmed for Illumina adaptor sequences and low-quality base-calls using Trimmomatic v.0.38 [78]. The trimmed reads with high quality were then aligned to the P. taeda reference genome sequence (Pita_v2.01; Treegenes database [79]) using HISAT2 v.2.0.0 [80] with parameters “--known-splicesite-infile”, “--dta” and “--rna-strandness RF”. In order to ensure the presence of F. circinatum biomass in the samples, the reads were also mapped to its publicly available genome sequence (accession number JAG-GEA0000000000). The SAM files from the pine mapping were processed with the SAMtools utility [81] for converting to binary alignment map (BAM) format, sorting by coordinates and removing duplicates. The transcripts for each sample were reconstructed separately by StringTie v.2.1.4 [82] using the “-G option” with the annotation file of P. taeda (Pita.2_01.entap_annotations.tsv; Treegenes database [79]). This file was previously fixed with Gffread utility v.0.12.1 [83] for the correct understanding by StringTie program. After the transcriptome assembly, the nine resulting GTF files were merged to generate a non-redundant set of transcripts with unique identifiers using the StringTie “-merge” parameter, where only transcripts with expression levels > 0.1 FPKM were included. Finally, this newly experiment-level transcriptome was further compared with the P. taeda reference annotation GTF file (Pita_v2.01; Treegenes database [79]) using the software Gffcompare v.0.12.1 [83], classifying transcripts in different class codes according to their nature/origin.

LncRNAs identification

Based on all the assembled transcripts, the known transcripts marked with the class code “=” were excluded before conducting the potential long non-coding RNAs identification. The remaining transcripts were subjected to the coding potential predictor FEELnc v.0.2 tool [84] as well as several filters to ensure reliability of lncRNAs. Firstly, the FEELnc filter module was used to remove short transcripts (< 200 nt) and keep only monoexonic transcripts with antisense localization. After that, the sequences of the resulting transcripts were extracted with Gffread v.0.12.1 [83] and the fasta file output was piped to the Eukaryotic Non-Model Transcriptome Annotation Pipeline (EnTAP) v.0.9.2 [85] for transcript annotation. Briefly, GeneMarkS-T v.5.1 [86] was used for ORF prediction and the sequence aligner DIAMOND v.1.9.2 [87] conducted the similarity search with default settings (E-value < 10⁻⁵) using the NCBI non-redundant protein database (release-201). After that, the assignment of protein domains (Pfam), GO terms and KEGG pathways was performed using EggNOG v.1.0.3 [88]. Finally, EnTAP filtered contaminants to retain only high-quality transcripts. Subsequently, the FEELnc codpot module was used with the shuffling mode to calculate a coding potential score (CPS) for the un-annotated transcripts using a random forest algorithm trained with multi k-mer frequencies and relaxed ORFs. The specificity threshold was set at 0.95 in order to increase the robustness of the final set of novel lncRNAs. The remaining transcripts were designated as lncRNAs and further classified according to the ‘Gffcompare’ output as lincRNAs categorized with class code ‘u’, lncNAT from the class code ‘x’, and intronic transcripts that were those with class code ‘i’ [89].

In order to investigate the conservation of the pine lncRNAs, two recently released and updated databases of known plant lncRNAs were used [40]. All the transcripts designated as lncRNA were aligned against CANTATA database [90] and GreeNc database [91] using the blastn algorithm (E-value < 10⁻⁵) of the BLAST v.2.9.0 software suite [92, 93]. Moreover, the transcripts were also aligned to the Rfam (version 14.1) and miRBase (version 21) non-coding RNA databases with designated threshold value (E-value < 10⁻⁵) using the blastn algorithm in order to detect housekeeping non-coding RNAs including tRNAs, rRNAs and snoRNAs, and miRNA precursors.

Differential expression analysis

StringTie together with the “-e” parameter was employed to estimate expression for all transcripts of the experiment-level transcriptome [82]. The output file was reformatted using the “prepDE.py” script for further expression analysis [94]. DESeq2 v.1.24.1 [95] was used to identify DE lncRNA transcripts based on the matrix of the estimated counts. DEGs were identified equally. The pairwise comparison of inoculated and control plants were evaluated using Wald tests. To visualize the similarity of the replicates and identify any sample outliers, the PCA was constructed using the rlog-transformed expression values. Transcripts were considered as differentially expressed if the adjusted p-values (padj) for multiple testing, using Benjamini–Hochberg to estimate the false discovery rate (FDR) [96], was less than 0.05 and the |log2Fold Change| ≥ 1.

Potential target gene prediction and functional enrichment

Based on the genome location of the lncRNAs relative to the neighbouring genes, the nearest protein-coding genes
transcribed within a 10 kb window upstream or 100 kb downstream were considered as potential cis-regulated target genes. These genes were identified using the FEELnc classifier module [84] and annotated using the EnTAP pipeline [85] as described above but implemented with the RefSeq complete protein database (release-201) and the UniProtKB/Swissprot database (release-2020_05).

Functional enrichment analysis of the target genes associated with the DE IncRNAs was conducted. DE IncRNA transcripts were divided into up- and down-regulated subsets for efficient functional analysis [97]. Using all genes as background, GO and KEGG enrichment analysis were conducted by GOSeq v.1.38.0 based on the Wallenius non-central hyper-geometric distribution that allows the adjustment for transcript length bias [98]. The GO terms and KEGG pathways with corrected p-values lower than 0.05 were considered to be enriched in the group. Redundant gene ontology categories were parsed using Revigo [99].

Co-expression analysis and identification of hub genes
In order to predict the co-expression modules and determine the GO terms that differentiate the transcriptome induced by F. circinatum, a weighted gene co-expression network analysis approach implemented in the R-based Co-Expression Modules identification Tool (CEMiTool) package v.1.8.3 [100] was conducted in R software. Network analysis was carried out on the expression data for three gene sets: DE IncRNAs, DEGs and targeted genes predicted by FEELnc. A variance stabilizing transformation (VST) was used and transcripts were filtered to reduce correlation between variance and gene expression. The Spearman’s method was used for calculating the correlation coefficients and a soft thresholding power (β) of 6 was selected. The co-expressed modules were subjected to over-representation analysis (ORA) based on the hypergeometric test [101] using the GO terms to determine the most significant module functions (q-value <= 0.05) [90]. Moreover, genes with the highest connectivity, known as hub genes and considered functionally important genes [102] were identified in each module.

Abbreviations
ncRNA: Non-coding RNA; miRNAs: MicroRNAs; IncRNAs: Long non-coding RNA; NGS: Next generation sequencing; PPC: Pine Pitch Canker; RNA-Seq: RNA sequencing; dpi: Days post inoculation; polyA: Polyadenylated; lncRNAs: Long intergenic non-coding RNAs; IncNATs: Long non-coding natural antisense transcripts; rRNA: Ribosomal RNAs; tRNA: Transfer RNAs; snoRNAs: Small nuclear RNAs; bp: Base pair; FPKM: Fragment per kilobase of exon per million mapped reads; PCA: Principal component analysis; DE: Differentially expressed; kb: Kilobases; mRNA: Messenger RNA; RLKs: Receptor-like protein kinases; PMEs: Pectin methylesterases; 4CL3: 4-coumarate-CoA ligase 3; DEGs: Differentially expressed genes; ETR2: Ethylene receptor 2; PAR1: Photoassimilate-responsive protein 1; PDC1: Pyruvate decarboxylase 1; FC: Fold-change; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; GSEA: Gene set enrichment analysis; NES: Normalised Enrichment Score; BH: Benjamin-Hochberg; RBPs: RNA-binding proteins; PPR2: Pentatricopeptide repeat-containing protein; RRM: RNA recognition motif; ROS: Reactive Oxygen Species; GST: Glutathione S-transferases; NO: Nitric Oxide; PP2C: Type 2 C protein phosphatase; NERD: Needed for rdr2-independent DNA methylation; PMEI: Pectin methylesterase inhibitor; CCOMT: Caffeoyl-CoA 3-O-methyltransferase; ADH1: Alcohol dehydrogenase 1; PDA: Potato Dextrose Agar; PDB: Potato Dextrose Broth; TAE: Tris-Acetate-EDTA; RIN: RNA integrity number; BAM: Binary Alignment Map; EnTAP: Eukaryotic Non-Model Transcriptome Annotation Pipeline; CPS: Coding Potential Score; ORF: Open Reading Frame; p.adjust: Adjusted p-value; FDR: False Discovery Rate; VST: variance stabilizing transformation; ORA: Over-representation analysis.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12864-022-08408-9.

Additional files

- Additional file 1.
- Additional file 2.
- Additional file 3.
- Additional file 4.
- Additional file 5.
- Additional file 6.
- Additional file 7.
- Additional file 8.
- Additional file 9.
- Additional file 10.
- Additional file 11.
- Additional file 12.
- Additional file 13.
- Additional file 14.
- Additional file 15.
- Additional file 16.

Acknowledgements
The authors thank Marcos García for informatics support and Sergio Díez for helping in R scripting. In addition, we would like to thank Jonatan Niño for assisting in R scripting. In addition, we would like to thank Jonatan Niño for reviewing the text. Special thanks to Alex Douglas for hosting CZ-B at the University of Aberdeen (Scotland) and teaching her bioinformatics. The data analysis has been carried out using the resources of the “Centro de Supercomputación de Castilla y León” (SCAYLE) under the valuable technical support of Carmen Calvo and Jesús Lorenzana.

Authors’ contributions
CZ-B and JM-G performed the experimental work. CZ-B conducted the bioinformatics analysis and interpreted data analyses. AS-V supervised the bioinformatics analysis. JJD obtained funding to support the research. CZ-B wrote the manuscript with input from JM-G, AS-V and JJD, who supervised aspects of this research. The authors read and approved the final manuscript.

Funding
This work was supported by project PID2019-110459RB-I00 funded by MICINN (Spain) as well as the project VA208P20 funded by JCYL (Spain), both co-financed by FEDER (EU) budget. University of Valladolid supported CZ-B (Rector’s Resolution of November 11, 2016).

Availability of data and materials
Pine seeds and fungal isolate used in this study will be provided upon request in order to support reproducibility. Likewise, the datasets generated and
analysed during the current study are available in the SRA database (https://www.ncbi.nlm.nih.gov/sra) of NCBI repository. The SRA accession numbers are SRX15100123-31 (BioProject PRJNA742852; https://www.ncbi.nlm.nih.gov/search/all?term=PRJNA742852). The reference genome (Pita_v2.01) is publicly available in Treegenes database (https://treegenedb.org/FTP/Genomes/Pita_v2.01/genome/Pita_2_01.fa.gz).

Declarations

Ethics approval and consent to participate
Not applicable. Pine seedlings used in this study were sourced from the provenance O1, Galicia litoral provided by the Consellería do Medio Rural (Xunta de Galicia, Spain). The voucher specimen of *F. radiata* had been deposited to our lab but not to any publicly available herbarium. *Fusarium circinatum* isolate used in this study was obtained with the pertinent permission from the competent administration (Government of Cantabria), an institution that was also collaborating in the research study that was ongoing at the time. This isolate is maintained by our group.

Consent for publication
Not applicable.

Competitive interests
The authors declare that they have no conflicts of interest.

Author details
1 Department of Vegetal Production and Forest Resources, University of Valadolid, Av Madrid 44, 34004 Palencia, Spain. 2 Sustainable Forest Management Research Institute, University of Valladolid—INA, 34004 Palencia, Spain. 3 Departamento de Producción Animal, Facultad de Veterinaria, University of León, Campus de Vegazana s/n, 24071 León, Spain.

Received: 9 October 2021   Accepted: 14 February 2022

Published online: 10 March 2022

References

1. Eddy SR. Non-coding RNA genes and the modern RNA world. Nat. Rev. Genet. 2001;2:919–929. https://doi.org/10.1038/3503511.
2. Kapanov P, Cheng L, Dike S, Nix DA, Duttagupta R, Willingham AT, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science. 2007;316:1484–8. https://doi.org/10.1126/science.1138341.
3. Quan M, Chen J, Zhang D. Exploring the secrets of long noncoding RNAs. Int J Mol Sci. 2015;16:5467–96. https://doi.org/10.3390/ijms16054667.
4. Gil N, Ulltby I. Regulation of gene expression by c-acting long non-coding RNAs. Nat Rev Genet. 2020;21:102–17. https://doi.org/10.1038/s41576-019-0184-5.
5. Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. RNA Biol. 2013;10:924–33. https://doi.org/10.4161/mb.24604.
6. Wang KC, Chang HY. Molecular Mechanisms of Long Noncoding RNAs. Mol Cell. 2011;43:904–14. https://doi.org/10.1016/j.molcel.2011.08.018.
7. Liu X, Hao L, Li D, Zhu L, Hu S. Long Non-coding RNAs and Their Biological Roles in Plants. Genomics, Proteomics Bioinforma. 2015;13:137–47. https://doi.org/10.1016/j.gpb.2015.02.003.
8. Sanchita, Trivedi PK, Asif MH. Updates on plant long non-coding RNAs (lncRNAs): the regulatory components. Plant Cell Tissue Organ Cult. 2020;142:259–69. https://doi.org/10.1007/s11210-019-01726-z.
9. Chen J, Quan M, Zhang D. Genome-wide identification of novel long non-coding RNAs in *Populus tomentosa* tension wood, opposite wood and inducing PR gene expression. Phytopathology. 2020;110:873–80. https://doi.org/10.1094/Phyto-12-19-0044-R.
10. Qin T, Zhao H, Cui P, Albesher N, Xiaona L. A nucleus-localized long non-coding RNA enhances drought and salt stress tolerance. Plant Physiol. 2017;175:1321–36. https://doi.org/10.1104/pp.17.00574.
11. Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga M, Rubio-Somoza I, et al. Target mimicry provides a new mechanism for regulation of microRNA activity. Nat Genet. 2007;39:1033–7. https://doi.org/10.1038/ng2079.
12. 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*. Front Plant Sci. 2017;8:1035.
13. Seo J, Sun HX, Park BS, Huang CH, Yeh SD, Jung C, et al. ELF18-INDUCED LONG-NONCODING RNA associates with mediator to enhance expression of innate immune response genes in Arabidopsis. Plant Cell. 2017;29:1024–38. https://doi.org/10.1105/tpc.16.00886.
14. Hou X, Cui J, Liu W, Jiang N, Zhou X, Qi H, et al. LncRNA39026 enhances tomato resistance to *Phytophthora infestans* by decoying miR168a and inducing PI gene expression. Phytopathology. 2020;110:873–80. https://doi.org/10.1094/Phyto-12-19-0044-R.
15. Zhang L, Wang M, Li N, Wang H, Qi P, Pei L, et al. Long noncoding RNAs involved in resistance to *Verticillium dahliae*, a fungal disease in cotton. Plant Biotechnol J. 2018;16:1172–85. https://doi.org/10.1111/pbi.12861.
16. Yu Y, Zou YF, Feng YZ, He H, Lian JP, Yang YW, et al. Transcriptional landscape of pathogen-responsive IncRNAs in rice unveils the role of ALF1 in jasmonate pathway and disease resistance. Plant Biotechnol J. 2020;18:679–90. https://doi.org/10.1111/pbi.13234.
17. Tripathi R, Chakraborty P, Varadawaj PK. Unraveling long non-coding RNAs through analysis of high-throughput RNA-sequencing data. Non-coding RNA Res. 2017;2:111–8.
18. Joshi RK, Megha S, Basu U, Rahman MH, Kav NN V. Genome Wide Identification and Functional Prediction of Long Non-Coding RNAs Responsive to *Sclerotinia sclerotiorum* Infection in *Brassica napus*. PLoS One. 2016;11:e0158784. https://doi.org/10.1371/journal.pone.0158784.
19. Li W, Li C, Li S, Peng M. Long noncoding RNAs that respond to *Fusarium oxysporum* infection in “Cavendish” banana (*Musa acuminate*). Sci Rep. 2017;7:1–13. https://doi.org/10.1038/s41598-017-07197-3.
20. Zhang W, Zhai X, Cao Y, Dong Y, Fan G. Long non-coding RNAs responsive to *Witches’ Broom* disease in *Paulownia tomentosa*. Forests. 2017;8:348. https://doi.org/10.3390/f8090348.
21. Zhang W, Liu Y, Li L, Li D, Zhang Q, Guo Y, et al. Whole transcriptome sequencing of *Pseudomonas syringae pv. actinomycetemcomitans* infected kiwifruit plants reveals species-specific interaction between long non-coding RNA and coding genes. Sci Rep. 2017;7:1–15. https://doi.org/10.1038/s41598-017-05377-y.
22. Feng S, Fang H, Liu X, Dong Y, Wang Q, Yang QK. Genome-wide identification and characterization of long non-coding RNAs conferring resistance to *Colletotrichum gloeosporioides* in walnut (* Juglans regia*). BMC Genomics. 2021;22:15. https://doi.org/10.1186/s12864-020-07310-6.
23. Wingfield MJ, Hammerbacher A, Ganley RI, Steenkamp ET, Gordon TR, Wingfield RD, et al. Pitch canker caused by *Fusarium cinctum* - A growing threat to pine plantations and forests worldwide. Australas Plant Pathol. 2008;37:319–34. https://doi.org/10.1071/AP08036.
24. Drenkhans R, Ganley B, Martín-García J, Vahalik P, Adamson K, Adamčíková K, et al. Global Geographic Distribution and Host Range of *Fusarium cinerea*, the Causal Agent of Pine Pitch Canker. Forests. 2020;11. https://doi.org/10.3390/f11070724.
25. Zamora-Ballesteros C, Diez JJ, Martín-García J, Witzel J, Solla A, Ahumada R, et al. Pine Pitch Canker (PPC). Pathways of Pathogen Spread and Preventive Measures. Forests. 2019;10:1158. https://doi.org/10.3390/f10121158.
26. Martín-García J, Zas R, Solla A, Woodward S, Hantula J, Vainio EJ, et al. Environmentally-friendly methods for controlling pine pitch canker. Plant Pathol. 2019;68:843–860. https://doi.org/10.1111/ppa.13009.
27. Gordon TR, Swett CL, Wingfield MJ. Management of *Fusarium* diseases affecting conifers. Crop Prot. 2015;73:28–39. https://doi.org/10.1016/j.cropro.2015.02.018.
28. Visser EA, Wegzyn JL, Steenkamp ET, Myburg AA, Naidoo S. Combined de novo and genome guided assembly and annotation of the *Pinus patula* juvenile shoot transcriptome. BMC Genomics. 2015;16:1037. https://doi.org/10.1186/s12864-015-1227-7.
29. Carrasco A, Wegzyn JL, Durán R, Fernández M, Donoso A, Rodriguez V, et al. Expression profiling in *Pinus radiata* infected with *Fusarium*
33. Visser EA, Wegrzyn JL, Steenkamp ET, Myburg AA, Naidoo S. Dual RNA-seq analysis of the pine-Fusarium circinatum interaction in resistant (Pinus taeda) and susceptible (Pinus palustris) hosts. Microorganisms. 2019;7:7–9. https://doi.org/10.3390/microorganisms7090315.

34. Hernández-Escribano L, Visser EA, Iturritxa E, Raposo R, Naidoo S. Temporal physiological response of pine to Fusarium circinatum infection is dependent on host susceptibility level: the role of ABA catabolism. Tree Physiol. 2020;paa143. https://doi.org/10.1093/treephys/paa143.

35. Zamora-Ballesteros C, Alves A, Berenger H, Correia B, Gómez-Cadenas A, Diez JJ, et al. Phosphite shifts physiological and hormonal profile of Monterey pine and delays Fusarium circinatum progress. Plant Physiol Biochem. 2017;114:88–99. https://doi.org/10.1016/j.plaphy.2017.02.020.

36. Yang Y, Luo X, Sun F, Hu J, Zha X, Su W, et al. Overexpressing IncRNA LAIR increases grain yield and regulates neighbouring gene cluster expression in rice. Nat Commun. 2018;9:1–9. https://doi.org/10.1038/s41467-018-05829-7.

37. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, Geisinger K, et al. GREAT: Gene Regulator Analysis tool. Nucleic Acids Res. 2014;42:D859–65. https://doi.org/10.1093/nar/gkt1190.

38. Coimbra S, Massi M, Amado S, Almeida J, Alves A, Berenger H, et al. Genome-wide identification and evolution of plant LAIR and structural class I hemoglobin gene hb1 reduces cadmium levels by modulating cd transporter expression through decreasing nitric oxide and ROS level in Arabidopsis. Front Plant Sci. 2019;10:201. https://doi.org/10.3389/fpls.2019.00201.

39. Fan G, Cao Y, Wang Z. Regulation of long noncoding RNAs responsive to phytophthora infection in Paulownia tomentosa. Int J Genomics. 2018. https://doi.org/10.1155/2018/374352.

40. Dai M, Alam M, Lightfoot DA, Guha P, Aiazzi B. Classification and experimental identification of plant long non-coding RNAs. Genomics. 2019;111:997–1005. https://doi.org/10.1016/j.ygeno.2018.04.014.

41. Lee K, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet. 2010;11:733–9. https://doi.org/10.1038/nrg2825.

42. Cao J, Jiang M, Li P, Chu Z. Genome-wide identification and evolutionary analyses of the PP2C gene family with their expression profiling in response to multiple stresses in Brachypodium distachyon. BMC Genomics. 2016;17:175. https://doi.org/10.1186/s12859-016-1256-4.

43. Cabili M, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals 20,770 processed IncRNAs. Nat Biotechnol. 2011;29:1665–70. https://doi.org/10.1038/nbt.2139.

44. Wang M, Yuan D, Tu L, Gao W, He Y, Hu H, et al. Long non-coding RNAs responsive to Fusarium oxysporum infection in Arabidopsis thaliana. New Phytol. 2014;201:574–84. https://doi.org/10.1111/nph.12537.

45. Shi W, Quan M, Du Q, Zhang D. The Interactions between the Long Non-coding RNAs and the defense responses in Arabidopsis thaliana. Front Plant Sci. 2017;8:1035. https://doi.org/10.3389/fpls.2017.01035.

46. Zhao Q, Guo HW. Paradigms and paradox in the ethylene signaling pathway and interaction network. Mol Plant. 2011;4:626–34. https://doi.org/10.1093/mp/ssr042.

47. Wang M, Yuan D, Tu L, Gao W, He Y, Hu H, et al. Long non-coding RNAs and their proposed functions in fibre development of cotton (Gossypium spp.). New Phytol. 2015;207:1181–97. https://doi.org/10.1111/nph.13429.

48. Tian J, Song Y, Yang X, Cao D, Chen J, Xie J, et al. Population genomic analysis of gibberellin-responsive long non-coding RNAs in Populus. Nat Genet. 2016;48:2467–82. https://doi.org/10.1038/ng.3607.

49. 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. Front Plant Sci. 2017;8:1035. https://doi.org/10.3389/fpls.2017.01035.

50. Rai MI, Alam M, Lightfoot DA, Gurha P, Afzal AJ. Classification and experimental identification of plant long non-coding RNAs. Genomics. 2019;111:997–1005. https://doi.org/10.1016/j.ygeno.2018.04.014.

51. Li X, Wu HK, Southerton SG. Comparative genomics reveals conservative evolution of the xylem transcriptome in vascular plants. BMC Evol Biol. 2010;10:1–14. https://doi.org/10.1186/1471-2148-10-190.

52. Cabili M, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 2011;25:1915–27. https://doi.org/10.1101/gad.17446511.

53. Zaynab M, Fatima M, Abbas S, Sharif Y, Umar M, Zafar MH, et al. Role of secondary metabolites in plant defense against pathogens. Microb Pathog. 2018;124:198–202. https://doi.org/10.1016/j.micpat.2018.08.034.

54. Shuai P, Liang D, Tang S, Zhang Z, Ye CY, Su Y, et al. Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in Populus trichocarpa. J Exp Bot. 2014;65:4975–83. https://doi.org/10.1093/jxb/eru256.

55. Kovalchuk A, Kerio S, Oghenekaro AO, Jabier E, Raffaello T, Asiegbo FO. Antimicrobial Defenses and Resistance in Forest Trees: Challenges and Perspectives in a Genomic Era. Artic Annu Rev Phytopathol. 2013;51:221–44. https://doi.org/10.1146/annurev-phyto-080712-102307.

56. Zhao H, Stephen S, Taylor J, Hellwell CW, Wang M-B. Long noncoding RNAs responsive to Fusarium oxysporum infection in Arabidopsis thaliana. New Phytol. 2014;201:574–84. https://doi.org/10.1111/nph.12537.
71. Martín-García J, Paraschiv M, Flores-Pacheco JA, Chira D, Diez JJ, Fernández-Correll JC, Gordon TR, McCain AH, Fox JW, Koehler CS, Wood DL, et al.

72. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. J Am Stat Assoc. 1958;53:481. https://doi.org/10.1080/01621459.1958.10501482.

73. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. J Am Stat Assoc. 1958;53:481. https://doi.org/10.1080/01621459.1958.10501482.

74. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. J Am Stat Assoc. 1958;53:481. https://doi.org/10.1080/01621459.1958.10501482.

75. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. J Am Stat Assoc. 1958;53:481. https://doi.org/10.1080/01621459.1958.10501482.

76. Martín-Rodrigues N, Espinel S, Sanchez-Zabala J, Ortíz A, González-Martín E, Player J, et al.

77. Andrews S. FastQC a quality control tool for high throughput sequence data. Nature Methods. 2010;7:9. https://doi.org/10.1038/nmeth0910-901.

78. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

79. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–5. https://doi.org/10.1038/nbt.3137.

80. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

81. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–5. https://doi.org/10.1038/nbt.3137.

82. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–5. https://doi.org/10.1038/nbt.3137.

83. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–5. https://doi.org/10.1038/nbt.3137.

84. Wucher V, Legeai F, Hédan B, Rizk G, Lagoutte L, Leeb T, et al. FEELnc: A fast and sensitive tool for long non-coding RNA annotation and its application to the dog genome. Nucleic Acids Res. 2019;47:e114–22. https://doi.org/10.1093/nar/gky1141.

85. Hart AJ, Ginzburg S, Xu M, Sam, Fisher CR, Rahmatpour N, Mitton JB, et al. EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 2019;47:e114–22. https://doi.org/10.1093/nar/gky1141.

86. Tang S, Lomsadze A, Borodovsky M, Segal E. Integrative meta-analysis of transcriptomic responses to abiotic stress in crop. Proc Biol Sci. 2019;17(461):1–12. https://doi.org/10.1098/rspb.2018.2053.

87. Wang L, Courcelle E, Plagnol V, Speed TP. A method for generating reference sequences for eukaryotic genomes. Genome Biol. 2009;10:R137. https://doi.org/10.1186/gb-2009-10-3-r137.

88. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–5. https://doi.org/10.1038/nbt.3137.

89. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple J R Stat Soc Ser B. 1995;57:289–300.

90. Russo PST, Ferreira GR, Cardozo LE, Bürger MC, Arias-Carrasco R, Maruyama SR, et al. EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 2019;47:e114–22. https://doi.org/10.1093/nar/gky1141.

91. Gallant AP, Pulido AH, De Lagrán IAM, Sanseverino W, Cigliano RA. GREENC: A wiki-based database of plant IncRNAs. Nucleic Acids Res. 2016;44:D1161–6. https://doi.org/10.1093/nar/gkv1215.

92. Kalvari I, Nawrocki EP, Onditeros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, et al. Pfam 14: Expanded coverage of metagenomic, viral and microRNA families. Nucleic Acids Res. 2021;49:D1192–200. https://doi.org/10.1093/nar/gkaa1047.

93. Kozomara A, Birchacan M, Griffiths-Jones S. miRBase: From microRNA sequences to function. Nucleic Acids Res. 2019;47:D155–62. https://doi.org/10.1093/nar/gky1141.

94. CCB, StringTie. Transcript assembly and quantification for RNA-Seq. 2019. http://ccb.jhu.edu/software/stringtie/index.shtml?manual. Accessed 7 Jul 2021.

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

96. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple J R Stat Soc Ser B. 1995;57:289–300.

97. Hong G, Zhang W, Li H, Shen X, Guo Z. Separate enrichment analysis of pathways for up- and downregulated genes. J R Soc Interface. 2014;11:20130950. https://doi.org/10.1098/rsif.2013.0950.

98. Young MD, Wakefield M, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11:14. https://doi.org/10.1186/gb-2010-11-2-r14.

99. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLoS One. 2011;6:e21800. https://doi.org/10.1371/journal.pone.0021800.

100. Russo PST, Ferreira GR, Cardozo LE, Bürger MC, Arias-Carrasco R, Maruyama SR, et al. EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 2019;47:e114–22. https://doi.org/10.1093/nar/gkaa1047.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.