**SOFTWARE NOTE**

**Plann: A Command-line Application for Annotating Plastome Sequences**

Daisie I. Huang\(^2,3\) and Quentin C. B. Cronk\(^2\)

\(^3\)Department of Botany, University of British Columbia, 3529-6270 University Blvd., Vancouver, British Columbia V6T 1Z4, Canada

\(^1\)Manuscript received 18 March 2015; revision accepted 5 July 2015. This contribution was written while D.I.H. was salaried via the Genome Canada Large-Scale Applied Research Project (grant no. 168BIO: PIs Carl Douglas and Shawn Mansfield, University of British Columbia). Q.C.B.C. acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants Program (grant no. RGPIN-2014-05820).

1. **Premise of the study:** Plann automates the process of annotating a plastome sequence in GenBank format for either downstream processing or for GenBank submission by annotating a new plastome based on a similar, well-annotated plastome.

2. **Methods and Results:** Plann is a Perl script to be executed on the command line. Plann compares a new plastome sequence to the features annotated in a reference plastome and then shifts the intervals of any matching features to the locations in the new plastome. Plann’s output can be used in the National Center for Biotechnology Information’s tbl2asn to create a Sequin file for GenBank submission.

3. **Conclusions:** Unlike Web-based annotation packages, Plann is a locally executable script that will accurately annotate a plastome sequence to a locally specified reference plastome. Because it executes from the command line, it is ready to use in other software pipelines and can be easily rerun as a draft plastome is improved.

**Key words:** chloroplast; GenBank; genome annotation; plastome; Sequin.

As next-generation sequencing becomes the primary source of molecular variation for phylogenetic analyses, sequencing and assembling whole plastomes is more feasible than ever. Recent papers (Kane et al., 2012; Bock et al., 2013; Ripma et al., 2014) have highlighted the ease of genome skimming techniques for assembling the relatively high-coverage plastome and chondriome (mitochondrial genome) from lower-level genomic DNA short-read sequencing. De novo assembly programs can easily assemble large contiguous sequences of plastome; these sequences can be aligned to the generally syntenic and conserved gene order of the plastome, which makes whole-plastome alignments ideal for phylogenomic and phylogeographic studies (Cronn et al., 2008; Parks et al., 2009, 2012; Straub et al., 2012; Njugu et al., 2013). Once a plastome is assembled, however, it still needs to be annotated. Generally, new plastomes are annotated in reference to known plastid genes and features, using available online tools such as DOGMA (Wyman et al., 2004) or CpGAVAS (Liu et al., 2012). These tools are excellent for annotating and visualizing a single plastome sequence from a single taxon, but can become tedious for annotation of plastomes for which a well-annotated sequenced relative already exists. In this situation, an interactive interface that repeatedly queries a broad range of annotated plastomes is needlessly complex.

Here we introduce Plann (Plastome Annotator), a command-line tool to automate the annotation of newly assembled plastomes based on a single well-annotated related reference plastome. Because it runs locally on the user’s machine (or on a Unix server), it can be easily run independently or pipelined into existing workflows. Its inputs are minimal: it requires only the new plastome sequence and the GenBank-formatted file of the related annotated plastome to match. Its output can be immediately used in the National Center for Biotechnology Information’s (NCBI) command-line tool tbl2asn to create a GenBank-ready submission. Features in the reference genome that were not matched in the new plastome are included as alignments in the report so that more attention can be paid to possible misassemblies in those regions. Plastome sequences are often used for deep phylogenetics in plants because of their highly conserved gene order and low mutation rate. In closely related taxa, the coding sequence of the genes is nearly identical, with few insertions or deletions; the variation is mostly present in the intergenic spacers. Because of these features, annotating the plastome using the template of a closely related pre-annotated plastome is a straightforward process.

**METHODS AND RESULTS**

The reference gene sequences to be used in the search can be obtained from a GenBank record that is already validated and has annotations describing the names and qualifiers for all of the genes in the plastome. Plann searches for sequences in the newly assembled plastome similar to those known genes and transforms those matched genes to their corresponding genomic locations. Because it only searches one sequence against one sequence, it is a very fast process: it should only take a few seconds to run.

Plann consists of Perl scripts contained in a GitHub repository (https://github.com/daisieh/plann/releases/tag/v1.1) and licensed under a BSD open-source license. It uses two freely available command-line tools from NCBI: BLASTN and tbl2asn. The graphical user interface (GUI) application Sequin, also available from NCBI, can be used to generate the template file required by tbl2asn and to validate the output of tbl2asn. It has been tested on Unix and...