

One pipeline that combines many gene-finding tools

January 12 2015

Reconstructing relationships among plant species is critical for a better understanding of fundamental aspects of plant biology, such as genome evolution, speciation, pollination biology, and many other areas.

However, evolutionary relationships within many plant groups have been difficult to reconstruct, often due to factors such as rapid diversification, hybridization, and incomplete lineage sorting. In such cases, large data sets including numerous genetic markers from both the nuclear and organellar genomes are necessary to accurately determine relationships. But developing numerous genetic markers for phylogenetic study is often difficult in the absence of extensive genomic resources, which are available for only a small portion of [plant species](#).

Fortunately, advances in sequencing technologies are helping researchers develop large data sets even in challenging genomes. One example of this is the MAKER2 Annotation Pipeline, a free and open-source pipeline that combines a variety of bioinformatic tools for genome analysis, better known as "genome annotation."

Using MAKER2, researchers at Ohio State University have developed a workflow that identifies numerous genetic markers for phylogenetic study from limited [genomic data](#). The approach, published in a recent issue of *Applications in Plant Sciences*, is exemplified using the flowering plant genus *Penstemon* (Plantaginaceae), whose phylogeny has been difficult to resolve due to its recent, rapid radiation.

Based on low-coverage (ca. 0.005×-0.007×) genomic data from six *Penstemon* samples, obtained via 454 sequencing, the researchers used MAKER2 to identify [genetic markers](#) useful for phylogenetic study from all three plant genomes. Primers for the selected loci were then designed using Primer-BLAST and Primer3Plus, which could then be used for PCR-based parallel sequencing.

Paul D. Blischak, lead author of the study, notes that the advantages MAKER2 offers for developing markers from low-coverage genomic data include the "ability to accurately identify gene regions, [to] collect all available information (gene predictions, BLAST hits, exon boundaries, etc.) for an identified gene into a single file for easy extraction, and its compatibility with visualization tools." With these features, he says, "the workflow makes it easy for researchers to select numerous loci useful for their phylogenetic study."

Although many researchers have used transcriptomic data for marker development, Blischak notes that using genomic data makes it easier to identify introns, which are more useful for reconstructing shallow or recent [evolutionary relationships](#).

"Targeting introns is a fairly common approach for marker development, which would likely be more difficult with transcriptome data because the only introns available are those that aren't spliced out."

The study demonstrates how MAKER2 functionality, which combines multiple programs into a single pipeline, can be used to create large-scale data sets, even for extremely low-coverage genomes. The authors also provide a sample protocol, sequence libraries, functional annotation files, and other resources.

More information: Paul D. Blischak, Aaron J. Wenzel, and Andrea D. Wolfe. 2014. Gene prediction and annotation in *Penstemon*

(Plantaginaceae): A workflow for marker development from extremely low-coverage genome sequencing. *Applications in Plant Sciences* 2(12): 1400044. [DOI: 10.3732/apps.1400044](https://doi.org/10.3732/apps.1400044)

Provided by Botanical Society of America

Citation: One pipeline that combines many gene-finding tools (2015, January 12) retrieved 27 April 2024 from <https://phys.org/news/2015-01-pipeline-combines-gene-finding-tools.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.