Applications in Plant Sciences

Applications in Plant Sciences 2014 2(9): 1400042

Protocol Note

Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics[1]

Kevin Weitemier[2,7], Shannon C. K. Straub[2,7], Richard C. Cronn[3], Mark Fishbein[4], Roswitha Schmickl[5], Angela McDonnell[4], and Aaron Liston[2,6]

[2]Department of Botany and Plant Pathology, Oregon State University, 2082 Cordley Hall, Corvallis, Oregon 97331 USA; [3]Pacific Northwest Research Station, USDA Forest Service, 3200 SW Jefferson Way, Corvallis, Oregon 97331 USA; [4]Department of Botany, Oklahoma State University, 301 Physical Sciences, Stillwater, Oklahoma 74078 USA; and [5]Institute of Botany, Academy of Sciences of the Czech Republic, CZ-25243 Průhonice, Czech Republic

- *Premise of the study:* Hyb-Seq, the combination of target enrichment and genome skimming, allows simultaneous data collection for low-copy nuclear genes and high-copy genomic targets for plant systematics and evolution studies.
- *Methods and Results:* Genome and transcriptome assemblies for milkweed (*Asclepias syriaca*) were used to design enrichment probes for 3385 exons from 768 genes (>1.6 Mbp) followed by Illumina sequencing of enriched libraries. Hyb-Seq of 12 individuals (10 *Asclepias* species and two related genera) resulted in at least partial assembly of 92.6% of exons and 99.7% of genes and an average assembly length >2 Mbp. Importantly, complete plastomes and nuclear ribosomal DNA cistrons were assembled using off-target reads. Phylogenomic analyses demonstrated signal conflict between genomes.
- *Conclusions:* The Hyb-Seq approach enables targeted sequencing of thousands of low-copy nuclear exons and flanking regions, as well as genome skimming of high-copy repeats and organellar genomes, to efficiently produce genome-scale data sets for phylogenomics.

Key words: genome skimming; Hyb-Seq; nuclear loci; phylogenomics; species tree; target enrichment.

The importance of incorporating low-copy nuclear genes in phylogenetic reconstruction is well-recognized, but has largely been constrained by technical limitations (Zimmer and Wen, 2013). These data are essential for reconstructing the evolutionary history of plants, including understanding the causes of observed incongruities among gene trees that arise from incomplete lineage sorting and introgressive hybridization. The combination of solution hybridization for target enrichment of specific genomic regions and the high sequencing throughput of current platforms (e.g., Illumina) provides the opportunity to sequence hundreds or thousands of low-copy nuclear loci appropriate for phylogenetic analyses in an efficient and cost-effective manner (Cronn et al., 2012; Lemmon and Lemmon, 2013). Most efforts to date for targeted sequencing of plant genomes for phylogenetics have been directed at the plastome (e.g., Parks et al., 2012; Stull et al., 2013). Recently, conserved orthologous sequences in Asteraceae (Chapman et al., 2007) were obtained via target enrichment for phylogenomics (Mandel et al., 2014).

Methods have been developed to target highly or ultra-conserved elements (UCEs) in animal genomes (Faircloth et al., 2012; Lemmon and Lemmon, 2013; McCormack et al., 2013). However, UCEs in plants are nonsyntenic, and are hypothesized to have originated via horizontal transfer from organelles or de novo evolution (Reneker et al., 2012). Whatever their origin, their potential for nonorthology among species makes them unsuitable as phylogenetic markers in plants. The frequency of polyploidy throughout angiosperm evolution (Jiao et al., 2011) also impedes obtaining a large set of conserved orthologous single-copy loci transferable across plant lineages, which in combination with the lack of orthologous UCEs, means that design of targeted sequencing strategies for plant nuclear genomes will necessarily be lineage-specific.

Here we present Hyb-Seq, a protocol that combines target enrichment of low-copy nuclear genes and genome skimming (Straub et al., 2012), the use of low-coverage shotgun sequencing to assemble high-copy genomic targets. Our protocol improves upon the methods of Mandel et al. (2014) by (1) utilizing the genome and transcriptome of a single species for probe design, which makes our approach more generally applicable to any plant lineage; (2) obtaining additional data from the procedure through combination with genome skimming; and (3) developing a data analysis pipeline that maximizes the data usable for phylogenomic analyses. Furthermore, we assemble sequences from the flanking regions (the "splash zone") of targeted exons, yielding noncoding sequence from introns or sequence 5′ or 3′ to genes, which are potentially

[1]Manuscript received 16 May 2014; revision accepted 25 June 2014.

The authors thank M. Dasenko, Z. Foster, K. Hansen, S. Jogdeo, Z. Kamvar, L. Mealy, M. Parks, M. Peterson, C. Sullivan, and L. Worchester for laboratory, sequencing, data analysis, and computational support. J. M. Rouillard with MYcroarray provided valuable technical support. The authors thank J. Mandel and another anonymous reviewer for comments on a previous version of this manuscript. This work was funded by the U.S. National Science Foundation (DEB 0919583).

[6]Author for correspondence: listona@science.oregonstate.edu

[7]These authors contributed equally to this work.

doi:10.3732/apps.1400042

