Applications in Plant Sciences

Applications in Plant Sciences 2014 2(1): 1300063

Protocol Note

A long PCR–based approach for DNA enrichment prior to next-generation sequencing for systematic studies[1]

Simon Uribe-Convers[2,3,5], Justin R. Duke[3], Michael J. Moore[4], and David C. Tank[2,3]

[2]Department of Biological Sciences and Institute for Bioinformatics and Evolutionary Studies, University of Idaho, 875 Perimeter Drive MS 3051, Moscow, Idaho 83844-3051 USA; [3]College of Natural Resources, University of Idaho, 875 Perimeter Drive MS 1133, Moscow, Idaho 83844-1133 USA; and [4]Department of Biology, Oberlin College, Science Center K111, 119 Woodland St., Oberlin, Ohio 44074-1097 USA

- *Premise of the study:* We present an alternative approach for molecular systematic studies that combines long PCR and next-generation sequencing. Our approach can be used to generate templates from any DNA source for next-generation sequencing. Here we test our approach by amplifying complete chloroplast genomes, and we present a set of 58 potentially universal primers for angiosperms to do so. Additionally, this approach is likely to be particularly useful for nuclear and mitochondrial regions.
- *Methods and Results:* Chloroplast genomes of 30 species across angiosperms were amplified to test our approach. Amplification success varied depending on whether PCR conditions were optimized for a given taxon. To further test our approach, some amplicons were sequenced on an Illumina HiSeq 2000.
- *Conclusions:* Although here we tested this approach by sequencing plastomes, long PCR amplicons could be generated using DNA from any genome, expanding the possibilities of this approach for molecular systematic studies.

Key words: angiosperms; chloroplast enrichment; long PCR; next-generation sequencing; plastome; universal chloroplast PCR primers.

Advancements in next-generation sequencing (NGS) technologies have permitted the assembly of large, genome-scale data sets that have shed light on the evolutionary history of many taxa (e.g., Parks et al., 2009; Moore et al., 2010; Xi et al., 2012; Eaton and Ree, 2013; Tennessen et al., 2013). For plant phylogenetics, there has been a major focus on methods for chloroplast phylogenomics (e.g., Parks et al., 2009; Moore et al., 2010), although methods for collecting phylogenomic data sets from the nuclear and mitochondrial genomes have also been developed (e.g., Straub et al., 2012; Eaton and Ree, 2013). Stull et al. (2013) developed a custom RNA probe set designed to capture angiosperm plastomes via solution-based hybridization. While their capture system was broadly successful, Stull et al. (2013) found that the most variable spacer regions were often captured at much-reduced coverage compared to more conserved regions, and were sometimes missed entirely if the target taxon was phylogenetically divergent from one of the 22 plastomes used in the bait design. Moreover, the current cost of the capture probes makes this method most efficient for projects dealing with hundreds of species. Another commonly employed method for plant phylogenomic studies is genome skimming (Straub et al., 2012), which takes advantage of the fact that organellar DNA and nuclear ribosomal DNA are present at high copy numbers in genomic DNA. However, a significant limitation of this method for systematic studies is that only high-copy number regions are recovered consistently across all samples, whereas regions with lower representation are only recovered in some samples and missed completely in others (Straub et al., 2011). This can be problematic for molecular systematic studies where missing data may result in misleading phylogenetic results (Lemmon et al., 2009). Moreover, being limited to high-copy regions in the genome becomes restrictive for experimental design as it excludes putatively highly informative regions in the genome such as single-copy nuclear genes (e.g., the single-copy orthologous genes [COSII] and the pentatricopeptide repeat [PPR] gene family; Wu et al., 2006, and Yuan et al., 2009, respectively).

As an alternative, we present an NGS approach that combines long PCR and Illumina sequencing to strategically compile phylogenomic data sets for molecular systematic studies. Long PCR, or long-range PCR, uses a combination of two polymerases—a nonproofreading polymerase at high concentration and a proofreading polymerase at a lower concentration—to amplify DNA fragments that range between 3 and 15 kilobases (kb), although cases of extremely large fragments (22–42 kb) have been reported (e.g., Cheng et al., 1994). Long PCR has been used extensively in human genome projects (e.g., Craig et al., 2008) and to sequence complete mitochondrial genomes (e.g., Knaus et al., 2011; Alexander et al., 2013), using both

[1]Manuscript received 26 July 2013; revision accepted 3 December 2013.

The authors thank R. Cronn, T. C. Peterson, D. F. Morales-Briones, S. J. Jacobs, H. E. Marx, and four anonymous reviewers for helpful comments on the manuscript, and the University of Idaho Institute for Bioinformatics and Evolutionary Studies (NIH/NCRR P20RR16448 and P20RR016454) and the iPlant Collaborative (NSF DBI-0735191) for bioinformatic resources. Funding for this work was provided by the National Science Foundation (DEB-1210895 to D.C.T. for S.U.C., and DEB-1253463 to D.C.T.) and by the University of Idaho Student Grant Program and Seed Grant Program to S.U.C. and D.C.T., respectively.

[5]Author for correspondence: uribe.convers@gmail.com

doi:10.3732/apps.1300063

