

# The missing links: Finding function in lincRNAs

July 19 2022, by Michael J. Haas

---



Kyle Palos examines some *Arabidopsis* growing in a BTI growth chamber.  
Credit: Boyce Thompson Institute.

Genomes contain regions between protein-coding genes that produce lengthy RNA molecules that never give rise to a protein. These long

intergenic non-coding RNAs (lincRNAs) are thought to have important functions, such as regulating responses to environmental change. However, a paucity of well-annotated lincRNA data, especially for crop plants, has precluded a deeper understanding of their roles.

Up until now, there have been no systematic genome-wide studies that both confirmed DNA sequences that produce lincRNAs and proposed functions for those lincRNAs. Plus, data are reported differently across studies, making direct comparisons among them difficult.

These barriers inspired researchers at the Boyce Thompson Institute to take a comprehensive look at the identity, production and function of lincRNAs in four species in the mustard family, including the model organism *Arabidopsis thaliana*, and *Brassica rapa*, a species that produces boy choy, turnips and other [food crops](#).

The group found locations across all four genomes that encoded lincRNAs, proposed functions for them, and confirmed the function of some lincRNAs involved in germination in *Arabidopsis*, creating an approach that could help researchers further understand the enigmatic molecules in all species, from crops to humans.

Their results were published in *The Plant Cell* on June 6.

"Our goal was to generate extensive and more actionable data for researchers to understand lincRNA function," said Kyle Palos, first author on the study and a postdoctoral fellow in the lab of BTI Assistant Professor Andrew Nelson.

"This project started small and mushroomed after we realized we couldn't begin to figure out lincRNA function without having thoroughly annotated genomes to know what lincRNAs were even present," said Nelson, who is the corresponding author of the paper. "Kyle really led

the charge on everything that went into this paper."

The team hypothesized that lincRNA production and function are limited to certain cell types and environmental conditions. The more common data sets don't cover that level of detail, "so it's easy to miss a lot due to limited sampling," Palos said. "Our comprehensive approach merges a [high-throughput](#), top-level analysis that identifies lincRNAs with a deeper dive into their likely functions, to give the full picture."

The study utilized a unified approach to gathering and annotating lincRNA data that could be easily adopted by other groups. According to Palos, this would facilitate comparisons across different experiments and species as the body of plant lincRNA data continues to grow.

The team uploaded its results to CyVerse, a free and open-science workspace where researchers can store, access and analyze data all in one place.

"We made it as simple as possible for others to search our results for lincRNAs involved in a plant trait or pathway of interest, and in responses to temperature and other environmental stressors," Palos said.

The team's methods could also help resolve long-standing questions with [genome-wide association studies](#) (GWAS) that identify correlations between plant traits and gene variants: What is going on with the variants that fall outside of protein-coding regions? Are these variants within other genes (i.e., lincRNAs) or regulatory elements?

"You need a properly annotated genome to know that, before you can determine the variant's effect and how to modulate it to produce your crop of choice," said Nelson, who is also an adjunct assistant professor in the School of Integrative Plant Science at Cornell University.

In the study, the team processed over 20,000 publicly available RNA sequence data sets from the four mustard species, supplemented with its own sequencing data, to identify thousands of lincRNAs, and then annotated the lincRNAs with genomic, structural and other information.

They assigned putative functions to the lincRNAs based on their similar expression patterns to protein-coding genes with known functions. Next, the team deleted a subset of lincRNAs that appeared to play roles in [seed germination](#) and development in Arabidopsis, which led to reductions in germination, thus validating their approach to determine lincRNA function.

In addition to ongoing studies of the germination-related lincRNAs, the team is applying its methods to lincRNAs in four more important crops for which a wealth of RNA sequence data is available—rice, maize, sorghum and *Setaria italica* (foxtail millet)—and has plans to expand into another nine well-sequenced species.

"Plant [genome research](#) often falls behind mammalian research, but with lincRNAs, we're still very much in the dark across all species," Nelson said. "Researching lincRNA in plants could have an impact on human health and crops alike, by helping us understand their fundamental properties, regardless of the species."

**More information:** Kyle Palos et al, Identification and functional annotation of long intergenic non-coding RNAs in Brassicaceae, *The Plant Cell* (2022). [DOI: 10.1093/plcell/koac166](https://doi.org/10.1093/plcell/koac166)

CyVerse results: [de.cyverse.org/data/ds/iplant/ ... blic?type=folder&re=](https://de.cyverse.org/data/ds/iplant/...blic?type=folder&re=)

Provided by Boyce Thompson Institute

Citation: The missing links: Finding function in lincRNAs (2022, July 19) retrieved 9 April 2024 from <https://phys.org/news/2022-07-links-function-lincrnas.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.