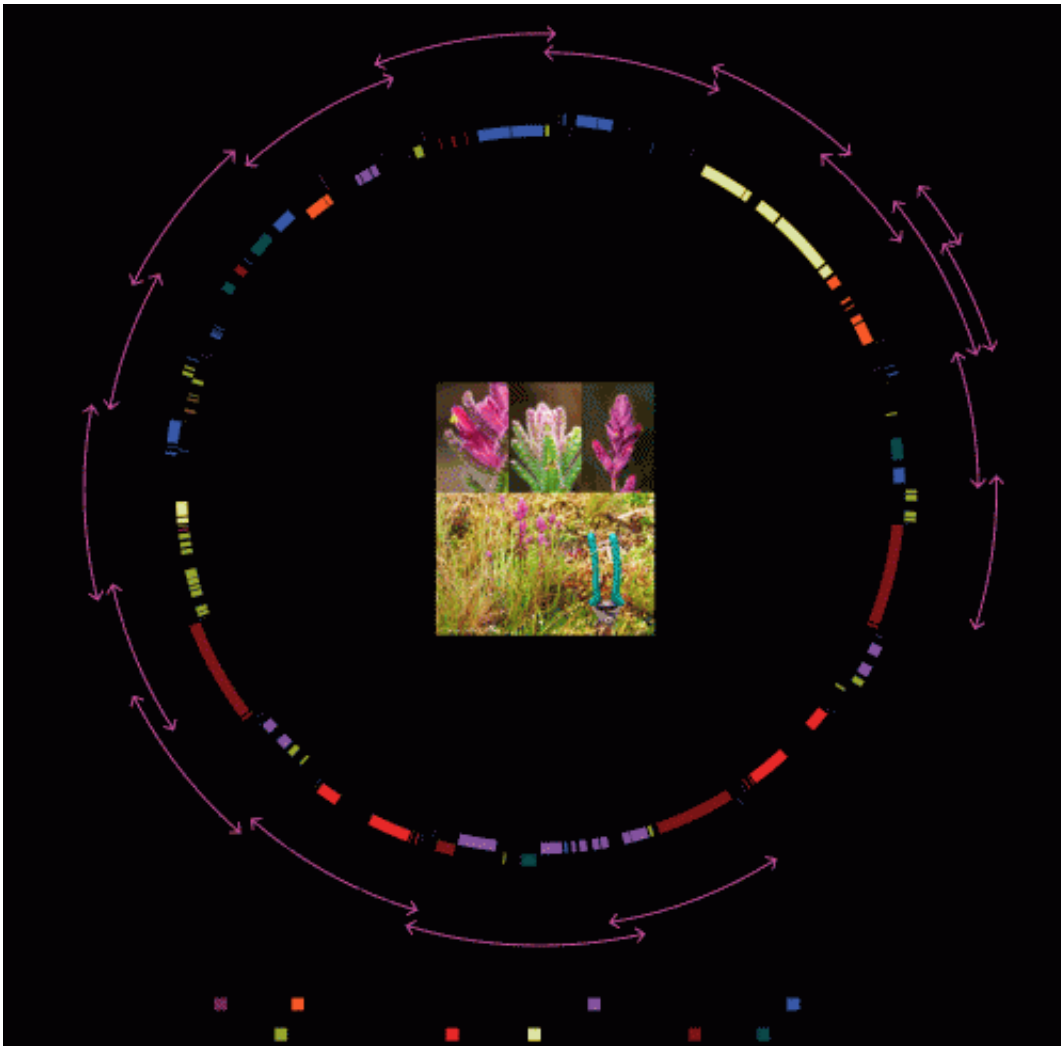


Big data: A method for obtaining large, phylogenomic data sets

January 9 2014



The final annotated chloroplast genome assembly of *Bartsia inaequalis* with the 16 overlapping primer combinations indicated. Note that the primer combinations for regions 11, 12, 13, and 16 amplify both inverted repeat A and B in a single reaction. Credit: Photo by Simon Uribe-Convers from: Uribe-

Convers, S., J. R. Duke, M. J. Moore, and D. C. Tank. 2014. A long PCR-based approach for DNA enrichment prior to next-generation sequencing for systematic studies. *Applications in Plant Sciences* 2(1): 1300063. doi:10.3732/apps.1300063

Traditional molecular systematic studies have progressed by sequencing genes one by one, a time- and cost-intensive task that has limited the amount of data a researcher could feasibly obtain. With the continual improvement of next-generation sequencing technologies, however, obtaining large molecular data sets is becoming much easier, and much cheaper. This increase in data means, in many cases, increased accuracy in reconstructing the evolutionary history of organisms.

As phylogenetic studies advance to include progressively more sequence data, new techniques are being developed to obtain such data sets. While it would be ideal to simply sequence entire genomes, this is not yet feasible across large numbers of taxa. Instead, current methods are being developed that allow researchers to target specific genomic regions of interest for the organisms being studied.

Scientists at the University of Idaho and Oberlin College have developed one such method to obtain large, phylogenomic data sets. "This method utilizes long PCR, or long-range PCR, to strategically generate DNA templates for next-generation sequencing," explains Simon Uribe-Convers, graduate student and lead author. The protocol is available for free viewing in the January issue of *Applications in Plant Sciences*.

Long-range PCR is a method that allows for the amplification of much larger fragments of DNA than is possible with traditional PCR—fragments larger than 40 kilobases have been reported in long PCR, versus fewer than 10 kilobases for traditional PCR. The authors of

this study have developed a universal primer set across flowering plants that amplifies 3–15 kilobase fragments, which can then easily be sequenced using recently developed next-generation sequencing technologies. Uribe-Convers and colleagues tested this approach by amplifying chloroplast genomes for 30 species across flowering plants. Surprisingly, the primers were even found to successfully amplify chloroplast regions in several pine species. To further test the compatibility of this approach with next-generation sequencing, 15 complete chloroplast genomes (often referred to as plastomes) were then sequenced.

Although this study focused on plastomes and utilized the popular Illumina sequencing platform, Uribe-Convers explains, "[t]his can easily be expanded to mitochondrial and nuclear regions, and can be used in combination with any next-generation sequencing platform. Furthermore, this approach is not restricted to plant studies, but will be useful for any organism."

With the development of new methods such as the one described by Uribe-Convers and colleagues, scientists can obtain large, phylogenomic data sets for large numbers of taxa. Long-range PCR, in concert with next-generation sequencing, provides researchers with the means to sequence entire plastomes, mitochondrial genomes, and large portions of the nuclear genome.

"This method has important implications for the way future systematic studies are conducted as it provides researchers with a way to strategically target regions of interest in their study organism, such as single-copy regions of the nuclear genome or portions of organellar genomes, to produce large data sets at low costs," says Uribe-Convers. "We want to help move the field of systematics into the realm of big data, and we hope that our approach contributes to that."

More information: Simon Uribe-Convers, Justin R. Duke, Michael J. Moore, and David C. Tank. 2014. A long PCR-based approach for DNA enrichment prior to next-generation sequencing for systematic studies. *Applications in Plant Sciences* 2(1): 1300063. [DOI: 10.3732/apps.1300063](https://doi.org/10.3732/apps.1300063)

Provided by American Journal of Botany

Citation: Big data: A method for obtaining large, phylogenomic data sets (2014, January 9) retrieved 23 April 2024 from <https://phys.org/news/2014-01-big-method-large-phylogenomic.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.