

Benchmark

Klaus Schliep

January 21, 2020

A small benchmark study

In this vignette we will benchmark several common functions for several packages computing phylogenetic diversity (Faith 1992), beta diversity and phylogenetic beta diversity.

Lets load first a all the packages we are going to use.

```
library(ape)
library(Matrix)
library(bench)
library(ggplot2)
# packages we benchmark
library(phyloregion)

##
## Attaching package: 'phyloregion'

## The following objects are masked from 'package:base':
##
##      colSums, rowSums

library(betapart)
library(picante)

## Loading required package: vegan
## Loading required package: permute
## Loading required package: lattice
## This is vegan 2.5-6
## Loading required package: nlme

library(vegan)
library(hilldiv)

## Registered S3 methods overwritten by 'huge':
##   method      from
##   plot.sim    BDgraph
##   print.sim   BDgraph

## Registered S3 method overwritten by 'geiger':
##   method      from
##   unique.multiPhylo ape

library(BAT)

##
```

```
## Attaching package: 'BAT'

## The following object is masked from 'package:betapart':
##
##      beta.multi

## The following object is masked from 'package:ggplot2':
##
##      alpha

## The following object is masked from 'package:base':
##
##      beta
```

We will use a small data set which comes with `phyloregion`.

```
data(africa)
# subset matrix
X_sparse <- africa$comm[1:30, ]
X_sparse <- X_sparse[, colSums(X_sparse)>0]
X_dense <- as.matrix(X_sparse)
Xt_dense <- t(X_dense)
```

```
object.size(X_sparse)
```

```
## 76504 bytes
```

```
object.size(X_dense)
```

```
## 134752 bytes
```

```
dim(X_sparse)
```

```
## [1] 30 401
```

To compare results and start the analysis often we need to reformat the input or results. E.g. for the `hilldiv` package the community matrix needs to be transposed. These transformations influence the timings usually only marginal. For the benchmarking we use the package `bench` as it returns timings and provides some information about memory allocations.

Beta diversity

For beta diversity there is only a small advantage:

```
chk_fun <- function(target, current)
  all.equal(target, current, check.attributes = FALSE)

fun_phyloregion <- function(x) as.matrix(phyloregion::beta_diss(x)[[3]])
fun_betapart <- function(x) as.matrix(betapart::beta.pair(x)[[3]])
fun_vegan <- function(x) as.matrix(vegan::vegdist(x, binary=TRUE))
fun_BAT <- function(x) as.matrix(BAT::beta(x, func = "Soerensen")[1])
results <- bench::mark(fun_phyloregion(X_sparse), fun_betapart(X_dense),
  fun_vegan(X_dense), fun_BAT(X_dense), check=chk_fun)
summary(results)
```

```
## Registered S3 method overwritten by 'cli':
```

```
##   method      from
```

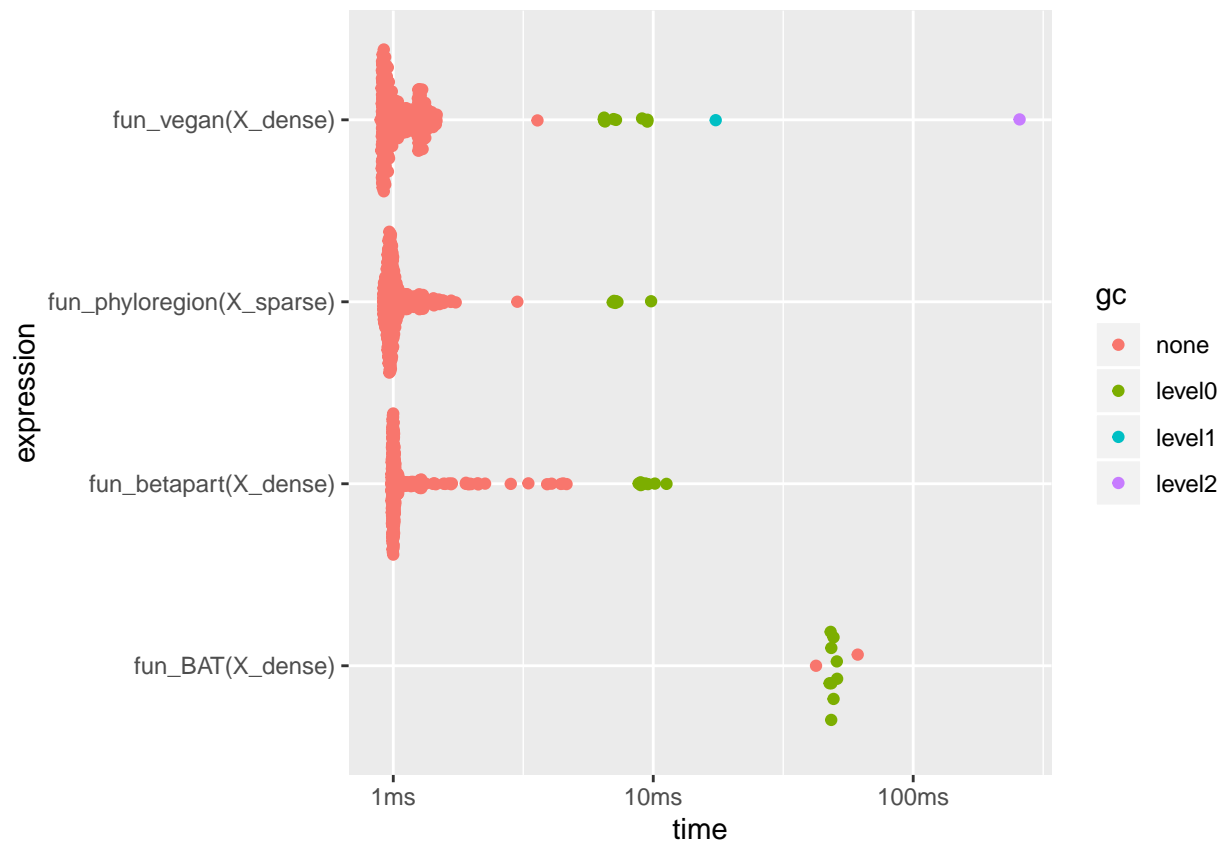
```
##   print.boxx spatstat
```

```
## # A tibble: 4 x 6
```

```
##   expression          min   median `itr/sec` mem_alloc `gc/sec`
##   <bch:expr>          <bch:tm> <bch:tm>    <dbl>  <bch:byt>    <dbl>
## 1 fun_phyloregion(X_sparse) 913.6us 987.19us   951.    620.52KB    13.2
## 2 fun_betapart(X_dense)    980.8us 1.01ms    887.    601.23KB    24.4
## 3 fun_vegan(X_dense)      893.8us 982.46us   931.     1.02MB    47.4
## 4 fun_BAT(X_dense)        42.3ms 51.74ms    19.3    31.76MB    87.0
```

```
autoplot(results)
```

```
## Loading required namespace: tidyr
```



Phylogenetic diversity (Faith 1992)

There is a large difference in memory allocations and running time for Faith's phylogenetic diversity.

```
tree <- africa$phylo
tree <- keep.tip(tree, colnames(X_sparse))

pd_picante <- function(x, tree){
  res <- picante::pd(x, tree)[,1]
  names(res) <- row.names(x)
  res
}

pd_hilldiv <- function(x, tree) hilldiv::index_div(x, tree, index="faith")
pd_phyloregion <- function(x, tree) phyloregion::PD(x, tree)

results <- bench::mark(pd_picante(X_dense, tree),
  pd_hilldiv(Xt_dense, tree=tree),
  pd_phyloregion(X_sparse, tree))
```

```
## Warning: Some expressions had a GC in every iteration; so filtering is disabled.
```

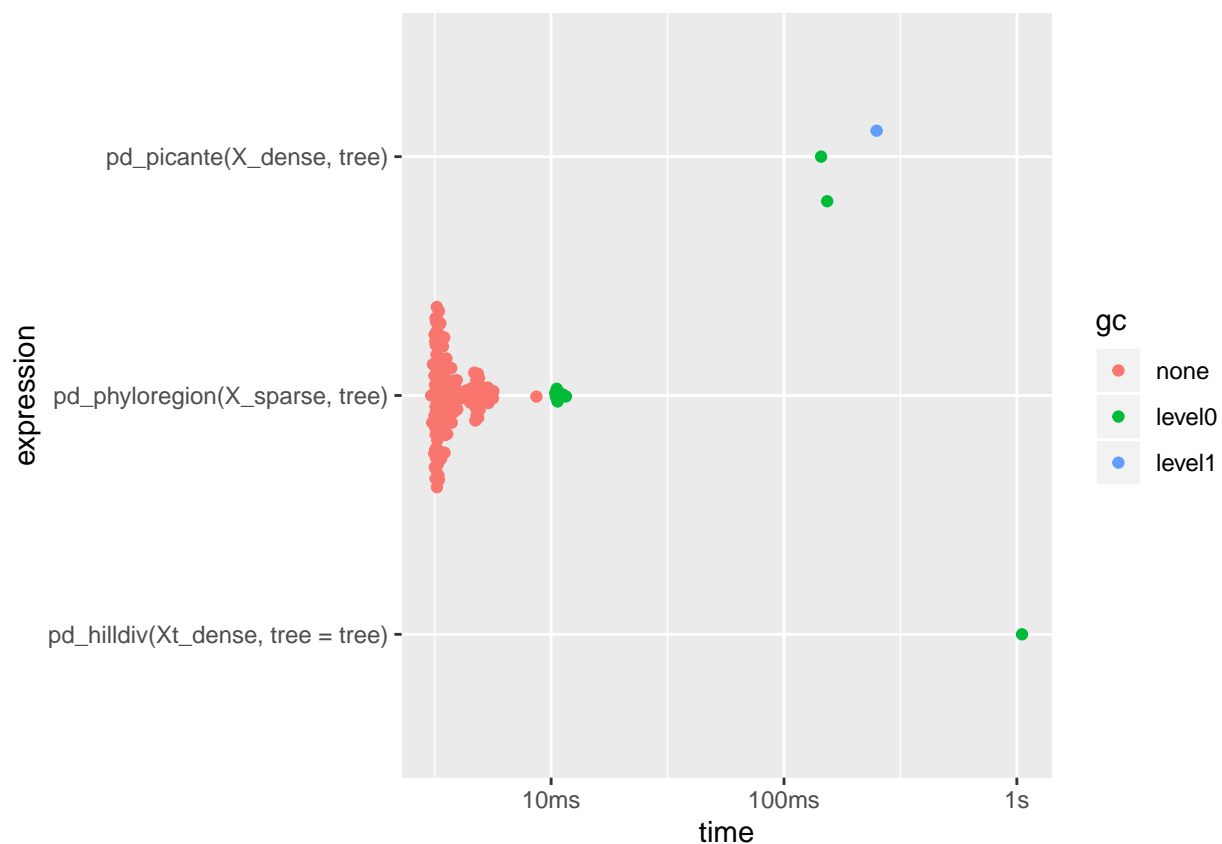
```
summary(results)
```

```
## Warning: Some expressions had a GC in every iteration; so filtering is disabled.
```

```
## # A tibble: 3 x 6
```

##	expression	min	median	`itr/sec`	mem_alloc
##	<bch:expr>	<bch:tm>	<bch:tm>	<dbl>	<bch:byt>
## 1	pd_picante(X_dense, tree)	144.35ms	153.3ms	5.48	59.5MB
## 2	pd_hilldiv(Xt_dense, tree = tree)	1.05s	1.05s	0.950	170.22MB
## 3	pd_phyloregion(X_sparse, tree)	3.06ms	3.47ms	243.	1.78MB
## #	... with 1 more variable: `gc/sec` <dbl>				

```
autoplot(results)
```



Phylogenetic beta diversity

Again there is a large difference in memory allocations and running time for the phylogenetic beta diversity.

```
fun_phyloregion <- function(x, tree) phylobeta(x, tree)[[3]]
fun_betapart <- function(x, tree) phylo.beta.pair(x, tree)[[3]]
fun_picante <- function(x, tree) 1 - phylosor(x, tree)
fun_BAT <- function(x, tree) beta(x, tree, func = "Soerensen")[[1]]

chk_fun <- function(target, current)
  all.equal(target, current, check.attributes = FALSE)
```

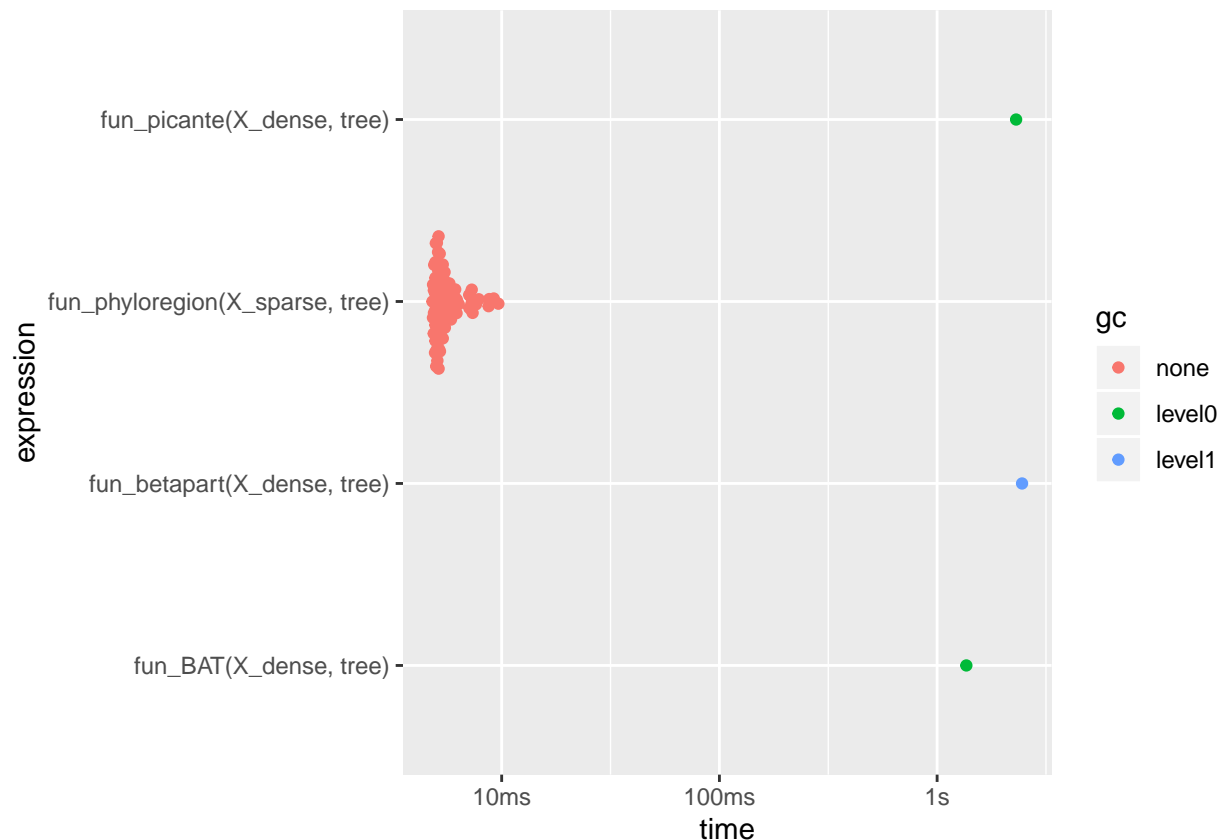
```

results <- bench::mark(fun_picante(X_dense, tree),
                       fun_betapart(X_dense, tree),
                       fun_BAT(X_dense, tree),
                       fun_phyloregion(X_sparse, tree), check=chk_fun)

## Warning: Some expressions had a GC in every iteration; so filtering is disabled.
summary(results)

## Warning: Some expressions had a GC in every iteration; so filtering is disabled.
## # A tibble: 4 x 6
##   expression          min   median `itr/sec` mem_alloc `gc/sec`
##   <bch:expr>      <bch:tm> <bch:tm>   <dbl>   <bch:byt>   <dbl>
## 1 fun_picante(X_dense, tree)    2.31s    2.31s     0.434    1.24GB     1.73
## 2 fun_betapart(X_dense, tree)    2.46s    2.46s     0.407    1.24GB     2.04
## 3 fun_BAT(X_dense, tree)        1.36s    1.36s     0.734   207.39MB     0.734
## 4 fun_phyloregion(X_sparse, tree) 4.81ms    5.28ms    174.      1.19MB      0
autoplot(results)

```



Note that the picante function returns a similarity matrix while the function in betapart and phyloregion return a dissimilarity matrix.

Session Infomation

```

sessionInfo()

## R version 3.6.2 (2019-12-12)

```

```

## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Debian GNU/Linux 10 (buster)
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/openblas/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/libopenblas-p-r0.3.5.so
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8      LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
##  [1] BAT_1.6.0          hilldiv_1.5.1      picante_1.8        nlme_3.1-143
##  [5] vegan_2.5-6        lattice_0.20-38    permute_0.9-5      betapart_1.5.1
##  [9] phyloregion_0.1.0  ggplot2_3.2.1      bench_1.1.1.9000   Matrix_1.2-18
## [13] ape_5.3
##
## loaded via a namespace (and not attached):
##  [1] backports_1.1.5      Hmisc_4.3-0        BDgraph_2.62
##  [4] fastmatch_1.1-0      plyr_1.8.5         igraph_1.2.4.2
##  [7] lazyeval_0.2.2       sp_1.3-2           splines_3.6.2
## [10] crosstalk_1.0.0      digest_0.6.23      htmltools_0.4.0
## [13] fansi_0.4.1          magrittr_1.5        checkmate_1.9.4
## [16] tensor_1.5           cluster_2.1.0       ks_1.11.6
## [19] fastcluster_1.1.25   pdist_1.2          prettyunits_1.1.0
## [22] jpeg_0.1-8.1         colorspace_1.4-1    xfun_0.12
## [25] dplyr_0.8.3          crayon_1.3.4        jsonlite_1.6
## [28] spatstat.data_1.4-0  spatstat_1.62-2     zeallot_0.1.0
## [31] survival_3.1-8       phangorn_2.6.0      glue_1.3.1
## [34] polyclip_1.10-0      gtable_0.3.0        geiger_2.0.6.2
## [37] webshot_0.5.2        maps_3.3.0          ggm_2.3
## [40] abind_1.4-5          scales_1.1.0        mvtnorm_1.0-12
## [43] miniUI_0.1.1.1       Rcpp_1.0.3          xtable_1.8-4
## [46] progress_1.2.2       htmlTable_1.13.3    magic_1.5-9
## [49] foreign_0.8-72       subplex_1.5-4       mclust_5.4.5
## [52] deSolve_1.27.1       Formula_1.2-3       stats4_3.6.2
## [55] htmlwidgets_1.5.1    RColorBrewer_1.1-2  lavaan_0.6-5
## [58] acepack_1.4.1         farver_2.0.3        pkgconfig_2.0.3
## [61] deldir_0.1-23        nnet_7.3-12         utf8_1.1.4
## [64] tidyselect_0.2.5     rlang_0.4.2         manipulateWidget_0.10.0
## [67] reshape2_1.4.3       later_1.0.0         munsell_0.5.0
## [70] tools_3.6.2          cli_2.0.1           fdrtool_1.2.15
## [73] evaluate_0.14        geometry_0.4.5      stringr_1.4.0
## [76] fastmap_1.0.1        goftest_1.2-2       yaml_2.2.0
## [79] knitr_1.27           rgl_0.100.30        purrr_0.3.3
## [82] glasso_1.11          pbapply_1.4-2       whisker_0.4
## [85] mime_0.8             profmem_0.5.0       compiler_3.6.2

```

## [88] rstudioapi_0.10	beeswarm_0.2.3	png_0.1-7
## [91] e1071_1.7-3	ggsignif_0.6.0	spatstat.utils_1.15-0
## [94] huge_1.3.4	tibble_2.1.3	pbivnorm_0.6.0
## [97] stringi_1.4.5	qgraph_1.6.4	rgeos_0.5-2
## [100] psych_1.9.12.31	vctr_0.2.1	pillar_1.4.3
## [103] lifecycle_0.1.0	data.table_1.12.8	corpcor_1.6.9
## [106] raster_3.0-7	httpuv_1.5.2	R6_2.4.1
## [109] latticeExtra_0.6-29	promises_1.1.0	KernSmooth_2.23-16
## [112] gridExtra_2.3	vipor_0.4.5	codetools_0.2-16
## [115] rcd_1.2-2	MASS_7.3-51.4	gtools_3.8.1
## [118] assertthat_0.2.1	rjson_0.2.20	withr_2.1.2
## [121] mnormt_1.5-5	mgcv_1.8-31	parallel_3.6.2
## [124] hms_0.5.3	quadprog_1.5-8	grid_3.6.2
## [127] rpart_4.1-15	tidyr_1.0.0	coda_0.19-3
## [130] class_7.3-15	nls2_0.2	rmarkdown_2.0
## [133] FSA_0.8.26	hypervolume_2.0.12	d3Network_0.5.2.1
## [136] ggpubr_0.2.4	shiny_1.4.0	base64enc_0.1-3
## [139] ggbeeswarm_0.6.0		