

Technical considerations in Hi-C scaffolding and evaluation of chromosome-scale genome assemblies

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

¹RIKEN Center for Biosystems Dynamics Research

²Database Center for Life Science

May 24, 2021

Abstract

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

Hosted file

Manuscript-Final.pdf available at <https://authorea.com/users/386784/articles/523251-technical-considerations-in-hi-c-scaffolding-and-evaluation-of-chromosome-scale-genome-assemblies>