

# Simplifying simple sequence repeats

January 31 2020

---

Simple sequence repeats (SSRs) are regions of DNA with high diversity, and they have long been a mainstay for botanists examining the genetic structure of plant populations. However, as the cost of sequencing DNA continues to plummet and genetic technologies advance, newer techniques for mapping genetic diversity such as genotyping-by-sequencing (GBS) or RAD-seq have begun to rival the traditional use of SSRs. In research presented in a recent issue of *Applications in Plant Sciences*, Dr. Mark Chapman optimized the process of identifying SSRs from genomic and transcriptomic data, helping to assure the continued use and relevance of SSRs in the age of high-throughput sequencing (HTS).

Sequence data generated using HTS can be used to identify candidate SSRs, for which researchers can design primers to examine genetic structure in a species. However, little work has been done to calibrate or optimize this process, both in terms of guidelines for reasonable parameters to specify, or what kind or depth of sequencing may be sufficient and appropriate to identify a workable set of SSRs.

"I've used transcriptome data for over a decade to generate [molecular markers](#) and have often wondered whether using genomes or transcriptomes would be preferable," said Dr. Chapman, Associate Professor in Ecology and Evolutionary Biology at the University of Southampton. This study found that each [data source](#) had its benefits; genomic data may be preferable in species with low polymorphism, but transcriptomic data usually assembles into longer sequences more amenable to designing primers, and these primers may be more

transferable across species.

"In addition, I always generate thousands of markers and only use a dozen or so, so I've always wondered what depth of sequencing would one have to generate to be sure of identifying a small number of markers for a basic population genetic study," said Dr. Chapman.

Researchers on a budget may look to generate the minimum necessary [sequence data](#) for SSR identification. Now these researchers have some guidance as to how many reads are sufficient: this study found that small assemblies of two million read pairs could generate about 200-2000 potential markers from the genome assemblies and about 600-3650 from the transcriptome assemblies.

As the cost of sequencing falls below the cost of labor for sample preparation, researchers are increasingly using newer techniques such as GBS and RAD-seq to map [genetic diversity](#) in populations. However, Dr. Chapman still sees a place for SSRs in the future of population genetics research. "SSRs have advantages over those other technologies that are unlikely to change even if costs go down, for example, the SSRs can be designed from specific genes of interest," said Dr. Chapman. "Also GBS and RAD-seq aren't really being explored for polyploids, whereas SSR scoring in polyploids can be done, with a bit of background information or careful design of primers. The untailed approach of GBS and RAD-seq is likely to resolve a lot of unscorable alleles in polyploids."

SSRs are a relatively inexpensive and efficient way to map genetic diversity in populations. The deluge of genetic data available from HTS can help to efficiently identify sets of SSRs, but until now there have not been clear guidelines for researchers seeking to do this work. In optimizing protocols and laying out major considerations in generating SSRs from genomic and transcriptomic data, Dr. Chapman has helped to

bring SSR studies up to date.

**More information:** Mark A. Chapman, Optimizing depth and type of high-throughput sequencing data for microsatellite discovery, *Applications in Plant Sciences* (2019). [DOI: 10.1002/aps3.11298](https://doi.org/10.1002/aps3.11298)

Provided by Botanical Society of America

Citation: Simplifying simple sequence repeats (2020, January 31) retrieved 19 April 2024 from <https://phys.org/news/2020-01-simple-sequence.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.