



Transcriptome assembly quality affects phylogenomic inference



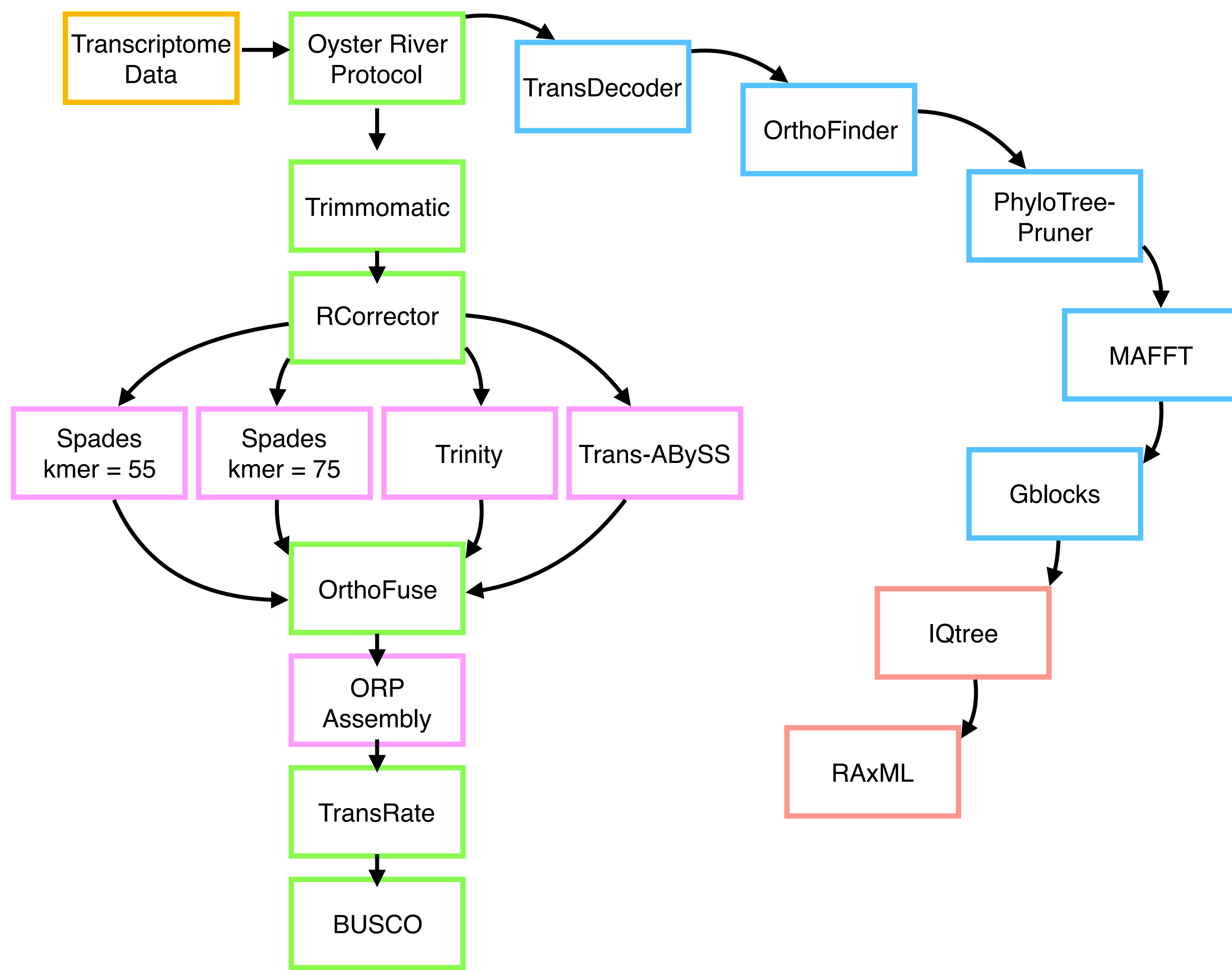
@jen_spillane

Jennifer L. Spillane, Troy M. LaPolice, Matthew D. MacManes, David C. Plachetzki

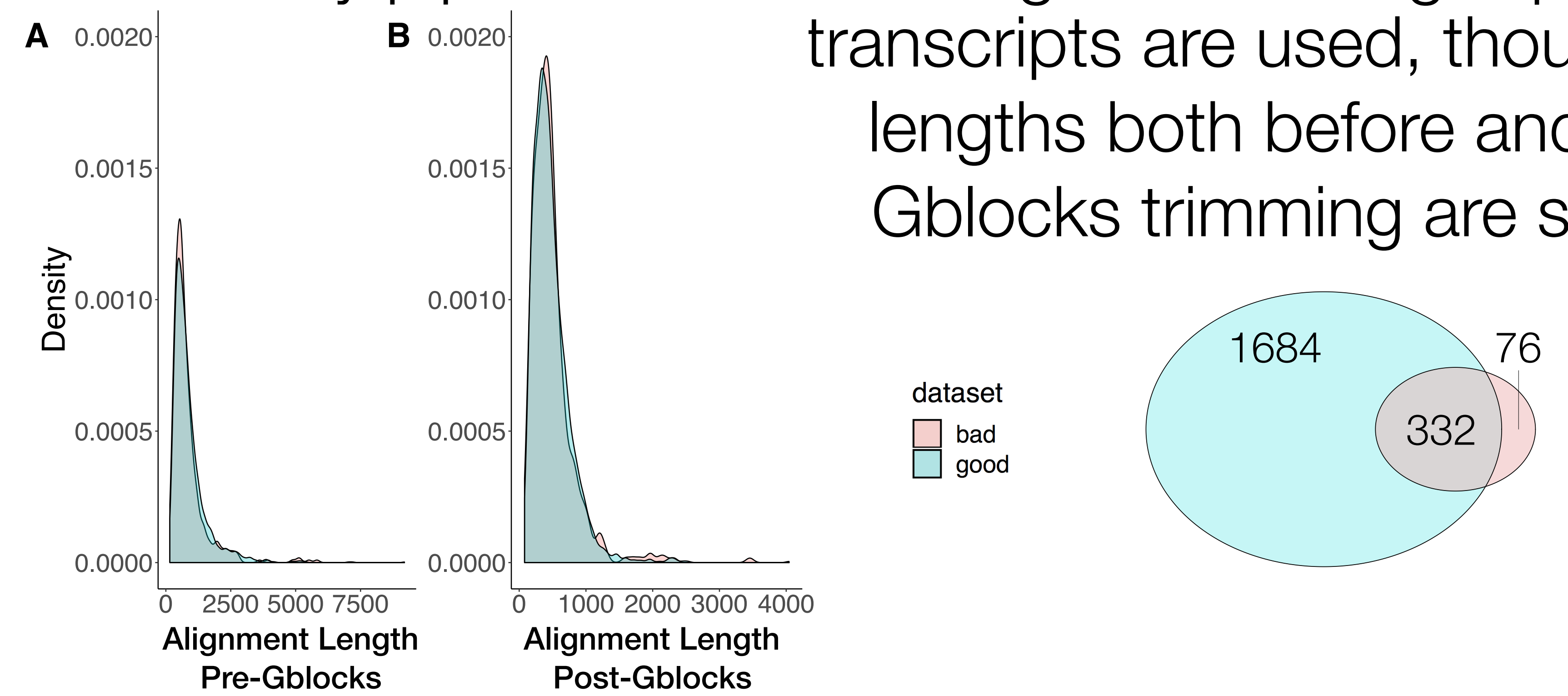
jlh1023@wildcats.unh.edu

University of New Hampshire

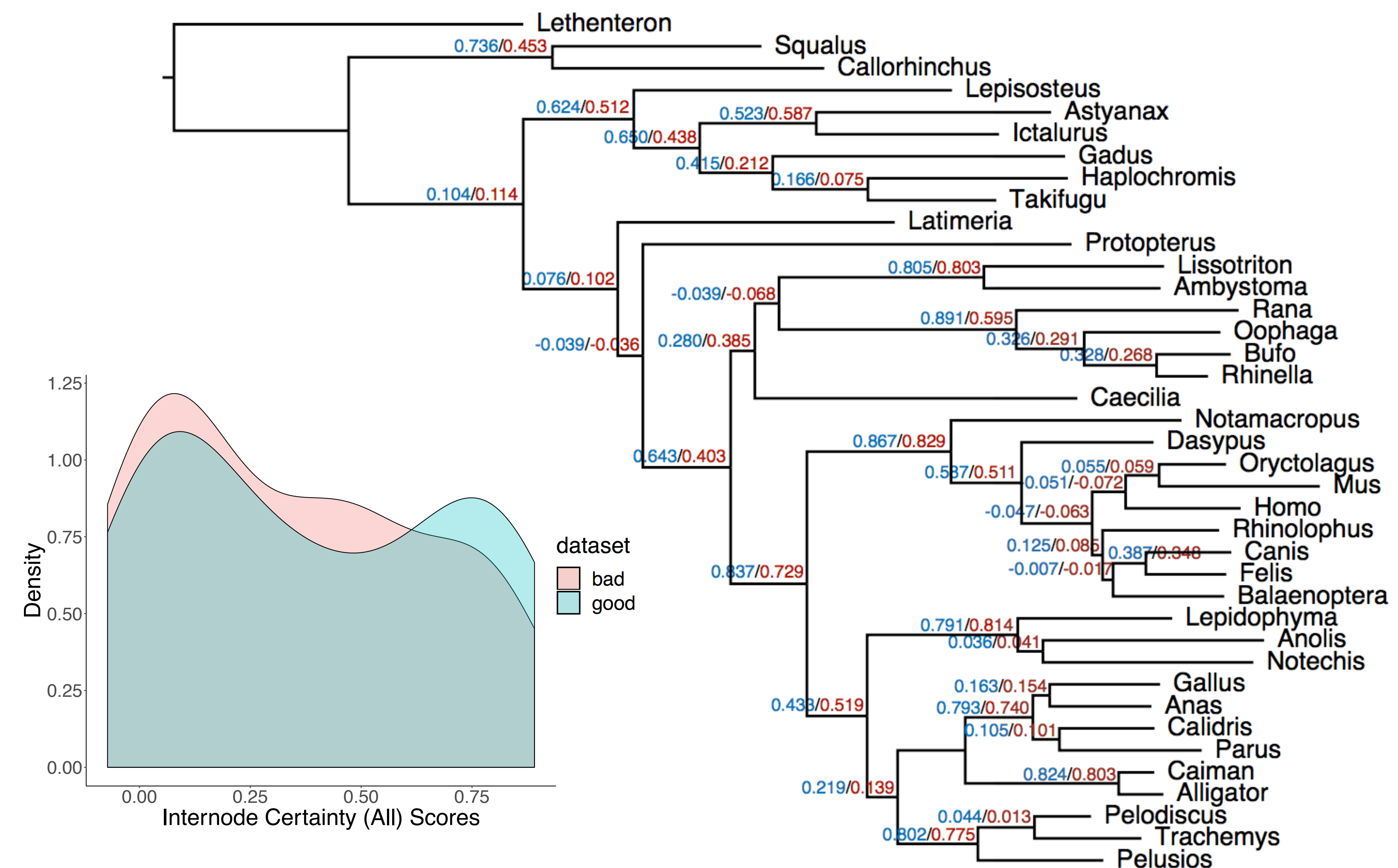
The workflow from publicly available transcriptome data to phylogenomic tree statistics.



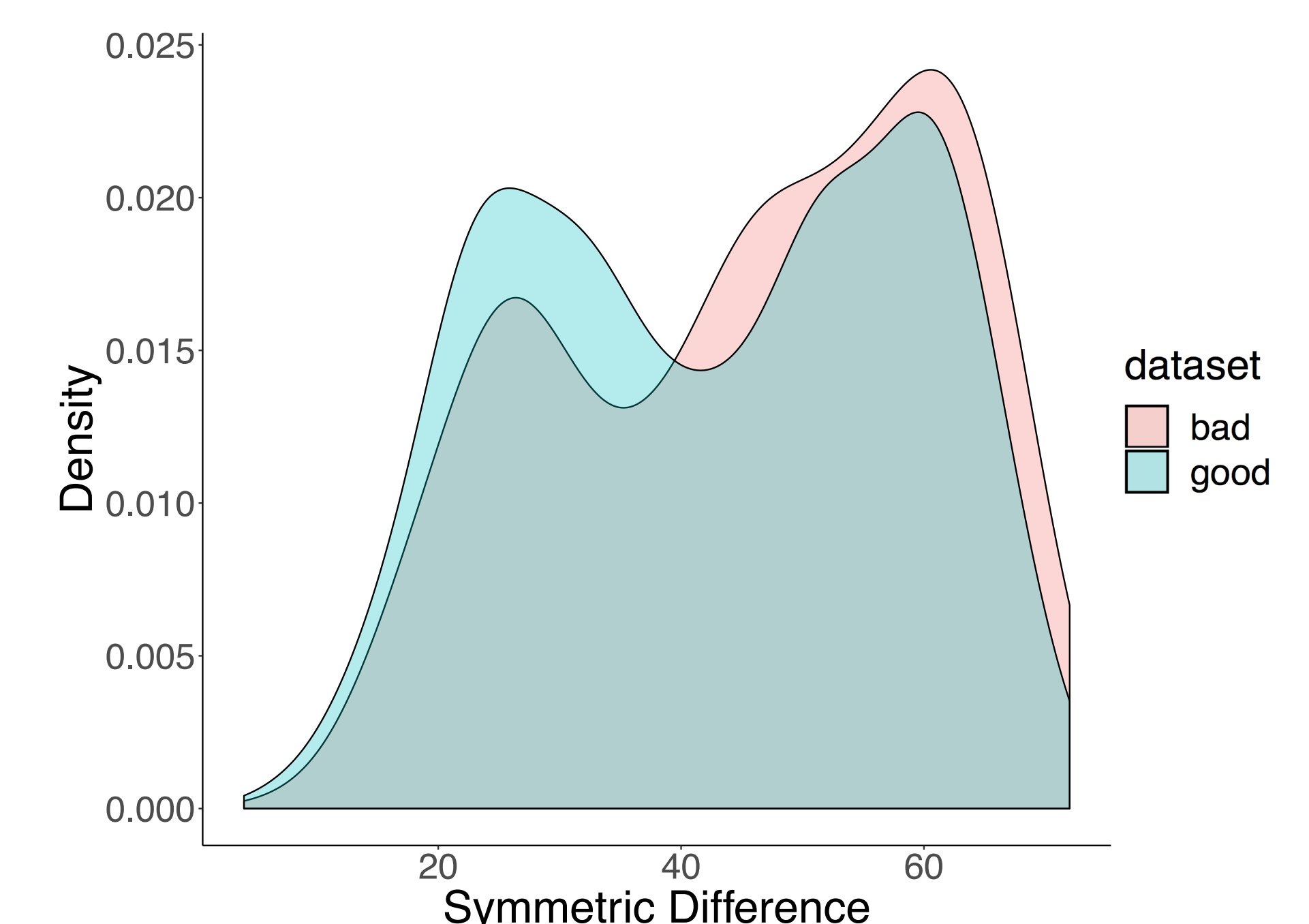
Number of partitions recovered through phylogenetic data assembly pipelines is fivefold higher when high quality transcripts are used, though their lengths both before and after Gblocks trimming are similar.



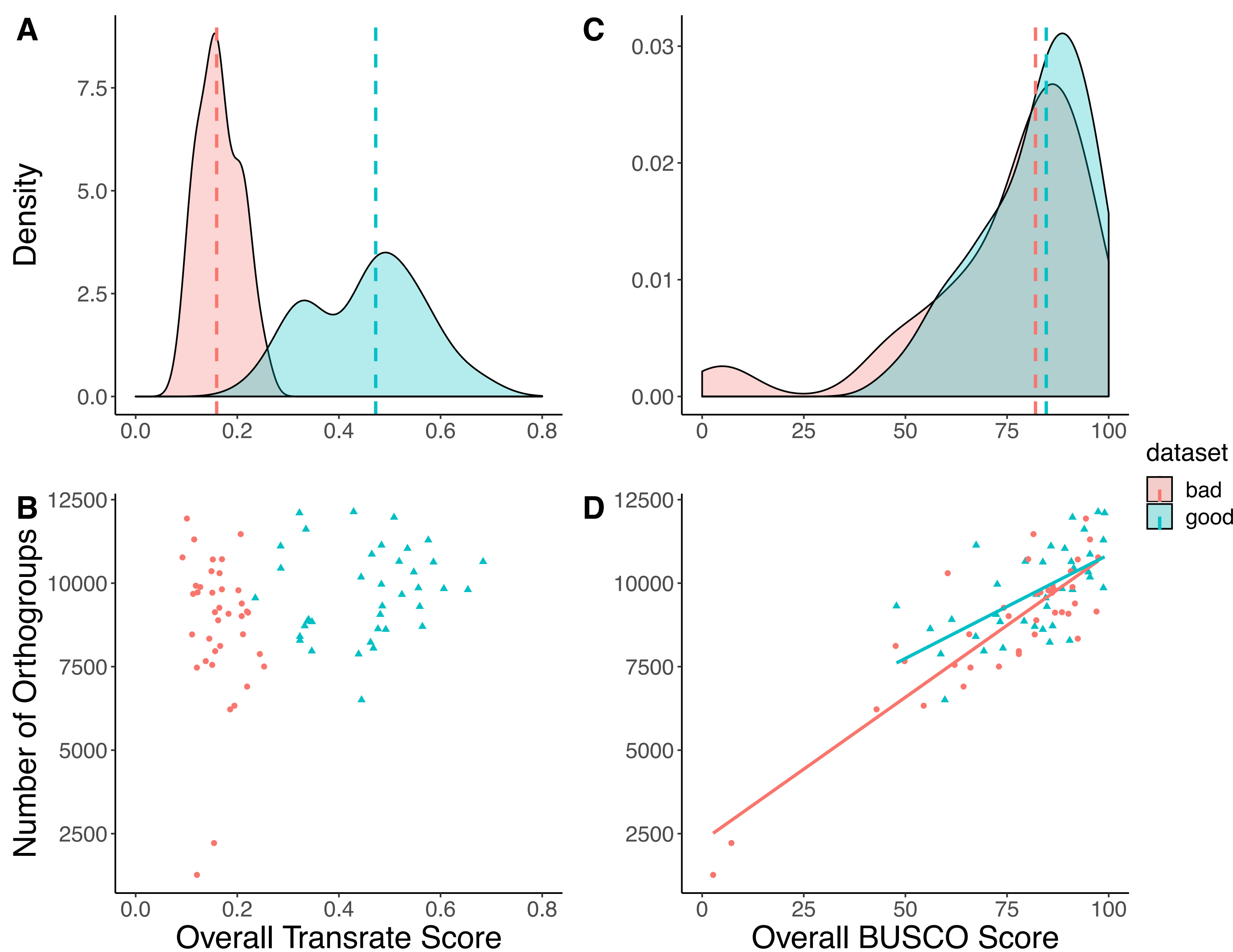
Partitions derived from the good dataset have higher internode certainty (all) values than those derived from the bad dataset when compared to the constraint tree.



Per partition Robinson-Foulds distances to the constraint tree are shorter in the good dataset compared with the bad.



Good and bad datasets were selected from five assemblies (in purple, above) based on Transrate score resulting in assemblies with both high and low completeness.



Phylogenomic datasets from high quality transcriptome assemblies are larger and more statistically consistent.