

## Characterization the coding and non-coding RNA components in the transcriptome of invasion weed *Alternanthera philoxeroides*

Yongxing ZHU<sup>1</sup>, Xinchun JIANG<sup>1</sup>, Xiaowen HAN<sup>1</sup>, Shuo HAN<sup>1</sup>,  
Zhongyi CHEN<sup>1</sup>, Junliang YIN<sup>1,2\*</sup>, Yiqing LIU<sup>1\*</sup>

<sup>1</sup>Hubei Key Laboratory of Waterlogging Disaster and Agricultural Use of Wetland, Yangtze University, Jingzhou 434000, Hubei, China; [yongxingzhu@yangtzeu.edu.cn](mailto:yongxingzhu@yangtzeu.edu.cn); [jxc417565835@163.com](mailto:jxc417565835@163.com); [1789878206@qq.com](mailto:1789878206@qq.com); [315946295@qq.com](mailto:315946295@qq.com); [czy@yangtzeu.edu.cn](mailto:czy@yangtzeu.edu.cn); [yinjunliang@yangtzeu.edu.cn](mailto:yinjunliang@yangtzeu.edu.cn) (\*corresponding author); [liung906@163.com](mailto:liung906@163.com) (\*corresponding author)

<sup>2</sup>Ministry of Agriculture Key Laboratory of Integrated Pest Management in Crops in Central China, Hubei Academy of Agricultural Sciences, 430064, Wuhan, China

### Abstract

*Alternanthera philoxeroides* is a notorious invasive weed worldwide, but it still lacks a genome information currently. In this study, we collected 4 groups of *A. philoxeroides* Illumina RNA-seq data (62.5 Gb) and performed a comprehensive *de novo* assembling. Totally, 421,372 unigenes were obtained with a total length of 230,842,460 bp, with 43,430 (10.31%) unigenes longer than 1000 bp. Then 119,222 (28.3%) unigenes were functional annotated and 235,885 (56.0%) were grouped into reliable lncRNAs reservoir. Besides, 534 tRNA and 234 rRNAs were identified in assembly sequences. Additionally, 131,624 microsatellites were characterized in 106,761 sequences. Then SSR primers were developed for the amplification of 40,752 microsatellites in 36,329 sequences. The miRNAs are key post-transcriptional regulators, about 86 candidate miRNA sequences were detected from *A. philoxeroides* assembly, and miRNA target genes prediction revealed possible functions of them in growth and development as well as stress responding processes. These results provide a vital basis for sequence-based studies of *A. philoxeroides* in the future, especially gene function analysis.

**Keywords:** alligator weed; invasive plant; reference unigenes; expression pattern profiling; function annotation; SSR marker

### Introduction

*Alternanthera philoxeroides* (Mart.) Griseb (*A. philoxeroides*), also known as alligator weed, belongs to family Amaranthaceae. It is a noxious semi-aquatic invasive weed native to South America and now is invasive in many parts of the world such as India, China, and Australia (Bai *et al.*, 2017). This invasion weed displays strong adaptive ability to a variety of habitats (e.g. aquatic, terrestrial habitats, farm lands) and fluctuating environments (Liu *et al.*, 2019). It inflicts serious damage to biodiversity and agricultural production to its invasion habitats (Wang *et al.*, 2018). In China, *A. philoxeroides* caused about 98 million US dollars lost annually, and is on the first shortlist of the invasive species that requires special control (Liu *et al.*, 2019).

Received: 27 Jan 2021. Received in revised form: 22 Feb 2021. Accepted: 26 Feb 2021. Published online: 01 Mar 2021.

From Volume 49, Issue 1, 2021, Notulae Botanicae Horti Agrobotanici Cluj-Napoca journal will use article numbers in place of the traditional method of continuous pagination through the volume. The journal will continue to appear quarterly, as before, with four annual numbers.

High-throughput sequencing has been widely used to identify transcripts involved in specific biological processes in plants, which is particularly useful for those species without whole genome sequencing (Li *et al.*, 2015; Zhu *et al.*, 2019a; Zhu *et al.*, 2019b). In *A. philoxeroides*, RNA-seq has been carried out in several studies to reveal its adaptations mechanisms to different habitats. For instance, Li *et al.* (2015) compared the gene expression profiles of *A. philoxeroides* under control and low potassium stress and discovered a large number of root-specific transcripts in response to low potassium stress. Gao *et al.* (2015) characterized transcript abundance variation in *A. philoxeroides* in contrasting hydrological conditions using RNA-seq, which improved our understanding of the biological processes underlying the phenotypic changes in this species. Liu *et al.* (2019) examined the differences in cold-responsive gene expression in *A. philoxeroides* from the northern edge and the central portion of its range, and suggested that genetic changes have accompanied the recent northward expansion of *A. philoxeroides* in China. However, since different RNA-seq projects were designed to solve specific problems, a single RNA-seq project may not enough to excavate as many transcripts as we expect, especially for *A. philoxeroides*, a species that lacks a genome information. Thus, combining different RNA-seq data together to perform *de novo* assembling is one robust strategy to obtain a more comprehensive transcripts reservoir (Iyer *et al.*, 2015).

Previous studies revealed that *A. philoxeroides* showed strong phenotypic plasticity with extremely low genetic diversity in China (Wang *et al.*, 2005; Geng *et al.*, 2019). Thus, to provide a more comprehensive transcriptome information of *A. philoxeroides*, we downloaded all accessible RNA-seq data of this species from NCBI and used them as input to carried out the *de novo* assembling with Trinity. Then expression profiling and function annotation were performed to obtain unigenes (the clustered contigs designated during the Trinity assembly). This study will provide useful gene resource for further study in elucidating the environmental acclimation mechanism and invasion management of *A. philoxeroides*.

## Materials and Methods

### *Illumina read processing and transcriptome assembly*

Through searching the SRA category of NCBI database by keyword “*Alternanthera philoxeroides*” and filtered by “RNA”, Illumina high throughput RNA-seq data belonging to 4 BioProjects were found and then downloaded to local disk. The accession number of these four data sets were PRJNA256237, PRJNA256235, PRJNA268359, and PRJNA263573 (detailed information can be found in Supplementary Data S1). The obtained SRA format data were transformed into fastq format by sratoolkit subroutine fastq-dump using command and parameters as “fastq-dump --gzip --offset 33 --defline-seq '@\$sn[\$rn]/\$ri' --defline-qual '+' --split-files file.sra”. Reads were trimmed by Trimmomatic to filter out the low-quality bases and reads using command and parameters as “trimmomatic PE -threads 40 -trimlog logfile -phred33 input\_forward.fq.gz input\_reverse.fq.gz output\_forward\_paired.fq.gz output\_forward\_unpaired.fq.gz output\_reverse\_paired.fq.gz output\_reverse\_unpaired.fq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36”. High quality reads obtained were input into Trinity (v2.0.6) to carry out the *de novo* assembling using command and parameters as “Trinity --seqType fq --max\_memory 200G --left reads\_1.fq --right reads\_2.fq --CPU 40”, and the obtained Trinity.fasta file was annotated by TransDecoder (v5.3.0), specifically subroutine TransDecoder.LongOrfs for ORF prediction (Command: “TransDecoder.LongOrfs -t target\_transcripts.fasta”) and TransDecoder.Predict for coded protein prediction (Command: “TransDecoder.Predict -t target\_transcripts.fasta”) (Haas *et al.*, 2013). Transcript expression level calculation was conducted by RSEM (v1.2.19) and default parameters were used for calculating (Li *et al.*, 2011).

### *Function annotation*

For unigenes annotation, the public databases including eggNOG (a database of orthologous groups and functional annotation), GO (Gene Ontology), KEGG (Kyoto Encyclopedia of Genes and Genomes),

KOG/COG (Clusters of Orthologous Groups of Proteins), Nr (NCBI non-redundant protein sequences), Pfam (Protein family), and Swiss-Prot (a manually annotated and reviewed protein sequence database, <https://www.uniprot.org/>) were used to perform gene function annotation, with a cut-off Expect value (E value) of  $10^{-10}$  (Yin *et al.*, 2018a; Zhu *et al.*, 2019c).

#### Non-coding RNAs annotation

tRNAscan-SE (v2.0) was used to identify tRNA and tRNAscan\_to\_GFF3.pl was used to produce GFF3 formatted annotation file (Lowe *et al.*, 1997). The barrnap (v0.9) was used to identify rRNA. CNCI (Sun *et al.*, 2013) and CPC ([http://cpc.gao-lab.org/programs/run\\_cpc.jsp](http://cpc.gao-lab.org/programs/run_cpc.jsp)) (Kong *et al.*, 2007) were used to assess the coding potential. The known plants mature miRNAs were downloaded from miRbase (v22.1, <http://www.mirbase.org/>) and queried to *de novo* assembly sequences using BLASTn (key parameters: -evalue 0.1 -word\_size 7) (Kozomara *et al.*, 20108). Then, hit regions were extracted by bedtools (v2.26.0) and further validated by ShortStack (v3.8.5) (Axtell *et al.*, 2013). Only N15 (passed all tests except that the miRNA-star was not sequenced) and Y (passed all tests) miRNA were kept as possible candidates.

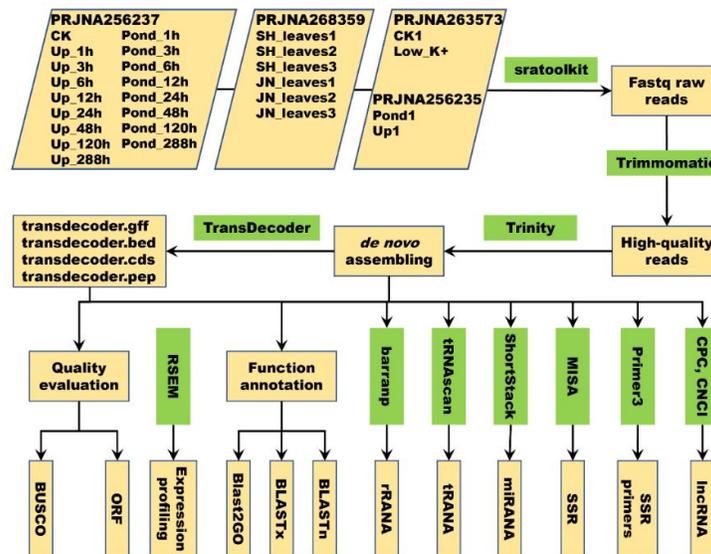
#### SSR marker development

MISA (v1.0) was used for microsatellites identification and localization (key parameters: definition: 1-10 2-6 3-5 4-5 5-5 6-5; interruptions: 100), p3\_in.pl was used to create the Primer3 input file, Primer3 was used to design SSR PCR primers from microsatellites DNA sequence, and p3\_out.pl was used to interpret the Primer3 output file (Beier *et al.*, 2017; Untergasser *et al.*, 2012).

## Results and Discussion

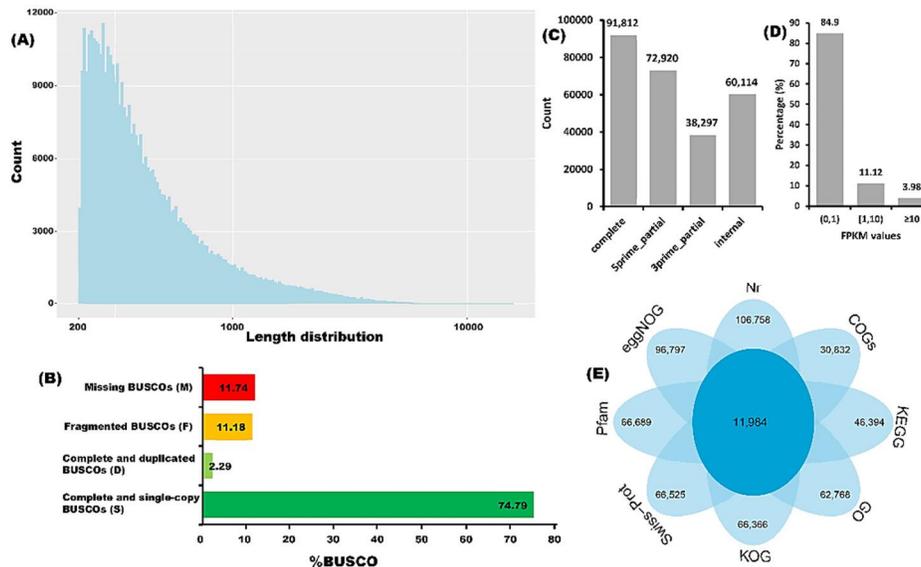
#### De novo assembling

To provide a comprehensive transcriptome data of *A. philoxeroides*, we collected all useful high throughput transcriptome sequencing data deposited in NCBI, and conducted the *de novo* assembling (Yin *et al.*, 2020). The overall process of reads processing, transcriptome assembling, expression profiling, quality evaluation, coding RNA function annotation and non-coding RNA detection, is outlined in Figure 1.



**Figure 1.** The workflow of *de novo* assembling and non-coding RNA annotation process. Illumina data belonging to four projects were used and their detailed information can be found in File S1.

Totally, 62.5 Gb data from four projects were used as input to perform the *de novo* assembling. Detailed information of these data can be found in File S1. As a result, 421,372 unigenes were obtained with a total length of 230,842,460 bp, an average length of 548 bp, and an N50 length of 621 bp. Specifically, 43,430 (10.31%) of 421,372 unigenes were longer than 1000 bp (Figure 2A, File S2). Similarly, in the RNA-seq study of Liu *et al.* (2019), a large number of unigenes (745,348) were detected and this may due to species specificity. To assess the assembling quality, Benchmarking Universal Single-Copy Orthologs (BUSCO) was used to estimate the completeness of the assembly sequences and results showed that the ratio of “Complete BUSCOs (C)” for assembled unigenes was 77.08% (Figure 2B). We then used assembly sequences as query to perform blastn against the previous assembly sequences by Liu *et al.* (2019). The result showed that 316,150 sequences assembled (75% of 421,372) in this study hit 723,070 of Liu *et al.* (2019) (97% of 745,348), which means most of the sequences reported by Liu *et al.* (2019) were assembled in this study. To be specific, 25% new sequences were assembled in the present study. Liu *et al.* (2019) used CEGMA to assess the completeness of their *de novo* assembly sequences and results showed that “complete” ratio was 92.7%, “missing” ratio was 2.7%, and “fragmented” ratio was 4.6%. In this research, BUSCO, the most popular and widely used assessment software for now, was used to assess the completeness of the *de novo* assembly results. For comparison, we further used BUSCO to assess the completeness of assembly sequences produced by Liu *et al.* (2019). As a result, the ratio of “Complete BUSCOs (C)” of assembly sequences produced by Liu *et al.* (2019) (85.4%) is slightly higher than ours (77.1%). However, most of sequences in the study of Liu *et al.* (2019) seem to be “duplicated” (75.1%), which is much higher than ours (2.3%). These results implied that the *de novo* assembly of this study contained more unique sequences with low duplication ratio, suggesting assembly sequences in this study are favourable and applicable for subsequent studies to functional characterize potential candidate genes involved in various metabolic pathways.



**Figure 2.** Statistic and evaluation of *de novo* assembly

**(A)** Length distribution of *de novo* assembling unigenes. **(B)** Completeness estimate result of BUSCO. **(C)** Results of the open reading frame (ORF) prediction analysis. The “complete” represents ORFs with 5prime start codon and 3prime stop codon; “5prime\_partial” represents ORFs with 5prime start codon but without 3prime stop codon; “3prime\_partial” represents ORFs with 3prime stop codon but without 5prime start codon; and “internal” represents ORFs without both 5prime start codon and 3prime stop codon. The detailed information can be found in File S3 to S5. **(D)** The expression levels distribution of assembled unigenes. Detailed information of samples and corresponding expression profiles can be found in File S6. **(E)** Venn diagram showing the annotation results of *A. philoxeroides* transcriptome against seven databases. Detailed annotation information can be found in File S7.

Then the TransDecoder was used to perform the gene structure annotation and the detailed information of output files including “transdecoder.gff3”, “transdecoder.cds”, and “transdecoder.pep” were listed in Supplementary Files 3 to 5. Totally, 263,143 ORFs were recognized in 155,422 unigenes (File S3). Among them, we detected 91,812 complete length ORFs, 72,920 5prime\_partial length, 38,297 3prime\_partial length, and 60,114 internal length with both 5prime\_partial and 3prime\_partial (Figure 2C). It seems that there is a significant bias between incomplete 5'UTR and 3'UTRs. The bias was also reported in other studies. For example, in the leaf of *Cucurbita pepo* L. infested by *Aphis Gossypii*, 25,000 and 8,220 truncated 5'UTR and 3'UTRs were detected, respectively (Vitiello *et al.*, 2018). We speculated that the reasons for this significant bias between incomplete 5'UTR and 3'UTRs may due to: (1) software (e.g. Trinity and TransDecoder) induced bias; (2) Illumina sequencing strategy induced bias, since left reads quality are generally better than right reads when pair-end sequencing; (3) and reasons we still don't know.

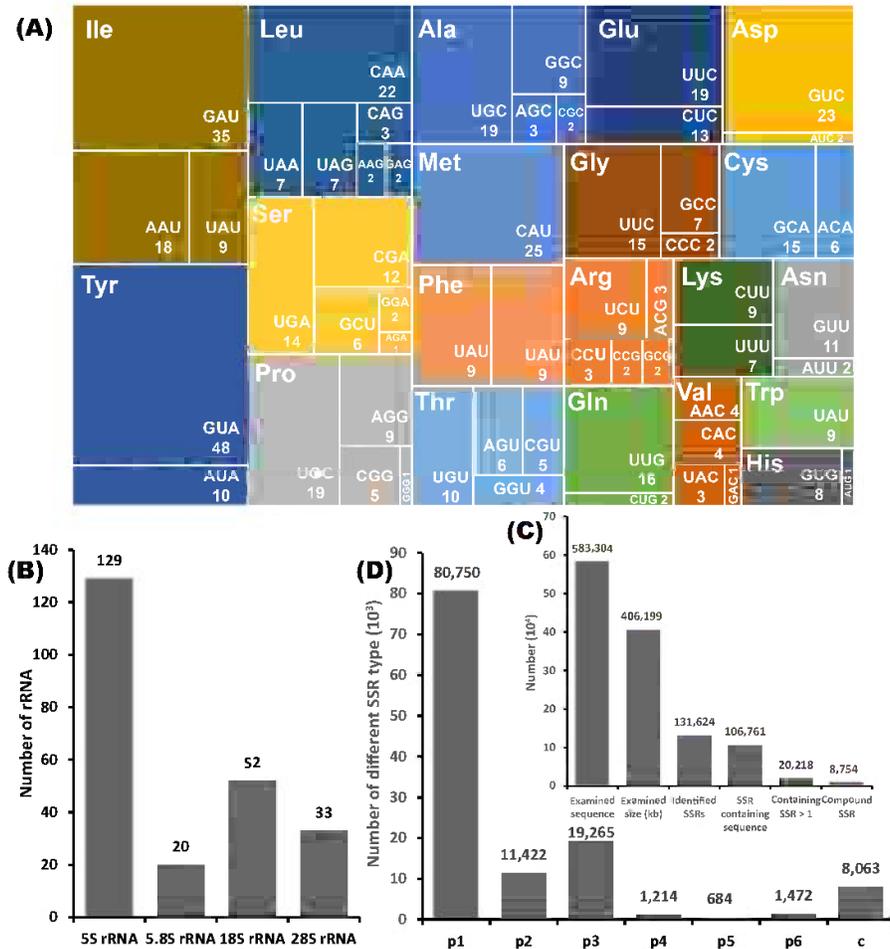
#### *Expression profiling and function annotation*

Expression levels of unigenes represented by FPKM values were calculated by RSEM and the details were listed in File S6. Then expression pattern profiling was performed and results showed that the expression levels of most unigenes were generally low (84.90% unigenes, FPKM Mean value < 1), while 11.12% unigenes showed medium expression level ( $1 \leq$  FPKM Mean value < 10), and 3.98% unigenes showed high expression level (FPKM Mean value  $\geq$  10) (Figure 2D).

Functional annotation and classification of the assembled transcripts provide insights into the diverse molecular and biological functions of represented genes (Xia *et al.*, 2011; Zhou *et al.*, 2018). We then performed BLAST to functional annotate the unigene sequences with the public annotation databases (Yin *et al.*, 2018b). Totally, 119,222 (28.3%) unigenes with annotated information were obtained, of which 106,758 (89.5%), 96,797 (81.2%), 66,689 (55.9%), 66,525 (55.8%), 66,366 (55.7%), 62,768 (52.6%), 46,394 (38.9%), and 30,832 (25.9%) were aligned to known proteins in Nr, eggNOG, Pfam, Swiss-Prot, KOG, GO, KEGG, and COGs databases, respectively (Figure 2E, File S7). The ratio of unigenes being functional annotated is acceptable since non-protein-coding RNAs (ncRNAs) constitute a substantial portion of transcribed sequences. For example, in human, over 68% of genes were classified as lncRNAs (long non coding RNAs) (Iyer *et al.*, 2015). And in *Arabidopsis* at least 50% of transcripts were lncRNAs and more lncRNAs will be detected with the development of the next-generation sequencing technologies and transcriptomic analysis approaches (Liu *et al.*, 2012; Yuan *et al.*, 2016; Severing *et al.*, 2018).

#### *The tRNA and rRNA annotation*

Despite the essential role of transfer RNA (tRNA) genes in protein translation in all living cells, little is known about the composition and amount of tRNAs in *A. philoxeroides* (Chan *et al.*, 2008). Here, tRNAscan-SE were used to scan the tRNA and 534 tRNA sequences were detected, including the tRNAs of Ala (33), Gly (24), Pro (34), Thr (25), Val (12), Ser (35), Arg (19), Leu (43), Phe (25), Asn (13), Lys (16), Asp (25), Glu (26), His (9), Gln (18), Ile (62), Met (26), Tyr (58), Cys (21), and Trp (11) (Figure 3A, File S8). The ribosomal RNA (rRNA) is another type of fundamental components in cells. Since rRNA genes are highly conserved sequences that are widespread across all genomes, they have extensive application values in the field of bioinformatics. For example, they can be used as the standard phylogenetic markers to characterize living species diversity and evolution (Karst *et al.*, 2018). Therefore, the barnnap was used to detect rRNA sequences and 234 rRNAs were detected, including 129 sequences of 5S rRNA, 20 of 5.8S rRNA, 52 of 18S rRNA, and 33 of 28S rRNA (Figure 3B, File S9).



**Figure 3.** The tRNA and rRNA categories and microsatellite containing sequences (A) The 20 amino acid and corresponding codons in *A. philoxeroides*. (B) The number of four type rRNAs. (C) Statistics of microsatellite (SSR) containing sequences. (D) The number of different types of SSRs in *A. philoxeroides*.

#### Microsatellites identification and SSR primers design

Microsatellite markers (SSR) derived from plant genomic sequences have been extensively used over the past decades due to their characteristic of co-dominant, locus-specific, and high polymorphism levels (Fang *et al.*, 2019; Zhang *et al.*, 2020). Transcriptomic sequences can provide a rich source for genetic markers development, such as simple sequence repeats (SSRs) in microsatellites and single nucleotide polymorphisms (SNPs), especially for plant species without reference genome (Vašek *et al.*, 2020). Previously, inter-simple sequence repeats (ISSR) markers were employed to interpret the genetic diversity of this invasive weed. However, due to the limited number of markers, the resolution of genetic diversity in natural populations was extremely low (Geng *et al.*, 2016). Here, we used MIRA to detect microsatellite sequences from *A. philoxeroides* and 131,624 microsatellites in 106,761 sequences were detected, among which 8,754 were compound formation microsatellites. To be specific, 20,218 sequences contain more than one microsatellite and each 3.1 kb of analysed sequence contains one SSR marker (Figure 3C, File S10). The most prevalent microsatellites were mononucleotide repeat pattern (80,750), followed by trinucleotide (19,265), dinucleotide (11,422), and compound (8,063) patterns. The most frequently observed microsatellite repeat was A/T, followed by G and AG/TC (Figure 3D, File S10). Then 36,447 sequences were chosen to design SSR amplification primer using Primer3, which resulted in each of 40,752 microsatellites in 36,329 sequences been covered by four pair of PCR primers (File S11). Successful control of *A. philoxeroides* depends on the fully

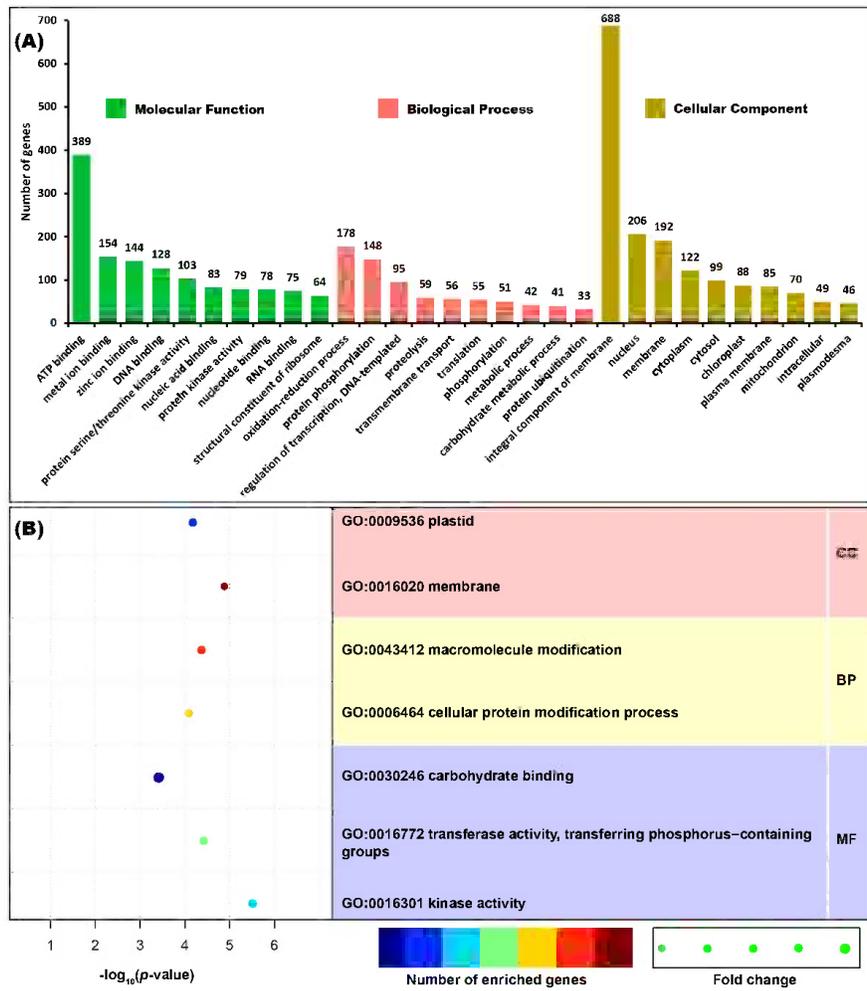
understanding of its genetic basis (Li and Ye, 2006). These large number of SSR markers developed in this study enable us to analyze population structure of *A. philoxeroides* at much higher resolution, which can accelerate the population genetic analysis of this invasion weed.

#### *The miRNA*

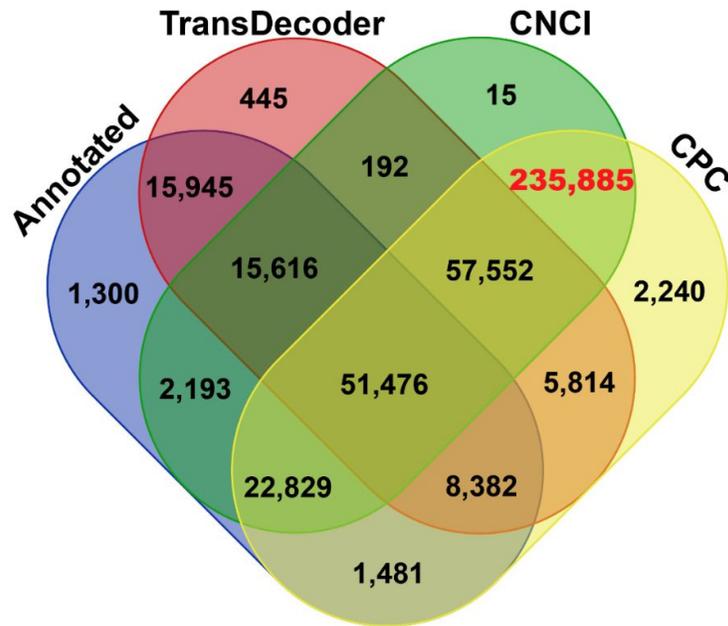
The miRNAs are key post-transcriptional regulators (Zhu *et al.*, 2019a). They are involved in a broad range of plants growth and development and stress responding processes through cleavage and/or inhibition translation of functional gene transcripts (Zhu *et al.*, 2019d). In this study, to explore the possible functions of *A. philoxeroides* miRNAs (Aph-miRNA), we detected miRNAs and predicted miRNA target genes. After BLASTn searching and ShortStack validation, about 86 candidate miRNA sequences were detected from *A. philoxeroides* assemblies (Files S12, S13). They were divided into 61 conserved miRNA families, including some well-studied members such as Aph-miR156 (regulation of flowering time), Aph-miR164 (response to water deficit), Aph-miR168 (influences phase transition, leaf epinasty, and fruit development), and Aph-miR394 (cold stress response) (Wang, 2014; Phookaew *et al.*, 2014; Xian *et al.*, 2014; Song *et al.*, 2016). Then, miRNA target genes were predicted by psRNATarget. Most of these target genes were transcripts coding transcription factors (File S14), such as squamosa promoter-binding protein (targets of miR156), NAC domain containing protein (miR164), nuclear transcription factors-YA4, -YA5, -YA6 (miR169), and GRAS family transcription factors (miR171). Besides transcription factors, miRNA targets including genes coding UDP-glucuronate decarboxylase 2 (miR164), AGO1-1 (miR168), protein phosphatase and kinase (miR390), F-box family protein (miR394), and proteins with unknown functions were predicted. Most conserved miRNA target genes, especially transcription factors (SBP, NAC, YA, GRAS), were found to be homologous to the conserved miRNA targets in Arabidopsis and other plants, which lay the foundation for further miRNA functional analysis in *A. philoxeroides* (Lakhotia *et al.*, 2014). Furthermore, gene ontology (GO) and enrichment analysis were performed by Blast2GO. Results demonstrated that within Cellular Component category most targets are genes coding membrane and nucleus components (Figure 4A). Within Molecular Function category, targets are mainly involved in ATP, metal ion, and DNA binding functions (Figure 4A). Within Biological Processes category, most targets participate in processes of oxidation-reduction, protein phosphorylation, and regulation of transcription (Figure 4A). Using GO annotations of all assembling unigenes as background, miRNA targets are enriched in cellular protein modification process (GO:0006464) and macromolecule modification (GO:0043412) under, kinase activity (GO:0016301), transferase activity (GO:0016772), carbohydrate binding (GO:0030246) under Molecular Function, and membrane (GO:0016020) and plastid (GO:0009536) under Cellular Component (Figure 4B).

#### *The lncRNA reservoir*

Considering the fact that large ratio of assembly sequences was not functionally annotated, we then used CNCI and CPC to confirm whether these sequences belong to lncRNA. Sequences that were identified as lncRNAs by both CNCI and CPC and did not contain TransDecoder interpreted protein coding regions were selected as reliable lncRNAs. Totally, 235,885 sequences (56% of 421,372 assembly sequences) were identified as lncRNAs (Figure 5, File S15). It is widely accepted that ncRNAs play important roles in transcriptional and/or post-transcriptional regulation (Liu *et al.*, 2012; Yuan *et al.*, 2016; Severing *et al.*, 2018). Accordingly, these large number of lncRNAs potentially involve in the regulation of physiological and biochemical processes of *A. philoxeroides*, and the function of which need to be deciphered in the future.



**Figure 4.** GO annotation and enrichment analysis of miRNA target genes (A) According to GO annotation, the putative targets of *A. philoxeroides* miRNAs are categorized into molecular function, biological process, and cellular component. (B) GO enrichment analysis of miRNA target genes. GO, Gene Ontology; CC, cellular component; MF, molecular function; BP, biological process.



**Figure 5.** The reservoir of lncRNAs in *A. philoxeroides* transcriptome

## Conclusions

In this study, we performed a comprehensive *de novo* assembling using Trinity based on 4 groups of released *A. philoxeroides* Illumina RNA-seq data (62.5 Gb). Totally, 421,372 unigenes were obtained. Completeness estimate showed the ratio of “Complete BUSCOs (C)” for assembled unigenes is 77.08%. The unigenes expression profiles were calculated by RSEM and their functions were annotated by public database including Nr, GO, KEGG, KOG/COG, Pfam, and Swiss-Prot. Further non-coding components analysis identified 534 tRNA and 234 rRNAs sequences. Microsatellites in transcriptome were scanned and detected 131,624 microsatellites in 106,761 sequences. Then 36,447 sequences were selected to design SSR primer, which resulted in each of 40,752 microsatellites in 36,329 sequences being covered by four pair of primers. Eighty-six candidate miRNA sequences from *A. philoxeroides* assembly were detected, and miRNA target genes prediction suggested their possible functions in plant growth and stress responding processes. Besides the 119,222 (28.3%) functional annotated unigenes, 235,885 (56.0%) were grouped into reliable lncRNAs reservoir. These results provide an important reference for sequence-based studies of *A. philoxeroides* and enriched resource for gene function investigations.

## Authors' Contributions

Conceptualization: JY, ZC and YL; Data curation: XH; Formal analysis: XH and SH; Funding acquisition: JY; Investigation: JY and XJ; Methodology: JY and XJ; Project administration: YZ and JY; Supervision: JY; Visualization: YZ; Writing - original draft: YZ; Writing - review and editing: ZY, JY and YL. All authors read and approved the final manuscript.

## Acknowledgements

This work was supported by the Open Project Program of Engineering Research Center of Ecology and Agricultural Use of Wetland, Ministry of Education (grant number KF201909); the Open Project Program of Ministry of Agriculture Key Laboratory of Integrated Pest Management in Crops in Central China (grant number 2019ZTSJJ8). We would like to thank the “Beijing Biomarker Technologies Co., Ltd” for assisting in the data analysis and “Shanghai Chenshan Botanical Garden” for sharing their RNA-seq data (PRJNA256235). We thank Prof. Yongli Qiao for beneficial comments on the initial project design and data analysis.

## Conflict of Interests

The authors declare that there are no conflicts of interest related to this article.

## References

- Axtell MJ (2013). Shortstack: Comprehensive annotation and quantification of small RNA genes. *RNA* 19:740-751. <https://doi.org/10.1261/rna.035279.112>
- Bai C, Wang P, Fan Q, Fu W, Wang L, Zhang Z, ... Wu J (2017). Analysis of the role of the drought-induced gene *DRI15* and salinity-induced gene *S11* in *Alternanthera philoxeroides* plasticity using a virus-based gene silencing tool. *Frontiers in Plant Science* 8:1579. <https://doi.org/10.3389/fpls.2017.01579>
- Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33:2583-2585. <https://doi.org/10.1093/bioinformatics/btx198>
- Chan PP, Lowe TM (2008). GtRNAdb: A database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Research* 37:D93-D97. <https://doi.org/10.1093/nar/gkn787>
- Fang Z, Sun C, Lu T, Xu Z, Huang W, Ma D, Yin J (2019). Molecular mapping of stripe rust resistance gene *Yrh922* in a derivative of wheat (*Triticum aestivum*)–*Psathyrostachys huashanica*. *Crop Pasture Science* 70:939-945. <https://doi.org/10.1071/CPI9317>
- Gao L, Geng Y, Yang H, Hu Y, Yang J (2015). Gene expression reaction norms unravel the molecular and cellular processes underpinning the plastic phenotypes of *Alternanthera philoxeroides* in contrasting hydrological conditions. *Frontiers in Plant Science* 6:991. <https://doi.org/10.3389/fpls.2015.00991>
- Geng Y, van Klinken RD, Sosa A, Li B, Chen J, Xu CY (2016). The relative importance of genetic diversity and phenotypic plasticity in determining invasion success of a clonal weed in the USA and China. *Frontiers in Plant Science* 7:213. <https://doi.org/10.3389/fpls.2016.00213>
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, ... Regev A (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8:1494-1512. <https://doi.org/10.1038/nprot.2013.084>
- Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, ... Chinnaiyan AM (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nature Genetics* 47(3):199. <https://doi.org/10.1038/ng.3192>
- Karst SM, Dueholm MS, McIlroy SJ, Kirkegaard RH, Nielsen PH, Albertsen M (2018). Retrieval of a million high-quality, full-length microbial 16s and 18s rRNA gene sequences without primer bias. *Nature Biotechnology* 36:190-195. <https://doi.org/10.1038/nbt.4045>
- Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G (2007). CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research* 35:W345-W349. <https://doi.org/10.1093/nar/gkm391>
- Kozomara A, Birgaoanu M, Griffiths-Jones S (2018). miRBase: from microRNA sequences to function. *Nucleic Acids Research*, 47, D155-D162. <https://doi.org/10.1093/nar/gky1141>
- Lakhotia N, Joshi G, Bhardwaj AR, Katiyar-Agarwal S, Agarwal M, Jagannath A, ... Kumar A (2014). Identification and characterization of miRNAome in root, stem, leaf and tuber developmental stages of potato (*Solanum*

- tuberosum* L.) by high-throughput sequencing. BMC Plant Biology, 14:6. <https://doi.org/10.1186/1471-2229-14-6>
- Li B, Dewey CN (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323-323. <https://doi.org/10.1186/1471-2105-12-323>
- Li J, Ye W (2006). Genetic diversity of alligator weed ecotypes is not the reason for their different responses to biological control. Aquatic Botany 85:155-158. <https://doi.org/10.1016/j.aquabot.2006.02.006>
- Li L, Xu L, Wang X, Pan G, Lu L (2015). *De novo* characterization of the alligator weed (*Alternanthera philoxeroides*) transcriptome illuminates gene expression under potassium deprivation. Journal of Genetics 94:95-104. <https://doi.org/10.1007/s12041-015-0493-1>
- Liu D, Horvath D, Li P, Liu W (2019). RNA sequencing characterizes transcriptomes differences in cold response between northern and southern *Alternanthera philoxeroides* and highlight adaptations associated with northward expansion. Frontiers in Plant Science 10:24. <https://doi.org/10.3389/fpls.2019.00024>
- Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenashuertero C, Chua NH (2012). Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. Plant Cell 24:4333-4345. <https://doi.org/10.1105/tpc.112.102855>
- Lowe TM, Eddy SR (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research 25:955-964. <https://doi.org/10.1093/nar/25.5.955>
- Phookaew P, Netpraphan S, Sojikul P, Narangajavana J (2014). Involvement of miR164- and miR167-mediated target gene expressions in responses to water deficit in cassava. Biologia Plantarum 58:469-478. <https://doi.org/10.1007/s10535-014-0410-0>
- Severing E, Faino L, Jange S, Busscher M, Kuijter-Zhang Y, Bellinazzo F, ... Pajoro A (2018). *Arabidopsis thaliana* ambient temperature responsive lncRNAs. BMC Plant Biology 18:145. <https://doi.org/10.1186/s12870-018-1362-x>
- Song JB, Gao S, Wang Y, Li BW, Zhang YL, Yang ZM (2016). miR394 and its target gene *LCR* are involved in cold stress response in *Arabidopsis*. Plant Gene 5:56-64. <https://doi.org/10.1016/j.plgene.2015.12.001>
- Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, ... Zhao Y (2013). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Research 41:e166. <https://doi.org/10.1093/nar/gkt646>
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012). Primer3-new capabilities and interfaces. Nucleic Acids Research 40:e115. <https://doi.org/10.1093/nar/gks596>
- Vašek J, Čilová D, Melounová M, Svoboda P, Vejl P, Štikarová R, ... Ovesná J (2020). New EST-SSR markers for individual genotyping of opium poppy cultivars (*Papaver somniferum* L.). Plants 9:10. <https://doi.org/10.3390/plants9010010>
- Vitiello A, Rao R, Corrado G, Chiaiese P, Digilio MC, Cigliano RA, D'Agostino N (2018). *De novo* transcriptome assembly of *Cucurbita pepo* L. leaf tissue infested by *Aphis gossypii*. Data 3(3):36. <https://doi.org/10.3390/data3030036>
- Wang B, Li W, Wang J (2005). Genetic diversity of *Alternanthera philoxeroides* in China. Aquatic Botany 81(3):277-283. <https://doi.org/10.1016/j.aquabot.2005.01.004>
- Wang J (2014). Regulation of flowering time by the miR156-mediated age pathway. Journal of Experimental Botany 65:4723-4730. <https://doi.org/10.1093/jxb/eru246>
- Wang T, Hu J, Wang R, Liu C, Yu D (2018). Tolerance and resistance facilitate the invasion success of *Alternanthera philoxeroides* in disturbed habitats: a reconsideration of the disturbance hypothesis in the light of phenotypic variation. Environmental and Experimental Botany 153:135-142. <https://doi.org/10.1016/j.envexpbot.2018.05.011>
- Xia Z, Xu H, Zhai J, Li D, Luo H, He C, Huang X (2011). RNA-seq analysis and *de novo* transcriptome assembly of *Hevea brasiliensis*. Plant Molecular Biology 77(3):299-308. <https://doi.org/10.1007/s11103-011-9811-z>
- Xian Z, Huang W, Yang Y, Tang N, Zhang C, Ren M, Li Z (2014). miR168 influences phase transition, leaf epinasty, and fruit development via *SLAGO1s* in tomato. Journal of Experimental Botany 65:6655-6666. <https://doi.org/10.1093/jxb/eru387>
- Yin J, Liu M, Ma D, Wu J, Li S, Zhu Y, Han B (2018a). Identification of circular RNAs and their targets during tomato fruit ripening. Postharvest Biology and Technology 136:90-98. <https://doi.org/10.1016/j.postharvbio.2017.10.013>

- Yin JL, Tian J, Li G, Zhu Y, Zhou X, He Y, ... Chen Z (2020). Carbohydrate, phytohormone, and associated transcriptome changes during storage root formation in *Alternanthera philoxeroides*. *Weed Science* 68:382-395. <https://doi.org/10.1017/wsc.2020.37>
- Yin J, Fang ZW, Sun C, Zhang P, Zhang X, Lu C, ... Zhu YX (2018b). Rapid identification of a stripe rust resistant gene in a space-induced wheat mutant using specific locus amplified fragment (SLAF) sequencing. *Scientific Reports* 8:3086. <https://doi.org/10.1038/s41598-018-21489-5>
- Yin JL, Wang LX, Zhao J, Li YT, Huang R, Jiang XC, ... Zhu YX (2020). Genome-wide characterization of the C<sub>2</sub>H<sub>2</sub> zinc-finger genes in *Cucumis sativus* and functional analyses of four CsZFPs in response to stresses. *BMC Plant Biology* 20: 359. <https://doi.org/10.1186/s12870-020-02575-1>
- Yuan J, Zhang Y, Dong J, Sun Y, Lim B, Liu D, Lu Z (2016). Systematic characterization of novel lncRNAs responding to phosphate starvation in *Arabidopsis thaliana*. *BMC Genomics* 17:655. <https://doi.org/10.1186/s12864-016-2929-2>
- Zhang X, Liu YH, Wang YH, Shen SK (2020). Genetic diversity and population structure of rhododendron rex subsp. Rex inferred from microsatellite markers and chloroplast DNA sequences. *Plants* 9:338. <https://doi.org/10.3390/plants9030338>
- Zhou R, Zhu Y, Zhao J, Fang Z, Wang S, Yin J, ... Ma D (2018). Transcriptome-wide identification and characterization of potato circular RNAs in response to *Pectobacterium carotovorum* subspecies *brasiliense* infection. *International Journal of Molecular Sciences* 19(1):71. <https://doi.org/10.3390/ijms19010071>
- Zhu Y, Jia JH, Yang L, Xia Y, Zhang H, Jia J, ... Liu L (2019a). Identification of cucumber circular RNAs responsive to salt stress. *BMC Plant Biology* 19:164. <https://doi.org/10.1186/s12870-019-1712-3>
- Zhu YX, Yin J, Liang Y, Liu J, Jia J, Huo H, ... Gong H (2019b). Transcriptomic dynamics provide an insight into the mechanism for silicon-mediated alleviation of salt stress in cucumber plants. *Ecotoxicology and Environmental Safety* 174:245-254. <https://doi.org/10.1016/j.ecoenv.2019.02.075>
- Zhu Y, Gong HJ, Yin JL (2019c). Role of silicon in mediating salt tolerance in plants: a review. *Plants* 8:147. <https://doi.org/10.3390/plants8060147>
- Zhu YX, Yang L, Liu N, Yang J, Zhou XK, Xia YC, ... Yin JL (2019d). Genome-wide identification, structure characterization, and expression pattern profiling of aquaporin gene family in cucumber. *BMC Plant Biology* 19: 345. <https://doi.org/10.1186/s12870-019-1953-1>



The journal offers free, immediate, and unrestricted access to peer-reviewed research and scholarly work. Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the articles, or use them for any other lawful purpose, without asking prior permission from the publisher or the author.



**License** - Articles published in *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* are Open-Access, distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) License.

© Articles by the authors; UASVM, Cluj-Napoca, Romania. The journal allows the author(s) to hold the copyright/to retain publishing rights without restriction.