

Practical guide for obtaining and validating chromosome-scale genome assemblies with Hi-C scaffolding

Kazuaki Yamaguchi¹, Mitsutaka Kadota¹, Osamu Nishimura¹, Yuta Ohishi¹, Yuki Naito², and Shigehiro Kuraku¹

¹RIKEN Center for Biosystems Dynamics Research

²Database Center for Life Science

January 3, 2021

Abstract

Recent development of ecological studies has been fueled by the introduction of massive information based on chromosome-scale genome sequences, even for species whose genetic linkage was previously not accessible. This was enabled mainly by the application of Hi-C, a method for genome-wide chromosome conformation capture which was originally developed for investigating long-range interaction of chromatin. Performing genomic scaffolding using Hi-C data is highly resource-demanding in elaborate laboratory steps for sequencing sample preparation, building primary genome sequence assembly as an input, and computation for genome scaffolding using Hi-C data, followed by careful validation. This article summarizes existing solutions for these steps and provides a test case of its application to a reptile species, the Madagascar ground gecko (*Paroedura picta*). Among frequently exerted metrics for evaluating scaffolding results, we investigate the validity of completeness assessment using single-copy reference orthologs and report problems with the widely used program pipeline BUSCO.

Hosted file

Text_ver2.7-KurakuFinal2.pdf available at <https://authorea.com/users/386784/articles/502254-practical-guide-for-obtaining-and-validating-chromosome-scale-genome-assemblies-with-hi-c-scaffolding>