

Gene regulatory network inference

Fabricio Almeida-Silva

2023-05-02

```
library(here)
library(BioNERO)
library(igraph)
library(ggstatsplot)
library(ggpubr)
library(patchwork)

set.seed(123)
dup_palette <- c("#1984c5", "#ffb400")
```

Overview

Here, we will describe the code to infer gene regulatory networks (GRN) for each species from the RNA-seq data obtained from EBI's Expression Atlas.

GRN inference

Here, we will define a function to process the input for GRN inference with *BioNERO* and infer a GRN for each species manually (not in a loop) to avoid errors in one data set leading to loss of all previously inferred GRNs.

```
# Load expression data
load(here("data", "expression_data.rda"))

# Load TFs
load(here("data", "tfs.rda"))

# Define function to infer GRNs
get_GRN <- function(species) {

  # Get expression matrix for species x
  exp <- expression_data[[species]]
  exp <- BioNERO::exp_preprocess(exp, Zk_filtering = FALSE)
  exp <- as.data.frame(exp)

  # Get vector of TFs for species x
  vtf <- tfs[[species]]$Gene
  vtf <- vtf[vtf %in% rownames(exp)] # keep only TFs in expression matrix
  vtf <- unique(vtf)

  if(length(vtf) > 1) {
    message(species, ":", length(vtf), " TFs")
```

```

} else {
  stop("Found less than 2 TFs.")
}

# Infer GRN
grn <- BioNERO::grn_infer(
  exp, method = "genie3", regulators = vtfss, nTrees = 1000
)
return(grn)
}

# Infer GRNs
grn_osa <- get_GRN("Osativa")
grn_zma <- get_GRN("Zmays")
grn_gma <- get_GRN("Gmax")
grn_sly <- get_GRN("Slycopersicum")
grn_ptr <- get_GRN("Ptrichocarpa")
grn_ath <- get_GRN("Athaliana")
grn_vvi <- get_GRN("Vvinifera")

```

We now have our GRNs as fully connected graphs. Let's filter them by picking only the top N edges (based on edge rankings). To decide N for each species, we will break each GRN into 20 increasingly larger graphs and assess the scale-free topology fit for each graph. Then, we will pick the maximum value of N that makes the graph fit a scale-free topology ($R^2 = 0.75$).

```

# Filter network based on scale-free topology fit
## Osativa
png(
  filename = here("products", "plots", "grn_filtering_osativa.png"),
  width = 8, height = 6, units = "in",
  res = 300
)
grn_filtered_osa <- grn_filter(grn_osa, nsplit = 20)
dev.off()

grn_filtered_osa <- grn_osa[1:1799884, ]

## Zmays
png(
  filename = here("products", "plots", "grn_filtering_zmays.png"),
  width = 8, height = 6, units = "in",
  res = 300
)
grn_filtered_zma <- grn_filter(grn_zma, nsplit = 20)
dev.off()

grn_filtered_zma <- grn_zma[1:2045899, ]

## Vvinifera
png(
  filename = here("products", "plots", "grn_filtering_vvinifera.png"),
  width = 8, height = 6, units = "in",
  res = 300
)
```

```

)
grn_filtered_vvi <- grn_filter(grn_vvi, nsplit = 20)
dev.off()

grn_filtered_vvi <- grn_vvi[1:1214095, ]

## Gmax
png(
  filename = here("products", "plots", "grn_filtering_gmax.png"),
  width = 8, height = 6, units = "in",
  res = 300
)
grn_filtered_gma <- grn_filter(grn_gma, nsplit = 20)
dev.off()

grn_filtered_gma <- grn_gma[1:2268359, ]

## Slycopersicum
png(
  filename = here("products", "plots", "grn_filtering_slycopersicum.png"),
  width = 8, height = 6, units = "in",
  res = 300
)
grn_filtered_sly <- grn_filter(grn_sly, nsplit = 20)
dev.off()

grn_filtered_sly <- grn_sly[1:1344344, ]


## Ptrichocarpa
png(
  filename = here("products", "plots", "grn_filtering_ptrichocarpa.png"),
  width = 8, height = 6, units = "in",
  res = 300
)
grn_filtered_ptr <- grn_filter(grn_ptr, nsplit = 20)
dev.off()

grn_filtered_ptr <- grn_ptr[1:1275872, ]


## Athaliana
png(
  filename = here("products", "plots", "grn_filtering_athaliana.png"),
  width = 8, height = 6, units = "in",
  res = 300
)
grn_filtered_ath <- grn_filter(grn_ath, nsplit = 20)
dev.off()

grn_filtered_ath <- grn_ath[1:2503623, ]

# Save filtered GRNs to an .rda file

```

```

grns <- list(
  Osativa = grn_filtered_osa,
  Zmays = grn_filtered_zma,
  Vvinifera = grn_filtered_vvi,
  Gmax = grn_filtered_gma,
  Slycopersicum = grn_filtered_sly,
  Ptrichocarpa = grn_filtered_ptr,
  Athaliana = grn_filtered_ath
)
grns <- lapply(grns, function(x) return(x[, 1:2])) # remove 'Weight' column

save(
  grns,
  file = here("products", "result_files", "grns.rda"),
  compress = "xz"
)

```

To summarize, the numbers of top N edges for each species were:

Species	Edge number (Millions)
Osativa	1.799884
Zmays	2.045899
Vvinifera	1.214095
Gmax	2.268359
Slycopersicum	1.344344
Ptrichocarpa	1.275872
Athaliana	2.503623

Exploring degree distributions

Finally, let's compare the degree distributions of WGD- and SSD-derived gene pairs.

```

load(here("data", "duplicated_genes.rda"))
load(here("products", "result_files", "grns.rda"))

# Create data frame of degree distribution for each species by duplication mode
degree_distros <- Reduce(rbind, lapply(seq_along(grns), function(x) {
  species <- names(grns)[x]
  net <- grns[[x]]
  dups <- duplicated_genes[[species]]

  # Get degree distribution as a data frame
  degree <- graph_from_data_frame(net[, 1:2]) %>% degree()
  degree_df <- data.frame(
    gene = names(degree),
    degree = degree
  ) %>% dplyr::inner_join(., dups, by = "gene") %>%
    dplyr::select(gene, degree, type, peak) %>%
    dplyr::mutate(species = species)

  return(degree_df)
}))
```

```

# Visualize degree distributions as a violin plot
plot_violin <- function(data) {

  p <- ggbeeswarm::ggbeeswarm(
    data = data, x = type, y = degree,
    type = "nonparametric", pairwise.display = "all",
    p.adjust.method = "BH"
  ) +
    ggplot2::scale_color_manual(values = dup_palette) +
    labs(x = "", y = "") +
    theme(plot.subtitle = element_text(size = 11))

  sub <- extract_subtitle(p)
  sub[c(2, 5, 6)] <- NULL

  p <- p + labs(subtitle = sub)

  return(p)
}

sdegree_distros <- split(degree_distros, degree_distros$species)

plot_degree_gma <- plot_violin(sdegree_distros$Gmax) +
  labs(title = "Glycine max")

plot_degree_ath <- plot_violin(sdegree_distros$Athaliana) +
  labs(title = "Arabidopsis thaliana")

plot_degree_ptr <- plot_violin(sdegree_distros$Ptrichocarpa) +
  labs(title = "Populus trichocarpa")

plot_degree_sly <- plot_violin(sdegree_distros$Slycopersicum) +
  labs(title = "Solanum lycopersicum")

plot_degree_vvi <- plot_violin(sdegree_distros$Vvinifera) +
  labs(title = "Vitis vinifera")

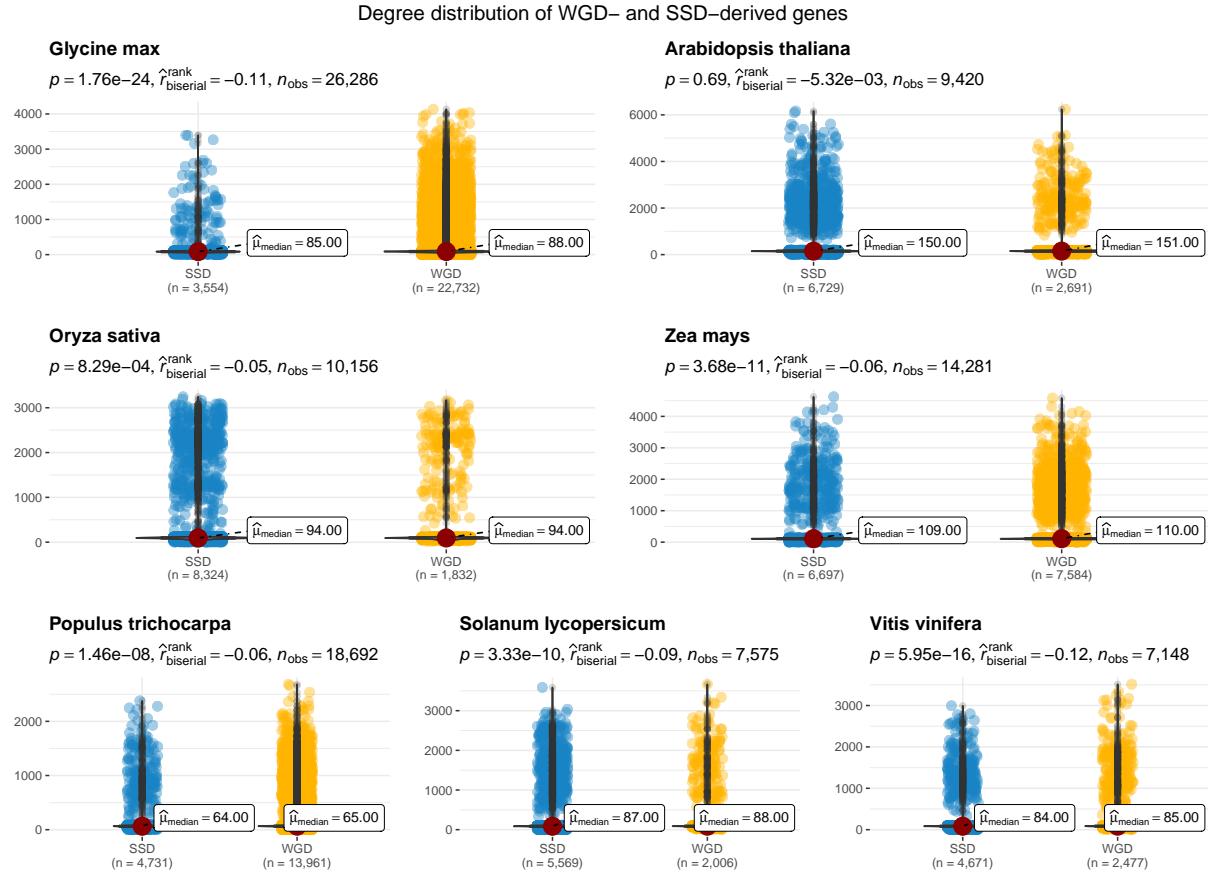
plot_degree_osa <- plot_violin(sdegree_distros$Osativa) +
  labs(title = "Oryza sativa")

plot_degree_zma <- plot_violin(sdegree_distros$Zmays) +
  labs(title = "Zea mays")

# Combine plots
p_deg_final <- wrap_plots(
  wrap_plots(plot_degree_gma, plot_degree_ath, nrow = 1),
  wrap_plots(plot_degree_osa, plot_degree_zma, nrow = 1),
  wrap_plots(plot_degree_ptr, plot_degree_sly, plot_degree_vvi, nrow = 1),
  nrow = 3
) +
  plot_annotation(
    "Degree distribution of WGD- and SSD-derived genes",
    theme = theme(plot.title = element_text(hjust = 0.5))
)

```

p_deg_final



As in the PPI network, although the P-values for some comparisons were significant, the effect size is negligible, indicating that degree does not influence motif frequencies. The small P-value is likely an artifact generated by the large sample size of the comparisons.

```
# Saving figure and degree distros object
## Degree distros
save(
  degree_distros,
  file = here("products", "result_files", "degree_distros_grn.rda"),
  compress = "xz"
)

## Plot
p_degree_distros_grn <- p_deg_final

save(
  p_degree_distros_grn,
  file = here("products", "plots", "p_degree_distros_grn.rda"),
  compress = "xz"
)
```

Session info

This document was created under the following conditions:

```
sessioninfo::session_info()

## - Session info -----
## setting value
## version R version 4.3.0 (2023-04-21)
## os       Ubuntu 20.04.5 LTS
## system  x86_64, linux-gnu
## ui       X11
## language (EN)
## collate en_US.UTF-8
## ctype    en_US.UTF-8
## tz       Europe/Brussels
## date     2023-05-02
## pandoc   2.19.2 @ /usr/lib/rstudio/resources/app/bin/quarto/bin/tools/ (via rmarkdown)
##
## - Packages -----
## package      * version  date (UTC) lib source
## abind        1.4-5    2016-07-21 [1] CRAN (R 4.3.0)
## annotate     1.78.0   2023-04-25 [1] Bioconductor
## AnnotationDbi 1.62.0   2023-04-25 [1] Bioconductor
## backports    1.4.1    2021-12-13 [1] CRAN (R 4.3.0)
## base64enc    0.1-3    2015-07-28 [1] CRAN (R 4.3.0)
## bayestestR   0.13.1   2023-04-07 [1] CRAN (R 4.3.0)
## Biobase      2.60.0   2023-04-25 [1] Bioconductor
## BiocGenerics 0.46.0   2023-04-25 [1] Bioconductor
## BiocManager   1.30.20  2023-02-24 [1] CRAN (R 4.3.0)
## BiocParallel  1.34.0   2023-04-25 [1] Bioconductor
## BiocStyle     2.28.0   2023-04-25 [1] Bioconductor
## BioNERO       * 1.8.0    2023-04-25 [1] Bioconductor
## Biostings    2.68.0   2023-04-25 [1] Bioconductor
## bit          4.0.5    2022-11-15 [1] CRAN (R 4.3.0)
## bit64        4.0.5    2020-08-30 [1] CRAN (R 4.3.0)
## bitops       1.0-7    2021-04-24 [1] CRAN (R 4.3.0)
## blob         1.2.4    2023-03-17 [1] CRAN (R 4.3.0)
## broom        1.0.4    2023-03-11 [1] CRAN (R 4.3.0)
## cachem       1.0.8    2023-05-01 [1] CRAN (R 4.3.0)
## car          3.1-2    2023-03-30 [1] CRAN (R 4.3.0)
## carData      3.0-5    2022-01-06 [1] CRAN (R 4.3.0)
## checkmate    2.2.0    2023-04-27 [1] CRAN (R 4.3.0)
## circlize     0.4.15   2022-05-10 [1] CRAN (R 4.3.0)
## cli          3.6.1    2023-03-23 [1] CRAN (R 4.3.0)
## clue         0.3-64   2023-01-31 [1] CRAN (R 4.3.0)
## cluster      2.1.4    2022-08-22 [4] CRAN (R 4.2.1)
## coda         0.19-4   2020-09-30 [1] CRAN (R 4.3.0)
## codetools    0.2-19   2023-02-01 [4] CRAN (R 4.2.2)
## colorspace   2.1-0    2023-01-23 [1] CRAN (R 4.3.0)
## ComplexHeatmap 2.16.0   2023-04-25 [1] Bioconductor
## correlation  0.8.4    2023-04-06 [1] CRAN (R 4.3.0)
## crayon       1.5.2    2022-09-29 [1] CRAN (R 4.3.0)
## data.table   1.14.8   2023-02-17 [1] CRAN (R 4.3.0)
## datawizard   0.7.1    2023-04-03 [1] CRAN (R 4.3.0)
```

```

##  DBI           1.1.3    2022-06-18 [1] CRAN (R 4.3.0)
##  DelayedArray  0.25.0   2022-11-01 [1] Bioconductor
##  digest        0.6.31   2022-12-11 [1] CRAN (R 4.3.0)
##  doParallel    1.0.17   2022-02-07 [1] CRAN (R 4.3.0)
##  dplyr         1.1.2    2023-04-20 [1] CRAN (R 4.3.0)
##  dynamicTreeCut 1.63-1   2016-03-11 [1] CRAN (R 4.3.0)
##  edgeR          3.42.0   2023-04-25 [1] Bioconductor
##  effectsize     0.8.3    2023-01-28 [1] CRAN (R 4.3.0)
##  evaluate       0.20     2023-01-17 [1] CRAN (R 4.3.0)
##  fansi          1.0.4    2023-01-22 [1] CRAN (R 4.3.0)
##  farver         2.1.1    2022-07-06 [1] CRAN (R 4.3.0)
##  fastcluster    1.2.3    2021-05-24 [1] CRAN (R 4.3.0)
##  fastmap        1.1.1    2023-02-24 [1] CRAN (R 4.3.0)
##  foreach        1.5.2    2022-02-02 [1] CRAN (R 4.3.0)
##  foreign        0.8-82   2022-01-13 [4] CRAN (R 4.1.2)
##  Formula        1.2-5    2023-02-24 [1] CRAN (R 4.3.0)
##  genefilter     1.82.0   2023-04-25 [1] Bioconductor
##  generics       0.1.3    2022-07-05 [1] CRAN (R 4.3.0)
##  GENIE3         1.22.0   2023-04-25 [1] Bioconductor
##  GenomeInfoDb   1.36.0   2023-04-25 [1] Bioconductor
##  GenomeInfoDbData 1.2.10   2023-04-28 [1] Bioconductor
##  GenomicRanges  1.52.0   2023-04-25 [1] Bioconductor
##  GetoptLong     1.0.5    2020-12-15 [1] CRAN (R 4.3.0)
##  ggnetwork      0.5.12   2023-03-06 [1] CRAN (R 4.3.0)
##  ggnewscale     0.4.8    2022-10-06 [1] CRAN (R 4.3.0)
##  ggplot2        * 3.4.2   2023-04-03 [1] CRAN (R 4.3.0)
##  ggpubr         * 0.6.0    2023-02-10 [1] CRAN (R 4.3.0)
##  ggrepel         0.9.3    2023-02-03 [1] CRAN (R 4.3.0)
##  ggsignif        0.6.4    2022-10-13 [1] CRAN (R 4.3.0)
##  ggstatsplot    * 0.11.1   2023-04-14 [1] CRAN (R 4.3.0)
##  GlobalOptions   0.1.2    2020-06-10 [1] CRAN (R 4.3.0)
##  glue            1.6.2    2022-02-24 [1] CRAN (R 4.3.0)
##  GO.db          3.17.0   2023-05-02 [1] Bioconductor
##  gridExtra       2.3      2017-09-09 [1] CRAN (R 4.3.0)
##  gtable          0.3.3    2023-03-21 [1] CRAN (R 4.3.0)
##  here            * 1.0.1   2020-12-13 [1] CRAN (R 4.3.0)
##  highr           0.10     2022-12-22 [1] CRAN (R 4.3.0)
##  Hmisc            5.0-1    2023-03-08 [1] CRAN (R 4.3.0)
##  htmlTable        2.4.1    2022-07-07 [1] CRAN (R 4.3.0)
##  htmltools        0.5.5    2023-03-23 [1] CRAN (R 4.3.0)
##  htmlwidgets     1.6.2    2023-03-17 [1] CRAN (R 4.3.0)
##  httr             1.4.5    2023-02-24 [1] CRAN (R 4.3.0)
##  igraph           * 1.4.2   2023-04-07 [1] CRAN (R 4.3.0)
##  impute          1.74.0   2023-04-25 [1] Bioconductor
##  insight          0.19.1   2023-03-18 [1] CRAN (R 4.3.0)
##  intergraph       2.0-2    2016-12-05 [1] CRAN (R 4.3.0)
##  IRanges          2.34.0   2023-04-25 [1] Bioconductor
##  iterators        1.0.14   2022-02-05 [1] CRAN (R 4.3.0)
##  KEGGREST         1.40.0   2023-04-25 [1] Bioconductor
##  knitr            1.42     2023-01-25 [1] CRAN (R 4.3.0)
##  labeling          0.4.2    2020-10-20 [1] CRAN (R 4.3.0)
##  lattice           0.20-45  2021-09-22 [4] CRAN (R 4.2.0)
##  lifecycle        1.0.3    2022-10-07 [1] CRAN (R 4.3.0)
##  limma            3.56.0   2023-04-25 [1] Bioconductor

```

```

##  locfit           1.5-9.7   2023-01-02 [1] CRAN (R 4.3.0)
##  magrittr        2.0.3    2022-03-30 [1] CRAN (R 4.3.0)
##  Matrix          1.5-1    2022-09-13 [4] CRAN (R 4.2.1)
##  MatrixGenerics 1.12.0   2023-04-25 [1] Bioconductor
##  matrixStats     0.63.0   2022-11-18 [1] CRAN (R 4.3.0)
##  memoise         2.0.1    2021-11-26 [1] CRAN (R 4.3.0)
##  mgcv            1.8-41   2022-10-21 [4] CRAN (R 4.2.1)
##  minet           3.58.0   2023-04-25 [1] Bioconductor
##  munsell         0.5.0    2018-06-12 [1] CRAN (R 4.3.0)
##  NetRep          1.2.6    2023-01-06 [1] CRAN (R 4.3.0)
##  network         1.18.1   2023-01-24 [1] CRAN (R 4.3.0)
##  networkD3       0.4      2017-03-18 [1] CRAN (R 4.3.0)
##  nlme            3.1-162  2023-01-31 [4] CRAN (R 4.2.2)
##  nnet             7.3-18   2022-09-28 [4] CRAN (R 4.2.1)
##  paletteer       1.5.0    2022-10-19 [1] CRAN (R 4.3.0)
##  parameters      0.21.0   2023-04-19 [1] CRAN (R 4.3.0)
##  patchwork       * 1.1.2   2022-08-19 [1] CRAN (R 4.3.0)
##  pillar           1.9.0    2023-03-22 [1] CRAN (R 4.3.0)
##  pkgconfig        2.0.3    2019-09-22 [1] CRAN (R 4.3.0)
##  plyr             1.8.8    2022-11-11 [1] CRAN (R 4.3.0)
##  png              0.1-8    2022-11-29 [1] CRAN (R 4.3.0)
##  preprocessCore   1.62.0   2023-04-25 [1] Bioconductor
##  purrr            1.0.1    2023-01-10 [1] CRAN (R 4.3.0)
##  R6                2.5.1    2021-08-19 [1] CRAN (R 4.3.0)
##  RColorBrewer     1.1-3    2022-04-03 [1] CRAN (R 4.3.0)
##  Rcpp              1.0.10   2023-01-22 [1] CRAN (R 4.3.0)
##  RCurl             1.98-1.12 2023-03-27 [1] CRAN (R 4.3.0)
##  rematch2          2.1.2    2020-05-01 [1] CRAN (R 4.3.0)
##  reshape2          1.4.4    2020-04-09 [1] CRAN (R 4.3.0)
##  RhpcBLASctl      0.23-42  2023-02-11 [1] CRAN (R 4.3.0)
##  rjson             0.2.21   2022-01-09 [1] CRAN (R 4.3.0)
##  rlang             1.1.1    2023-04-28 [1] CRAN (R 4.3.0)
##  rmarkdown         2.21     2023-03-26 [1] CRAN (R 4.3.0)
##  rpart             4.1.19   2022-10-21 [4] CRAN (R 4.2.1)
##  rprojroot         2.0.3    2022-04-02 [1] CRAN (R 4.3.0)
##  RSQLite            2.3.1    2023-04-03 [1] CRAN (R 4.3.0)
##  rstatix           0.7.2    2023-02-01 [1] CRAN (R 4.3.0)
##  rstudioapi        0.14     2022-08-22 [1] CRAN (R 4.3.0)
##  S4Vectors         0.38.0   2023-04-25 [1] Bioconductor
##  scales            1.2.1    2022-08-20 [1] CRAN (R 4.3.0)
##  sessioninfo       1.2.2    2021-12-06 [1] CRAN (R 4.3.0)
##  shape              1.4.6    2021-05-19 [1] CRAN (R 4.3.0)
##  statmod            1.5.0    2023-01-06 [1] CRAN (R 4.3.0)
##  statnet.common     4.8.0    2023-01-24 [1] CRAN (R 4.3.0)
##  statsExpressions  1.5.0    2023-02-19 [1] CRAN (R 4.3.0)
##  stringi            1.7.12   2023-01-11 [1] CRAN (R 4.3.0)
##  stringr            1.5.0    2022-12-02 [1] CRAN (R 4.3.0)
##  SummarizedExperiment 1.30.0  2023-04-25 [1] Bioconductor
##  survival          3.5-3    2023-02-12 [4] CRAN (R 4.2.2)
##  sva                3.48.0   2023-04-25 [1] Bioconductor
##  tibble             3.2.1    2023-03-20 [1] CRAN (R 4.3.0)
##  tidyverse          1.3.0    2023-01-24 [1] CRAN (R 4.3.0)
##  tidyselect         1.2.0    2022-10-10 [1] CRAN (R 4.3.0)
##  utf8               1.2.3    2023-01-31 [1] CRAN (R 4.3.0)

```

```
## vctrs          0.6.2    2023-04-19 [1] CRAN (R 4.3.0)
## WGCNA          1.72-1   2023-01-18 [1] CRAN (R 4.3.0)
## withr          2.5.0    2022-03-03 [1] CRAN (R 4.3.0)
## xfun           0.39     2023-04-20 [1] CRAN (R 4.3.0)
## XML            3.99-0.14 2023-03-19 [1] CRAN (R 4.3.0)
## xtable          1.8-4    2019-04-21 [1] CRAN (R 4.3.0)
## XVector         0.40.0   2023-04-25 [1] Bioconductor
## yaml            2.3.7    2023-01-23 [1] CRAN (R 4.3.0)
## zeallot         0.1.0    2018-01-28 [1] CRAN (R 4.3.0)
## zlibbioc        1.46.0   2023-04-25 [1] Bioconductor
##
## [1] /home/faalm/R/x86_64-pc-linux-gnu-library/4.3
## [2] /usr/local/lib/R/site-library
## [3] /usr/lib/R/site-library
## [4] /usr/lib/R/library
##
## -----
```