

Benchmarking computational efficiency of TreeSE methods

Giulio Benedetti and Leo Lahti

2021-11-25

Overview

This report contains benchmarking results for the melt operation with common microbiome data containers in R/Bioconductor.

The benchmarking tests utilize the following publicly available data sets. We thank the original authors for making these valuable data resources openly available. Check the links for details and original references:

- hitchip1006 (Lahti et al. 2014)
- SongQADData Song et al. (2016)
- GrieneisenTSDData Grieneisen et al. (2021) baboon data set

Data characteristics

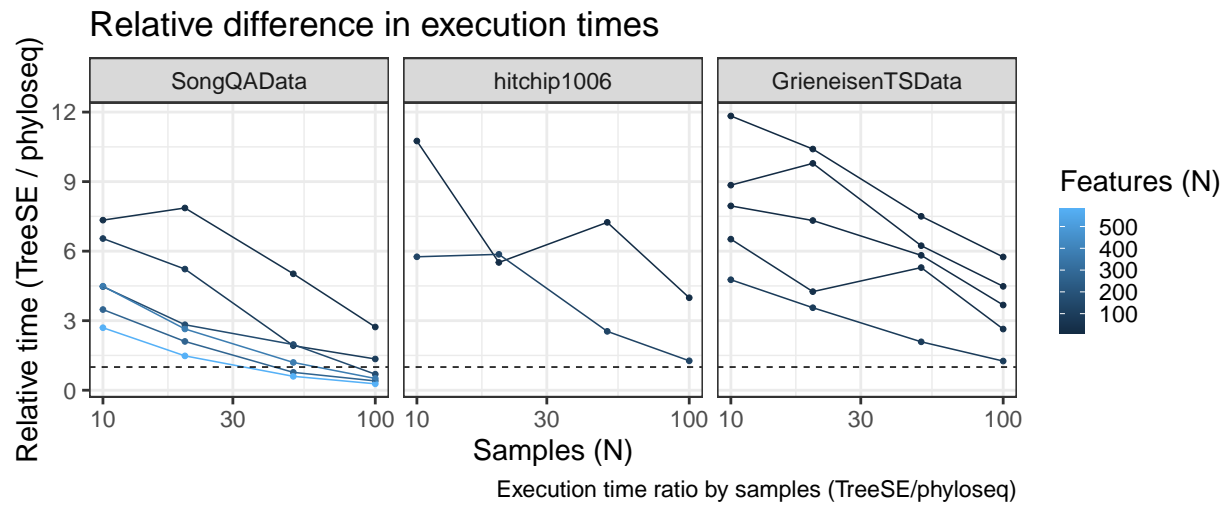
Full sample sizes by data set:

Dataset	N
hitchip1006	1151
SongQADData	1522
GrieneisenTSDData	16234

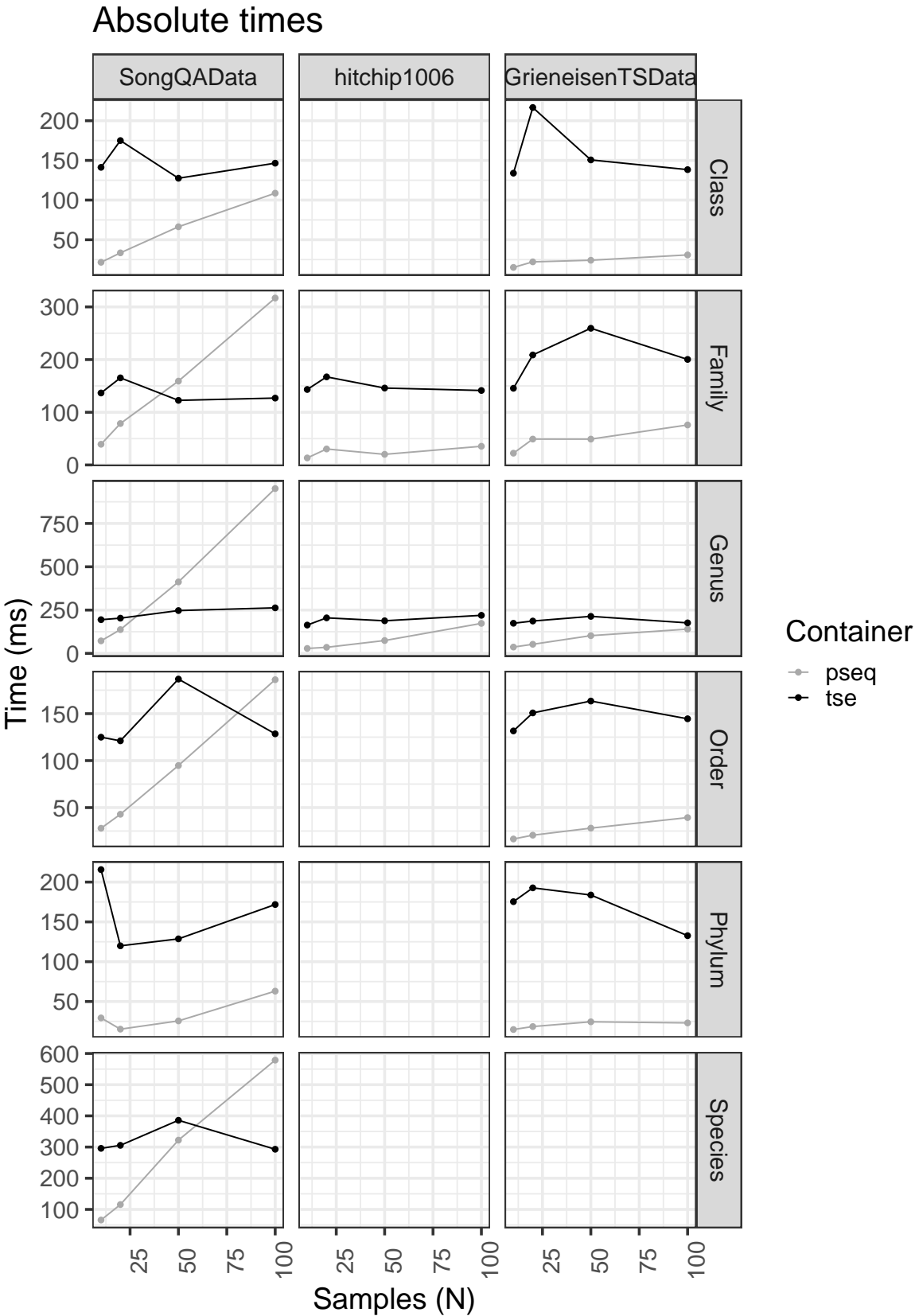
Feature counts by data set:

Rank	SongQADData	hitchip1006	GrieneisenTSDData
Phylum	39	0	12
Class	92	0	18
Order	167	0	24
Family	271	22	40
Species	375	0	0
Genus	583	130	92

Relative differences in execution time by sample size

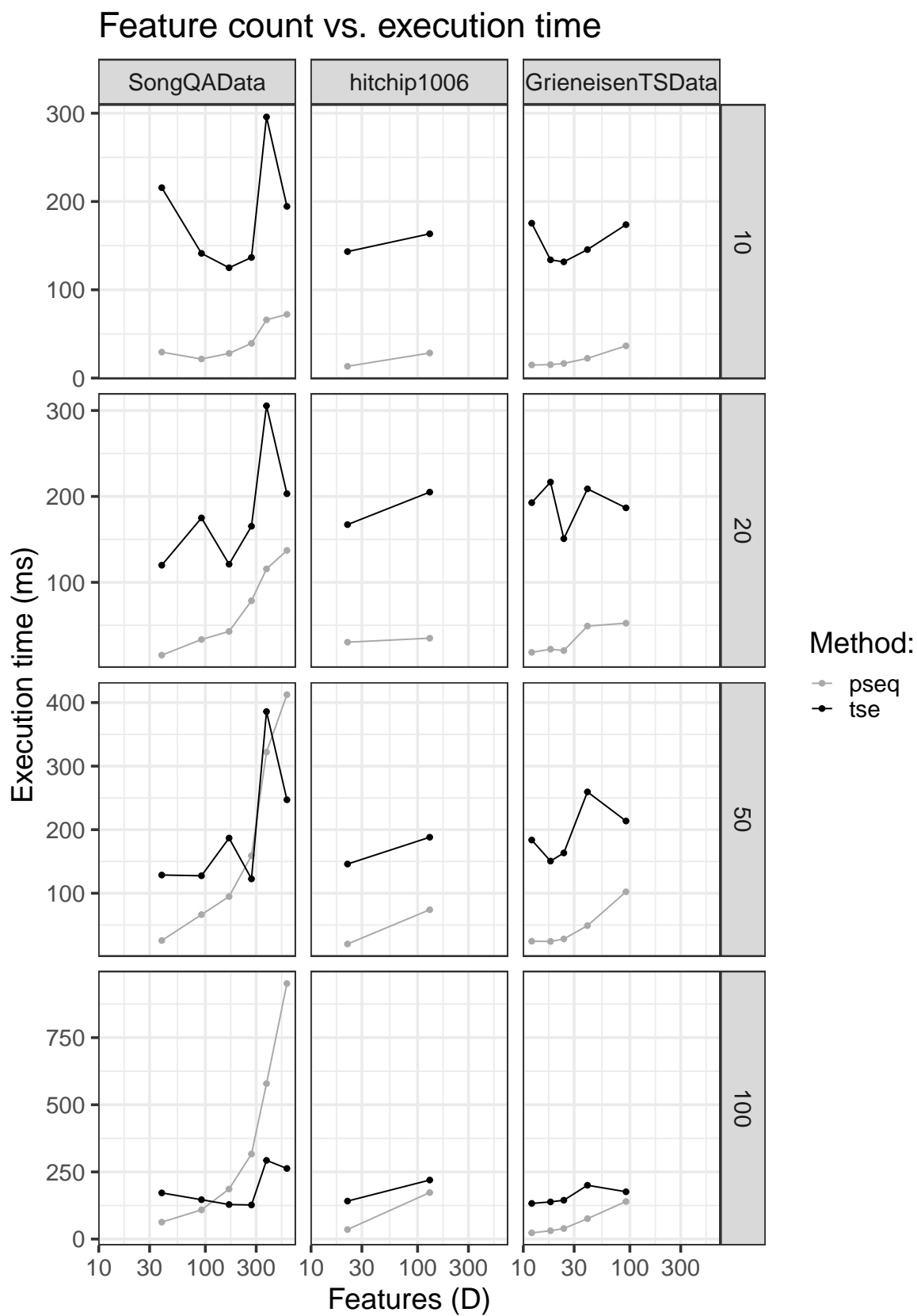


Absolute execution time by sample size



Execution time comparison between TreeSE and phyloseq.

Execution times vs number of features



Execution time as a function of number of features