Scientists improve yield predictions based on seedling data

January 8 2020

Using plant RNA data from 2-week-old corn seedlings, Shinhan Shiu has shown that farmers and scientists can improve adult crop trait predictions with accuracy that rivals current approaches using DNA, i.e. genetic data. Credit: Kurt Stepnitz

A doctor diagnosing a 50-year-old patient based on a blood test taken during the patient's infancy would be unthinkable.

Anecdotally speaking, however, that's what Michigan State University scientists have done with corn. Using plant RNA data from 2-week-old corn seedlings, Shinhan Shiu, professor of plant biology and computational mathematics, science and engineering, has shown that farmers and scientists can improve adult crop trait predictions with accuracy that rivals current approaches using DNA, i.e. genetic data.

"Traditional breeding methods take months to years, which can be saved if we can predict the desirable traits just from DNA and RNA without growing them, without having to measure the actual traits directly," said Shiu, senior author of the paper appearing in the current issue of *The Plant Cell*. "To continue the human medicine anecdote, it's like sequencing an infant's RNA and analyzing what sort of traits the infant may develop later in life."

Shiu has long been fascinated with using computational approaches to resolve evolution and genome biology questions. A well-recognized grand challenge in biology is how to connect information in the DNA, or genotype, with traits, or phenotype. Solving this mystery is fundamental to understanding how genetic information is translated into outward traits in any species, Shiu said.

Since RNA is a product of DNA, one step closer to the traits DNA ultimately influences, the RNA blueprints can potentially offer better predictions. Using machine learning approaches, Shiu and his colleagues have taken a step closer to connecting DNA, RNA and the underlying traits.

"This is helpful for new breeding programs and may have implications in new ways to do genetic testing," Shiu said. "We found that RNA measurements provide additional information that we cannot get from DNA alone." In terms of reproduction, for example, the team was able to make accurate flowering and yield predictions—even before the plants had developed their seed or flower organisms.

Traditional methods using genetic marker-based models identified only one of 14 known genes linked to flowering time as important. However, the gene expression-based model created by Shiu and his colleagues identified five.

Even with this increased accuracy, though, Shiu's team isn't saying the new method should replace the old.

"Our findings are complementary to genetic marker-based prediction and identifies gene expression-trait associations that are not explained by genetic markers," Shiu said. "Not only does this help in selection of breeding lines with desirable traits, but also enhances our understanding of the mechanisms involved in these processes."

Future research will work to improve the model's accuracy, efficiency and cost.