doi 10.18699/bgrs2024-12.3-23

AthRiboNC: unveiling the translational potential of Arabidopsis ncRNAs

Shen Y., Chen M.*

Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou, China

** mchen@zju.edu.cn*

Key words: ncRNAs; sORFs; Ribo-seq; *Arabidopsis thaliana*; coding potential

Motivation and Aim: Non-coding RNAs (ncRNAs) play significant regulatory roles in gene expression. Recent studies have discovered that some ncRNAs harbor small open reading frames (sORFs) with the potential to encode active peptides. However, research on ncRNA-encoded peptides in plants is not extensive, and there is a lack of systematic databases. Therefore, we developed AthRiboNC, which provides evidence related to Arabidopsis ncRNA-encoded peptides, with the hope of offering bioinformatics guidance for wet lab experimental research.

Methods and Algorithms: We collected Arabidopsis Ribo-seq data and sample processing metadata from 226 samples across 26 different BioProjects from the GEO and SRA databases [1]. The Ribo-seq data were aligned to the TAIR10 reference genome [2] and the PlantcircBase [3] circRNA library using HISAT2 [4]. FeatureCounts [5] was used to quantify the translation levels of miRNAs, lncRNAs, and circRNAs in different samples. ORFfinder and RiboTISH [6] were employed to identify sORFs in ncRNAs with translational potential. Co-expression network analyses were carried out using the WGCNA package in R [7] to relate the translation level of ncRNAs to sample treatment and to distinguish modules. Functional analyses were based on the hub genes of the modules. Finally, the front-end and back-end of the database website were built using Django, MySQL and jQuery.

Results: AthRiboNC provides a rich catalog of Arabidopsis ncRNAs with translational potential identified from Ribo-seq data. The database includes 4869 Arabidopsis ncRNAs detected by Ribo-seq, comprising 2743 lncRNAs, 255 miRNAs, and 1871 circRNAs. Users can conveniently browse and search for Arabidopsis ncRNAs with translational potential, their sORFs, and expression levels in samples within the database. The co-expression module offers additional downstream knowledge, such as members of co-expression modules, module-trait associations, module hub genes, and functional enrichment. Additionally, the website provides a simple BLAST function to align target sequences with all sORFs in AthRiboNC. AthRiboNC is available at https://bis.zju.edu.cn/athribonc/.

Conclusions: AthRiboNC serves as a pivotal resource for exploring the translational landscape of ncRNAs in Arabidopsis. It underlines the significance of Ribo-seq data for identifying functional peptides from ncRNAs and provides a valuable tool for wet lab researchers to investigate gene function and regulatory networks.

Funding: The study is supported by the National Natural Sciences Foundation of China (No. 32070677).

References

1. Kenneth K., Shutov O., Lapoint R. et al. The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Res.* 2022;50(D1):D387-D390

2. Cheng C.Y., Krishnakumar V., Chan A.P. et al. Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J*. 2017;89(4):789-804
3. Xu X., Du T., Mao W. et al. PlantcircBase 7.0: Full-length transcripts and conservation of plant circRNAs. *Plant Commun*. 2022;3(4):100343
4. Kim D., Paggi J.M., Park C. et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37(8):907-915
5. Liao Y., Smyth G.K., Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930
6. Zhang P., He D., Xu Y. et al. Genome-wide identification and differential analysis of translational initiation. *Nat Commun*. 2017;8(1):1749
7. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559