Seq2Sat & SatAnalyzer toolkit: towards comprehensive microsatellite genotyping from sequencing data

Peng Liu[1], Paul Wilson[2], Bridget Redquest[3], Sonesinh Keobouasone[1], and Micheline Manseau[1]

[1]Environment and Climate Change Canada National Wildlife Research Centre
[2]Trent University
[3]Trent University Department of Biology

September 12, 2023

Abstract

Accurate and efficient genotyping of microsatellite loci is essential for their application in population genetics and various demographic analysis. Protocols for next generation sequencing of microsatellite loci generate high-throughput and cross-compatible allele scoring characteristics: common issues associated with size separation on conventional capillary-based protocols. As a result, we have developed a novel, ultra-fast, all-in-one software Seq2Sat in C++ to support accurate automated microsatellite genotyping. It directly takes raw reads of microsatellite amplicons and subsequently performs read quality control before inferring genotypes based on depth of read, sequence composition and length. It does not produce any intermediate files, making I/O very efficient. Additionally, we developed a module in Seq2Sat for sex identification based on sex locus amplicons. We further developed a user-friendly website-based platform SatAnalyzer to conduct reads-to-report analyses by calling Seq2Sat to generate genotype tables and interactive genotype graphs for manual editing. SatAnalyzer also allows visualization of read quality and distribution across loci and samples to troubleshoot multiplex optimization and high-quality library preparation. To evaluate its performance, we benchmarked SatAnalyzer against conventional capillary gel electrophoresis and an existing microsatellite genotyping software MEGASAT. Results show that SatAnalyzer can achieve $>$ 0.993 genotyping accuracy and Seq2Sat is ˜ 5 times faster than MEGASAT despite many more informative tables and figures generated. Seq2Sat and SatAnalyzer are freely available at github (https://github.com/ecogenomicscanada/Seq2Sat) and dockerhub (https://hub.docker.com/r/rocpengliu/satanalyzer).

Hosted file

`seq2sat_satanalyzer_submitted_MER.docx` available at https://authorea.com/users/663762/articles/665800-seq2sat-satanalyzer-toolkit-towards-comprehensive-microsatellite-genotyping-from-sequencing-data