

A pipeline for effectively developing highly polymorphic SSR markers based on multi-sample genomic data

Hui Wang¹, Shenghan Gao², Yu Liu¹, Pengcheng Wang³, Zhengwang Zhang¹, and De Chen¹

¹Beijing Normal University

²State Key Laboratory of Microbial Resources

³Chinese Academy of Sciences Key Laboratory of Animal Ecology and Conservation Biology

September 24, 2021

Abstract

Simple sequence repeats (SSRs) are widely used genetic markers in ecology, evolution and conservation even in the genomics era, while a general limitation to their application is the difficulty of developing polymorphic SSR markers. Next-generation sequencing (NGS) offers the opportunity for the rapid development of SSRs; however, previous studies developing SSRs using genomic data from only one individual need redundant experiments to test the polymorphisms of SSRs. In this study, we designed a pipeline for the rapid development of polymorphic SSR markers from multi-sample genomic data. We used bioinformatic software to genotype multiple individuals using resequencing data, detected highly polymorphic SSRs prior to experimental validation, significantly improved the efficiency and reduced the experimental effort. The pipeline was successfully applied to a globally threatened species, the brown-eared pheasant (*Crossoptilon mantchuricum*), which showed very low genomic diversity. The 20 newly developed SSR markers were highly polymorphic, the average number of alleles was much higher than the genomic average. We also evaluated the effect of the number of individuals and sequencing depth on the SSR mining results, and we found that ten individuals and ~10X sequencing data were enough to obtain a sufficient number of polymorphic SSRs, even for species with low genetic diversity. Furthermore, the genome assembly of NGS data from the optimal number of individuals and sequencing depth can be used as an alternative reference genome if a high-quality genome is not available. Our pipeline provided a paradigm for the application of NGS technology to mining and developing molecular markers for ecological and evolutionary studies.

Hosted file

SSR20210809.docx available at <https://authorea.com/users/435307/articles/538421-a-pipeline-for-effectively-developing-highly-polymorphic-ssr-markers-based-on-multi-sample-genomic-data>



