

melt

Benchmarking computational efficiency of microbiome data containers

Giulio Benedetti and Leo Lahti

2021-12-10

Overview

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

The 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

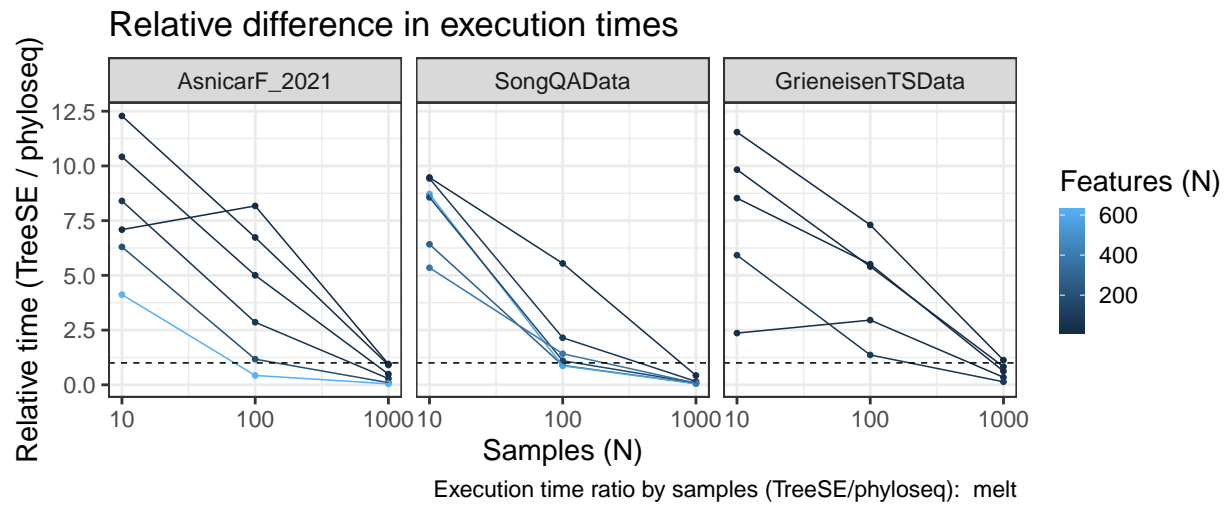
Full sample sizes by data set:

Dataset	N
AsnicarF_2021	1098
SongQADData	1522
GrieneisenTSDData	16234

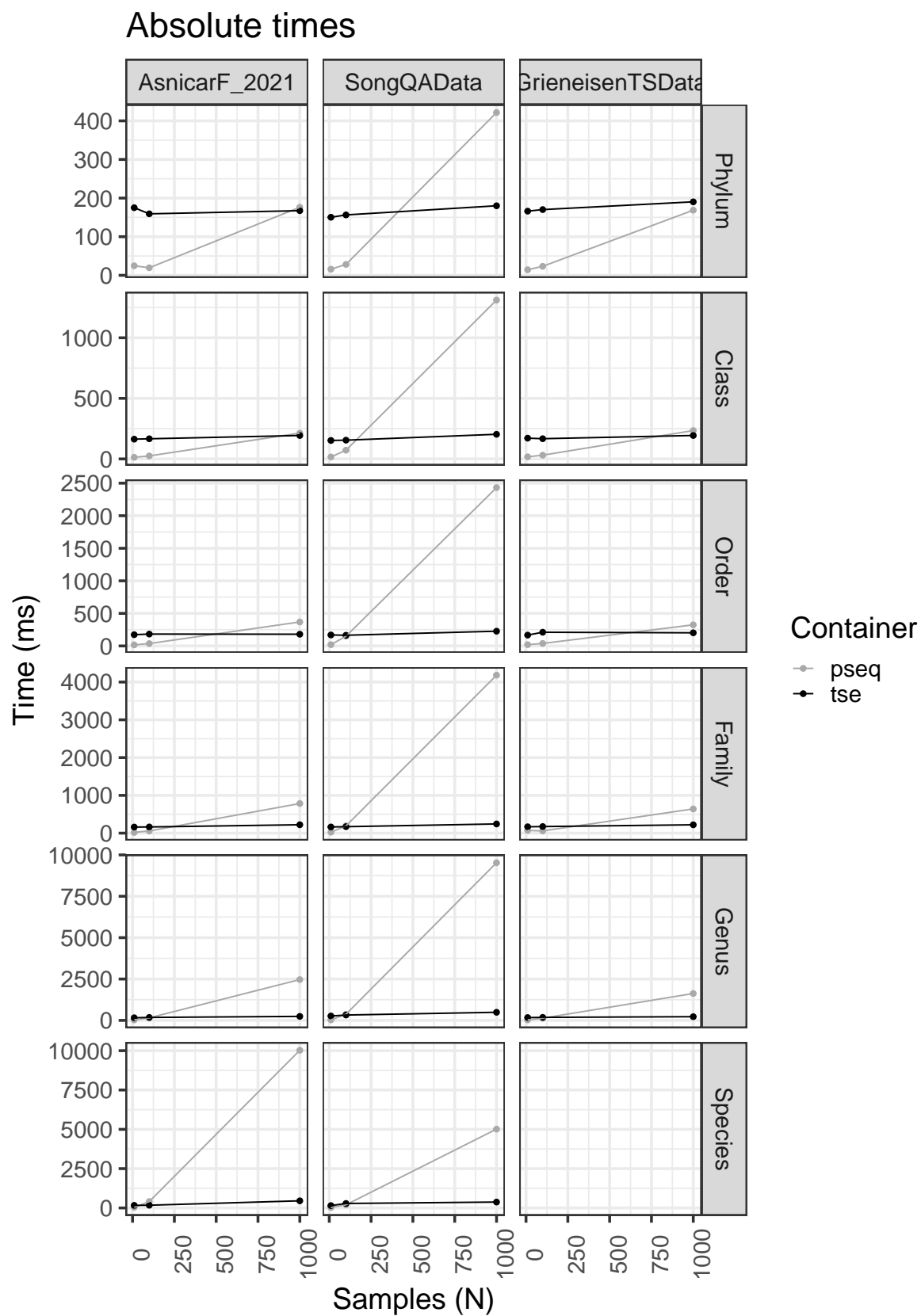
Feature counts by data set:

Rank	AsnicarF_2021	SongQADData	GrieneisenTSDData
Phylum	16	39	12
Class	23	92	18
Order	40	167	24
Family	74	271	40
Genus	205	583	92
Species	633	375	0

Relative differences in execution time by sample size

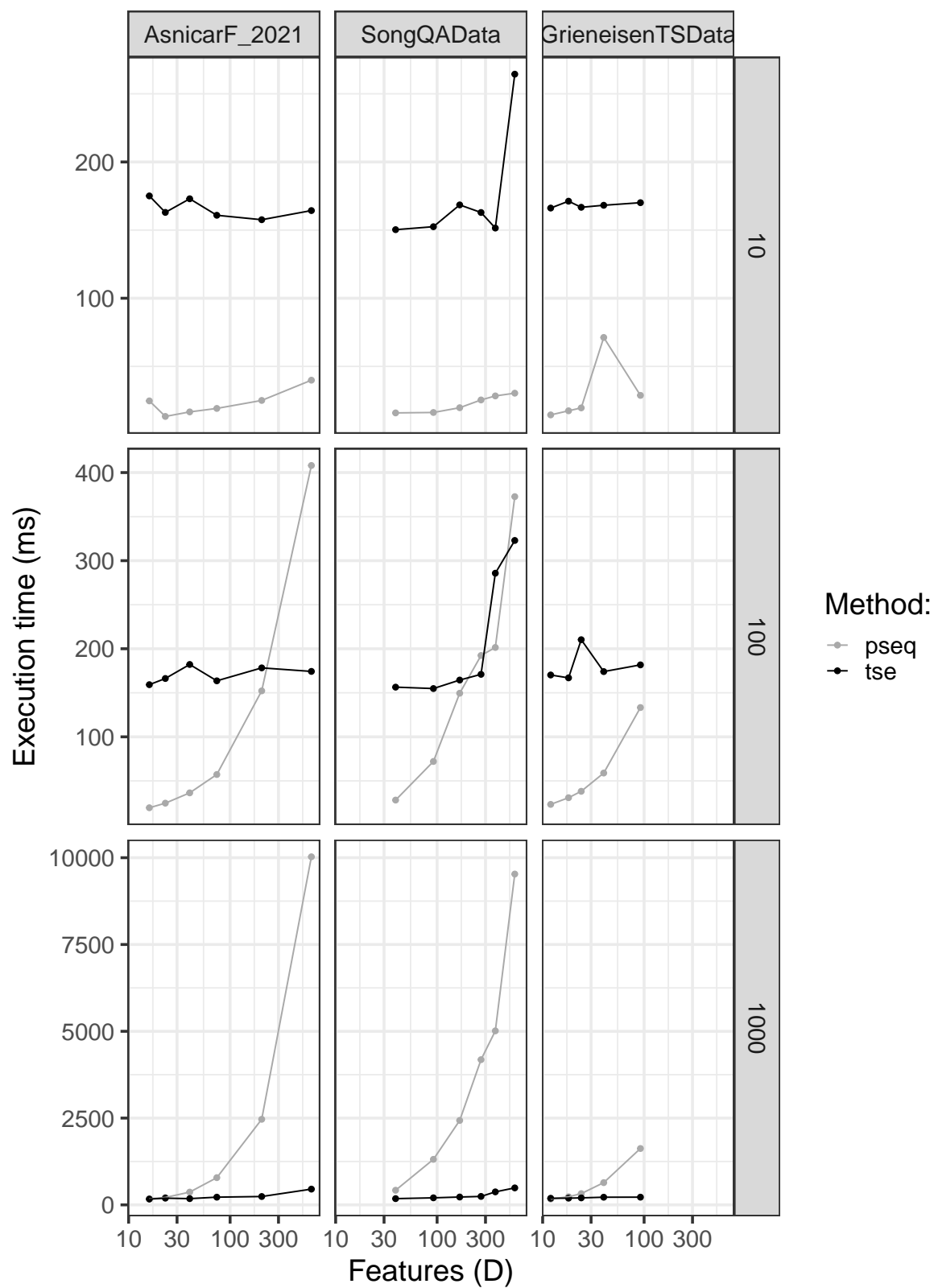


Absolute execution time by sample size



Execution times vs number of features

Feature count vs. execution time



Execution time by feature count: melt