

Minimizing method bias for merged Reduced Representation Library and Whole Genome sequencing dataset

Linett Rasmussen¹ and Filipe Vieira¹

¹University of Copenhagen Globe Institute

October 14, 2022

Abstract

During the era of genomics, many sequencing methods have been (and continue to be) developed. One family of methods that has recently gotten a lot of popularity is the reduced representation library method. Yet, there has not been much research into how to deal with combining reduced representation library data with other types, like (e.g.) whole genome sequencing data. Merging of different types of data can be difficult and lead to biases in downstream data analysis. This study attempts to identify the origin of some of these biases and propose strategies to minimize them, by using a dataset of six wolves that have been sequenced both with MobiSeq and standard whole genome sequencing. Taking the whole genome sequencing data as reference, we have taken a step-by-step approach to identify parameters that minimize the bias produced by MobiSeq. Our results show that missing data has a large effect on the data analysis. Therefore, we recommend that areas for variant calling should be limited to targeted regions of the reduced representation library method, and in some analysis the addition of requiring a minimum number of individuals can minimize the bias even more.

Hosted file

main_text_minimizing_method_bias_for_RRL_and_WGS_dataset final.docx available at <https://authorea.com/users/514811/articles/590400-minimizing-method-bias-for-merged-reduced-representation-library-and-whole-genome-sequencing-dataset>







