

Do environmental interventions ameliorate gut dysbiosis in Huntington disease mice?

Data Analysis (16s)

Saritha Kodikara

October 13, 2021

Contents

Executive Summary	2
1 Data preprocessing	2
2 Composition barplots for relative abundance data	3
3 Diversity	4
3.1 Alpha diversity	4
3.1.1 ANOVA- Alpha diversity	5
3.2 Beta diversity	6
3.2.1 Bray-Curtis distance	6
3.2.2 Unweighted Unifrac distance	8
4 Library size	10
5 PCA	11
5.1 PCA with filtered OTUs for females and Males separately	13
5.1.1 Female	13
6 sPLS-DA - Identifying signatures that discriminate Housing and Genotype (with 5 components)	14
6.1 Female	14
6.2 Male	15
7 sPLS-DA - Identifying signatures that discriminate Housing (with 2 components)	16
7.1 Female- HD	16
7.2 Female- WT	17
7.3 Male- HD	18
7.4 Male- WT	19

8 LMM for OTUs among Females and Males 20

9 Information about the used R session 20

Executive Summary

- Taxonomy classification of the OTUs were done using the ‘`assigntaxonomy()`’ function in ‘`dada2`’ library using the reference database “`silva_nr99_v138.1_wSpecies_train_set.fa.gz`” (<https://zenodo.org/record/4587955#.YH4rbS0RpKM>). The results were saved to ‘`taxonomy.csv`’ file.
- sPLS-DA on the centered log-ratio (CLR) transformed data were used to identify microbial drivers discriminating particular housing conditions.
- There was no clear discrimination in housing when all mice were considered together, indicating that the majority of the variation in the data was not due to the housing condition.
- When sPLS-DA was applied separately based on Sex and Genotype, it was found that the Bacteroidales, Lachnospirales, Oscillospirales were the main signatures that discriminate housing types (“SH”, “EE”, “EX”).

1 Data preprocessing

- Data in ‘`DMGPR00078.biom`’ was read using the ‘`read_biom()`’ function in ‘`biomformat`’ library.
- Filtered out the samples related to brain samples. (‘`DAMG_FM009 - Amplicon 16S Sample Metadata Template_EE Ex-HD & HD brain_CG.xlsx`’ excel file was used for this task).
- Sample information in ‘`DAMG_FM009 - Amplicon 16S Sample Metadata Template_EE Ex-HD & HD brain_CG.xlsx`’ excel file related with the gut were saved in ‘`Gut sample information (DAMG_FM009).csv`’.
- Filtered the OTUs with zero counts across all samples. As a result 207 OTUs out of 5346 OTUs were removed

2 Composition barplots for relative abundance data

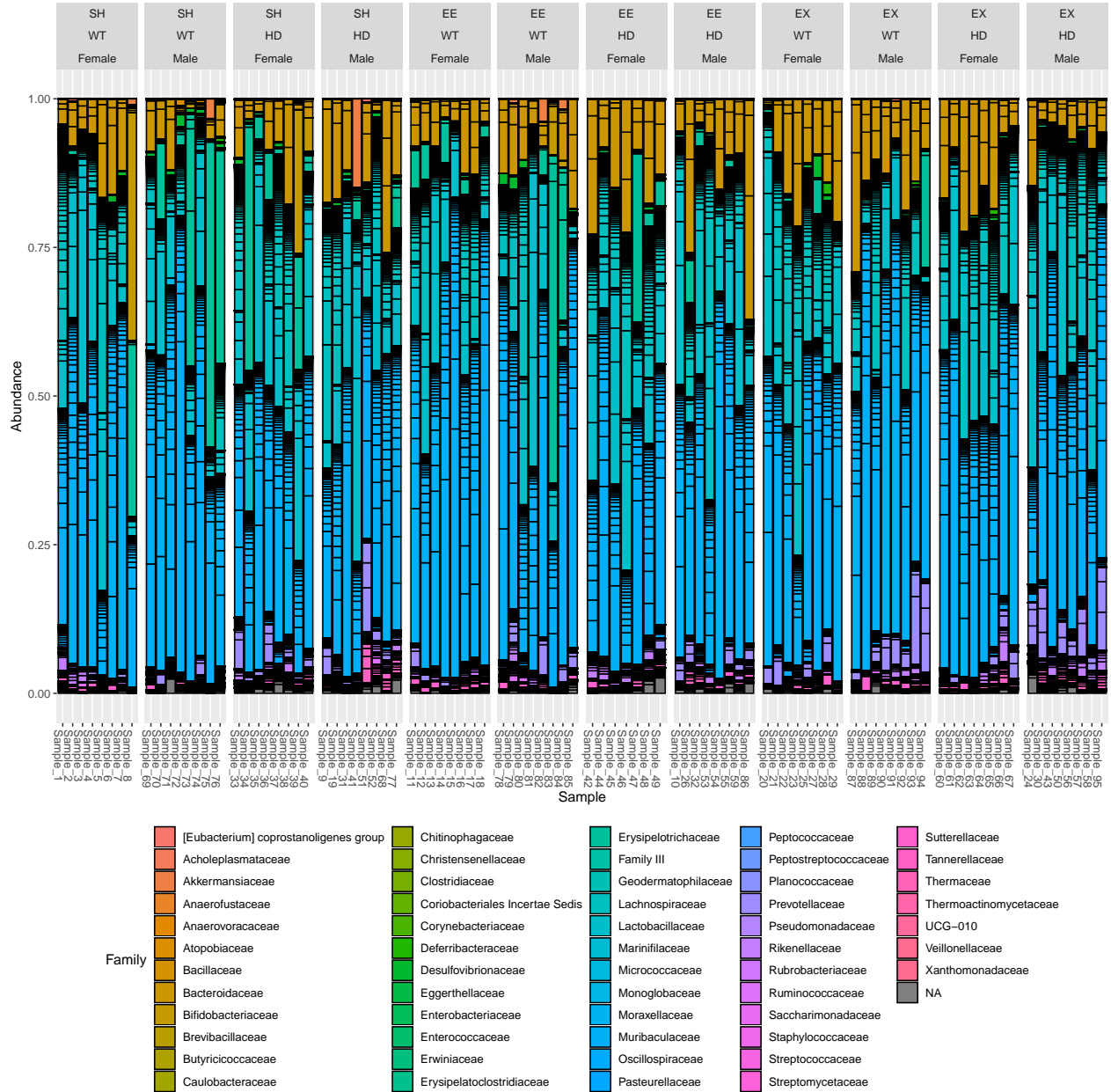


Figure 1: Microbial composition at Family level for each sample by its relative abundance

3 Diversity

3.1 Alpha diversity

- Following Kong et al.(2018), reads were rarefied to 15,000 reads to calculate several alpha- diversity metrics, including species richness (Observed), Shannon and Inverse Simpson metrics, using the ‘Phyloseq’ R package

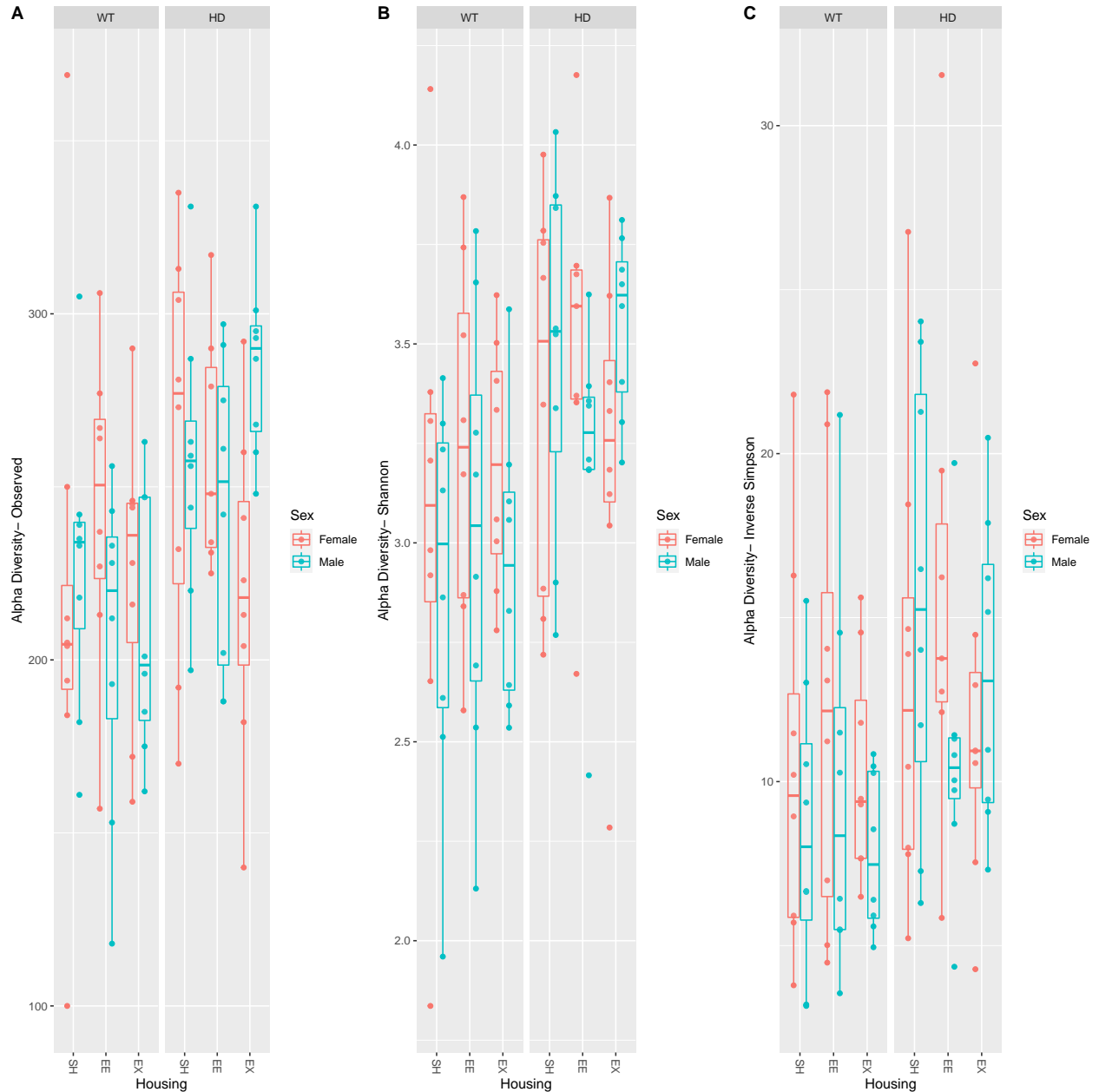


Figure 2: Alpha diversity metrics including Observed (A), Shannon (B) and Inverse Simpson (InvSimpson) (C) indices were probed for the two genotypes among different housing conditions.

- Figure 2 A, gives the species richness measure which represents the number of observed OTUs in a

given 16s sample. Based on the boxplots it can be concluded that the lowest median richness value was observed in Males-‘WT’-‘EX’ group.

- Figure 2 B and C represent the Shannon’s and Inverse Simpson diversity index, which takes into account both richness and the relative abundance of OTUs (i.e. evenness). For both these measures, HD group had high median evenness compared to its counterpart in ‘WT’.

3.1.1 ANOVA- Alpha diversity

- Based on the analysis of variance results, we can infer that there is a statistical difference in the alpha diversity among HD and WT.

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Housing      2    487   243.7    0.094   0.91
## Residuals    92 238000  2587.0

##           Df Sum Sq Mean Sq F value Pr(>F)
## Housing      2   0.034  0.01694    0.073   0.93
## Residuals    92 21.475  0.23342

##           Df Sum Sq Mean Sq F value Pr(>F)
## Housing      2   29.2   14.60    0.427  0.654
## Residuals    92 3144.5   34.18

##           Df Sum Sq Mean Sq F value Pr(>F)
## Sex          1     9     8.6    0.003  0.954
## Residuals    93 238479  2564.3

##           Df Sum Sq Mean Sq F value Pr(>F)
## Sex          1   0.151  0.1510    0.658  0.419
## Residuals    93 21.357  0.2296

##           Df Sum Sq Mean Sq F value Pr(>F)
## Sex          1   38.6   38.63    1.146  0.287
## Residuals    93 3135.0   33.71

##           Df Sum Sq Mean Sq F value Pr(>F)
## Genotype      1  27201  27201   11.97 0.000817 ***
## Residuals    93 211286   2272
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##           Df Sum Sq Mean Sq F value Pr(>F)
## Genotype      1   2.648  2.6485   13.06 0.000489 ***
## Residuals    93 18.860  0.2028
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##           Df Sum Sq Mean Sq F value Pr(>F)
## Genotype      1  291.4  291.39   9.402 0.00284 **
## Residuals    93 2882.3   30.99
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3.2 Beta diversity

3.2.1 Bray-Curtis distance

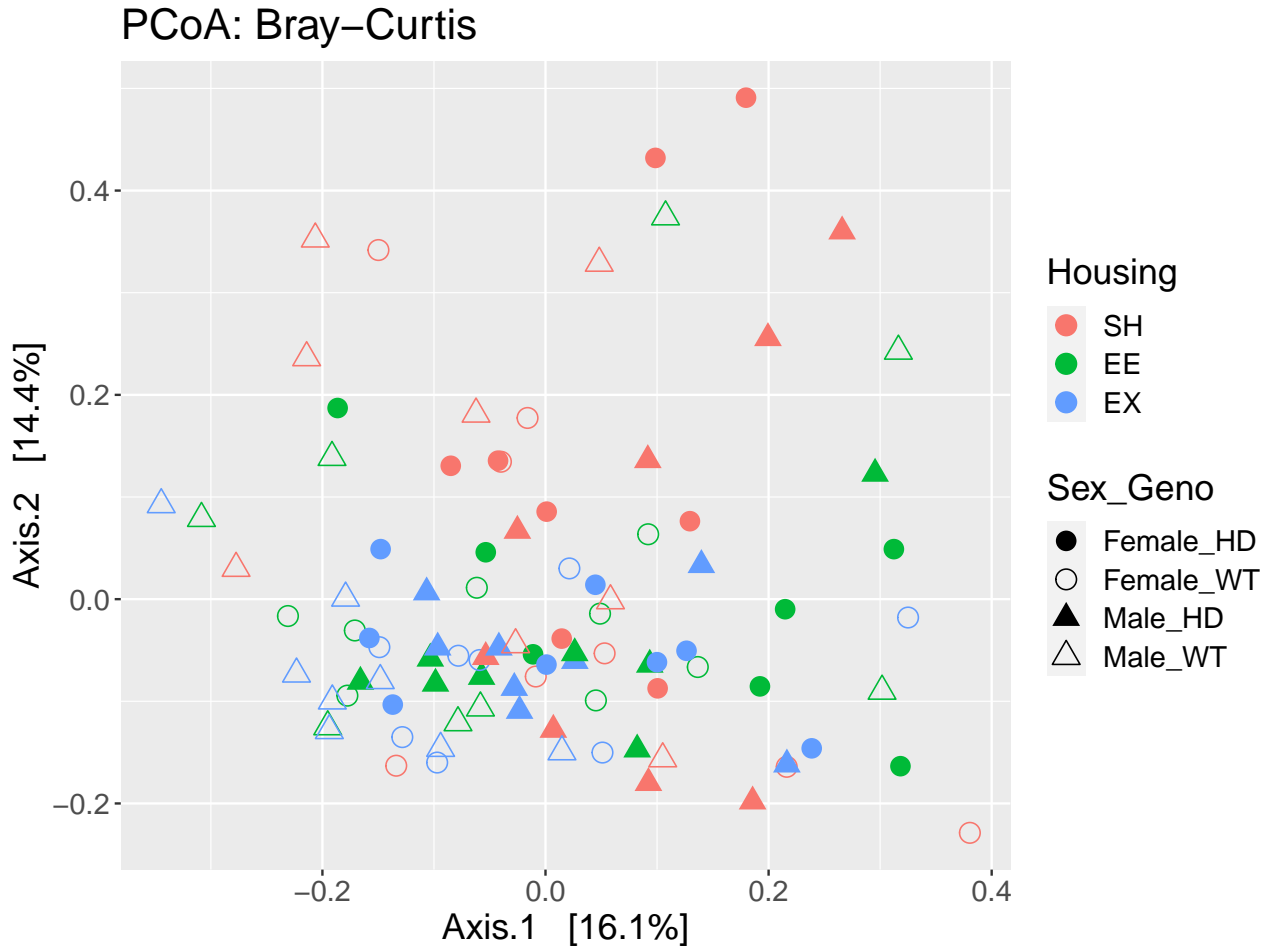


Figure 3: PCoA: Bray-Curtis by Genotype, Housing and Sex

```
##
## Call:
## vegan::adonis(formula = distBC ~ Genotype, data = data.frame(sample_data(physeq_ra)))
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##          Df SumsOfSqs MeanSqs F.Model    R2 Pr(>F)
## Genotype   1     0.465  0.46499   3.0032 0.03128 0.001 ***
## Residuals  93    14.399  0.15483         0.96872
## Total     94    14.864         1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Call:
## vegan::adonis(formula = distBC ~ Housing, data = data.frame(sample_data(physeq_ra)))
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##           Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## Housing    2    0.7221 0.36107  2.3488 0.04858 0.002 **
## Residuals 92   14.1423 0.15372          0.95142
## Total     94   14.8645          1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Call:
## vegan::adonis(formula = distBC ~ Sex, data = data.frame(sample_data(physeq_ra)))
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##           Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## Sex         1    0.2986 0.29858  1.9063 0.02009 0.022 *
## Residuals 93   14.5659 0.15662          0.97991
## Total     94   14.8645          1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- In Figure 3, Bray-Curtis distance was calculated and used in principal coordinates analysis (PCoA) to compare the microbial composition in one environment to another. Although the results in this figure do not indicate strong clustering according to any factor variable (i.e. Housing, Sex and Genotype), Bray-Curtis distance was found to be statistically different in each factor.

3.2.2 Unweighted Unifrac distance

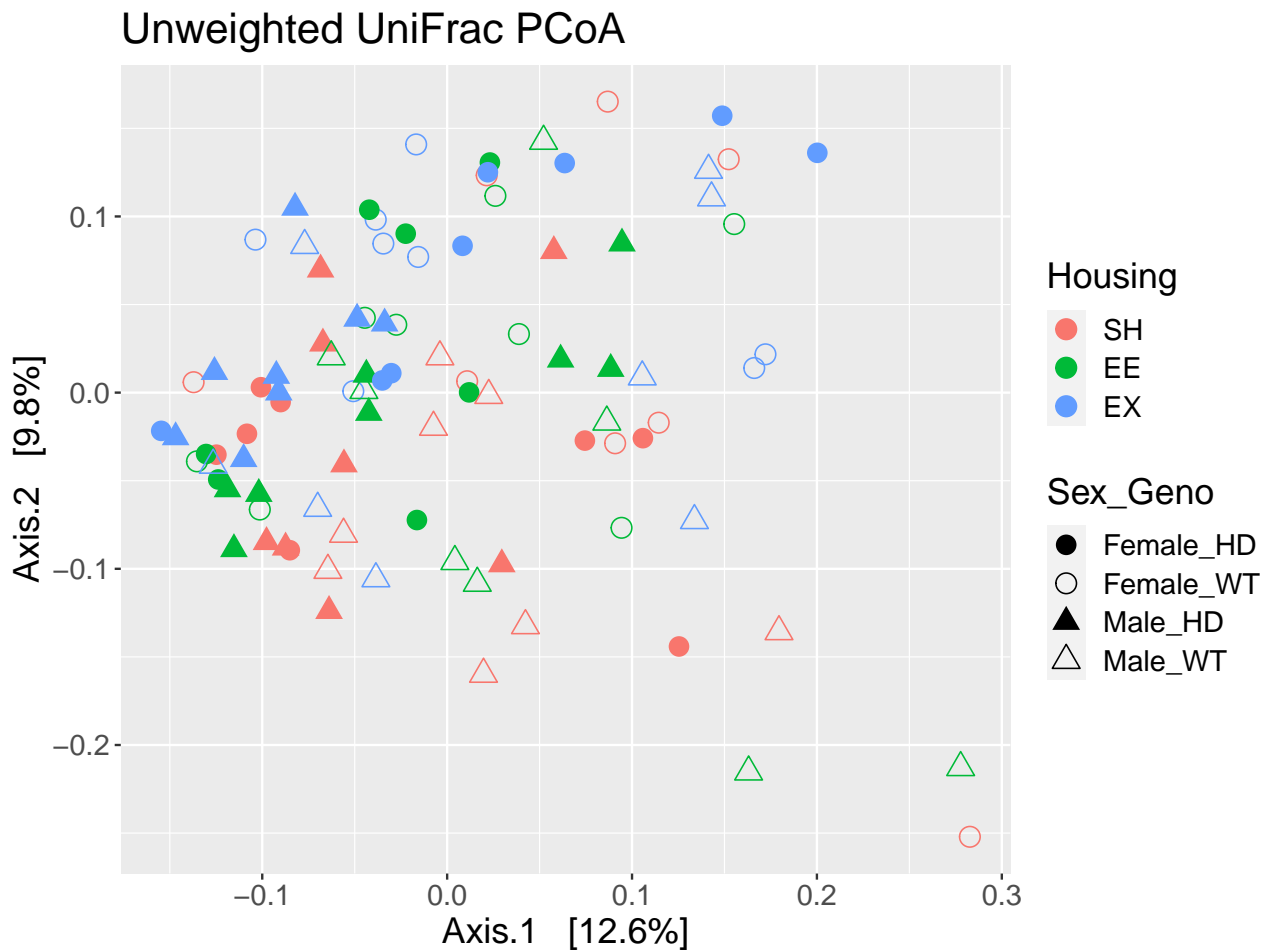


Figure 4: PCoA: Unweighted Unifrac distance by Genotype, Housing and Sex

```
##
## Call:
## vegan::adonis(formula = distUF ~ Genotype, data = data.frame(sample_data(physeq_ra)))
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##          Df SumsOfSqs MeanSqs F.Model    R2 Pr(>F)
## Genotype  1   0.1873 0.18734  2.3868 0.02502 0.004 **
## Residuals 93   7.2996 0.07849         0.97498
## Total    94   7.4869         1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Call:
```



```
## vegan::adonis(formula = distUF ~ Housing, data = data.frame(sample_data(physeq_ra)))
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##           Df SumsOfSqs  MeanSqs F.Model      R2 Pr(>F)
## Housing    2    0.2930 0.146480  1.8733 0.03913 0.002 **
## Residuals 92    7.1939 0.078195          0.96087
## Total     94    7.4869          1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Call:
## vegan::adonis(formula = distUF ~ Sex, data = data.frame(sample_data(physeq_ra)))
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##           Df SumsOfSqs  MeanSqs F.Model      R2 Pr(>F)
## Sex         1    0.1563 0.156336  1.9834 0.02088 0.005 **
## Residuals 93    7.3306 0.078823          0.97912
## Total     94    7.4869          1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- In Figure 4, unweighted UniFrac distance distance was calculates and used in principal coordinates analysis (PCoA) to compare the microbial composition in one environment to another. Compared to Bray-Curtis distance, the unweighted UniFrac distance accounts for the phylogenetic relationship between the OTUs. Using Adonis (Permutation multivariate ANOVA) from the ‘vegan’ R package, it was found that the distances were statistically different across factors as similar to Bray-Curtis distance.

4 Library size

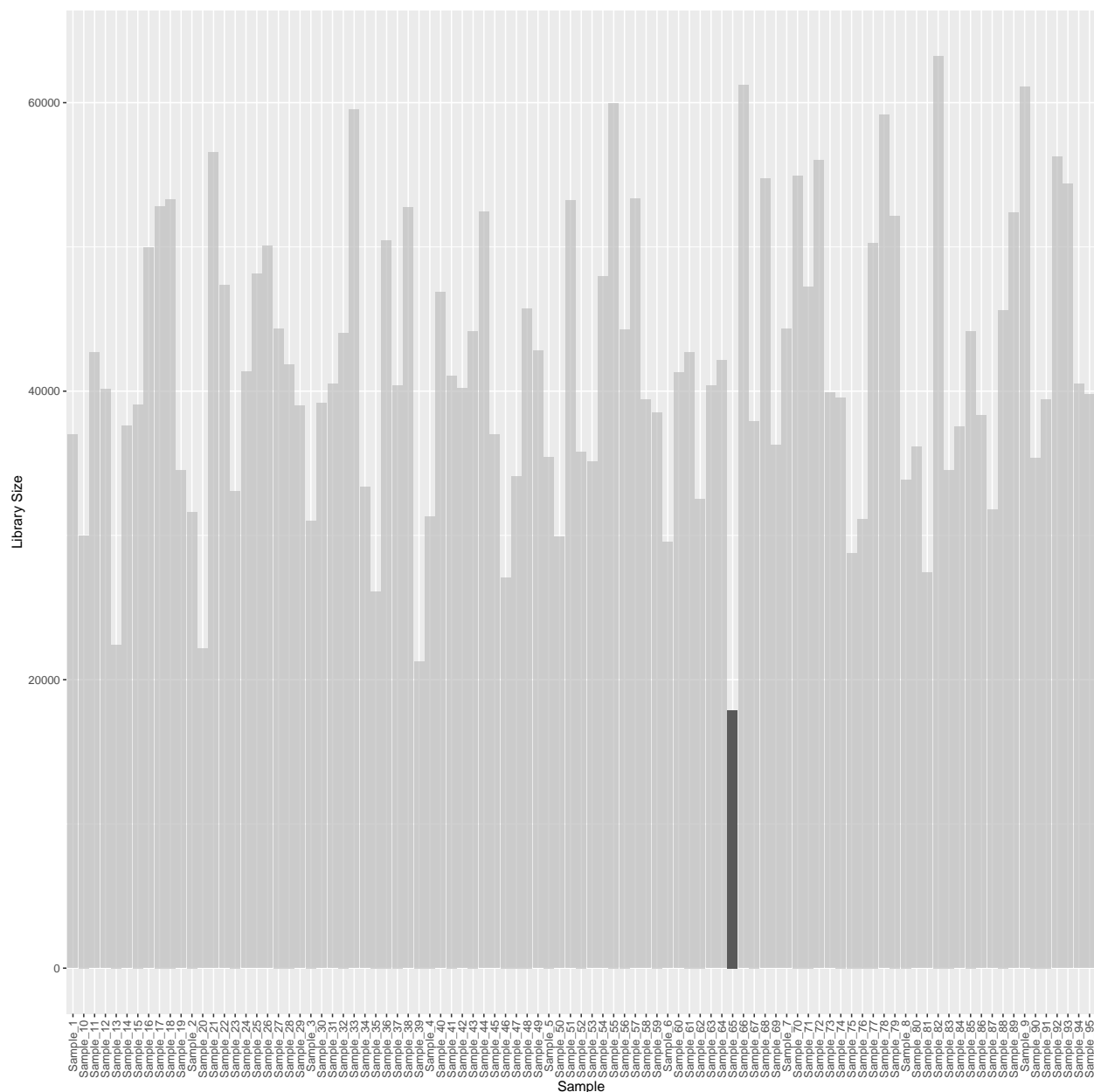


Figure 5: Bar plot for library size per sample

In Figure 5, the sum of all OTU counts per sample (i.e. library size) is represented in a barplot to identify any heterogeneous library sizes. Based on the results, ‘Sample 65’ has a relatively low library size. However, this sample was not removed from future analysis as there was no compelling reason to do so.

5 PCA

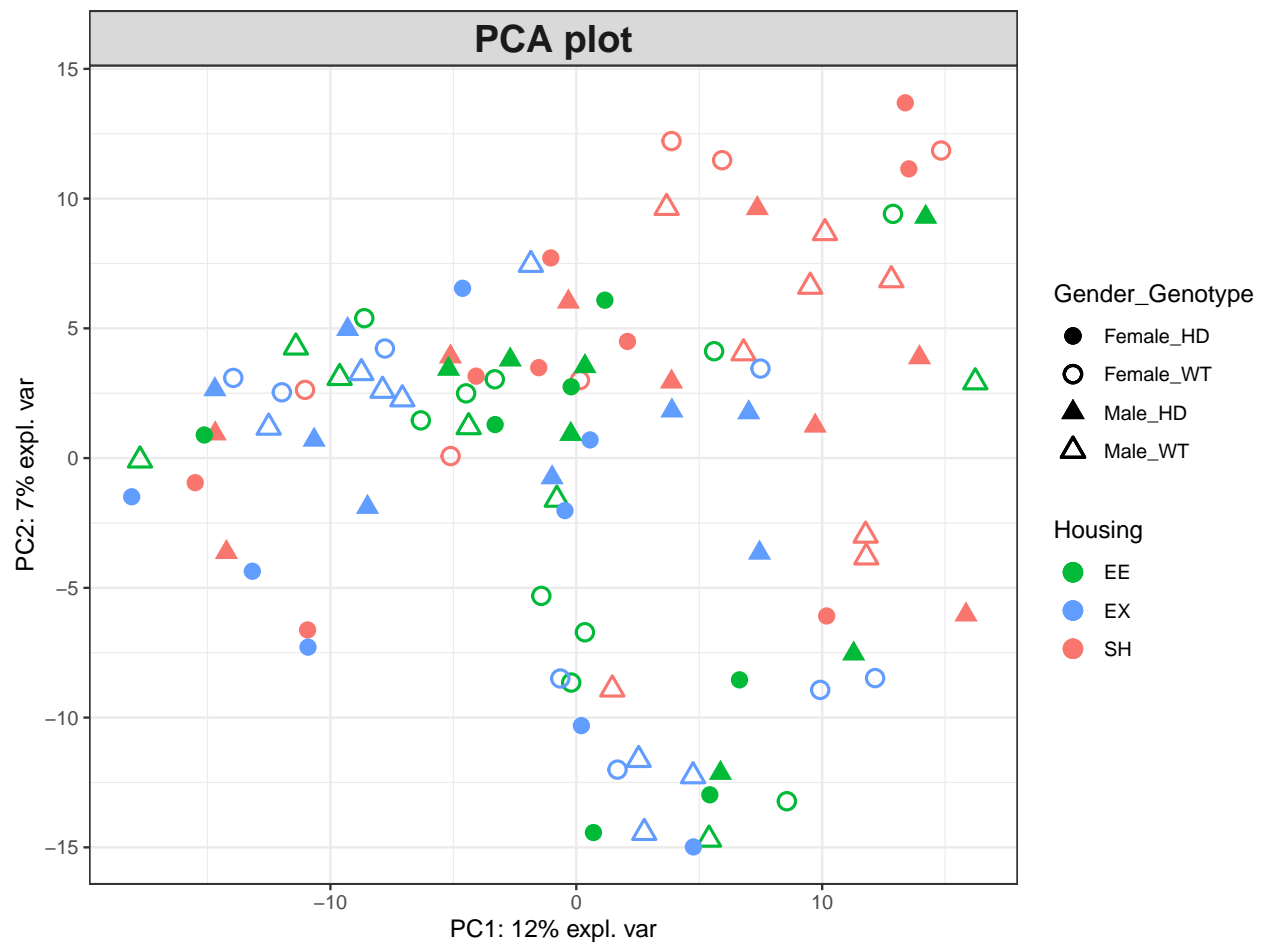


Figure 6: PCA plot with Housing, Sex and Genotype.

- Based on the PCA plot in 6, there is no clear separation among samples coming from different housing conditions.

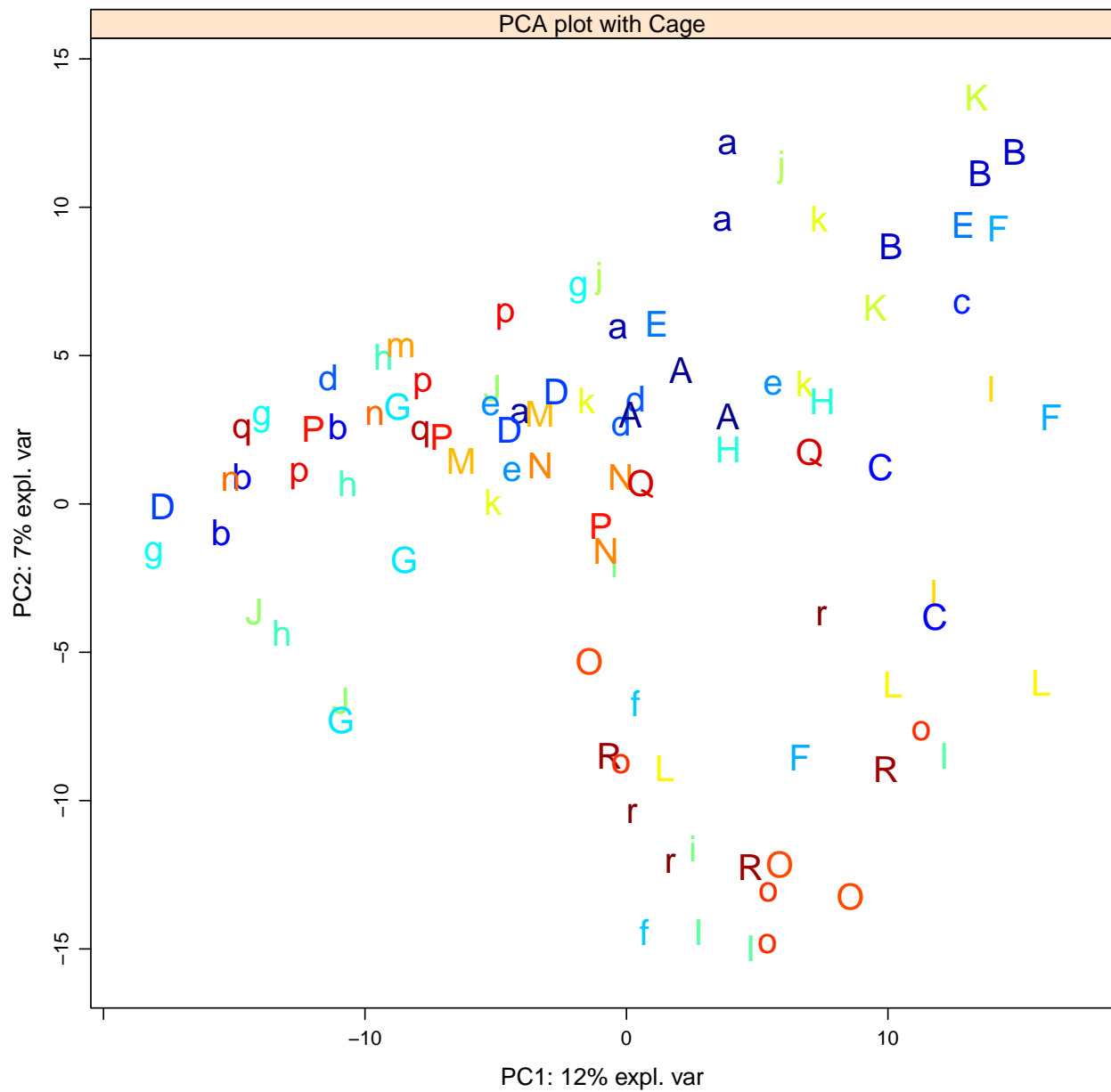


Figure 7: PCA plot with cage effect.

- There seems to be mild cage effect as some cages (i.e. 'A', 'B') tends to cluster together.

5.1 PCA with filtered OTUs for females and Males separately

5.1.1 Female

- Used a pre-filtering step to remove OTUs separately for males and females for which the sum of counts are below 0.01%. As a result for females there were 290 OTUs after the pre-filter and for males it was 304. There were 258 OTUs that were common to both sex.

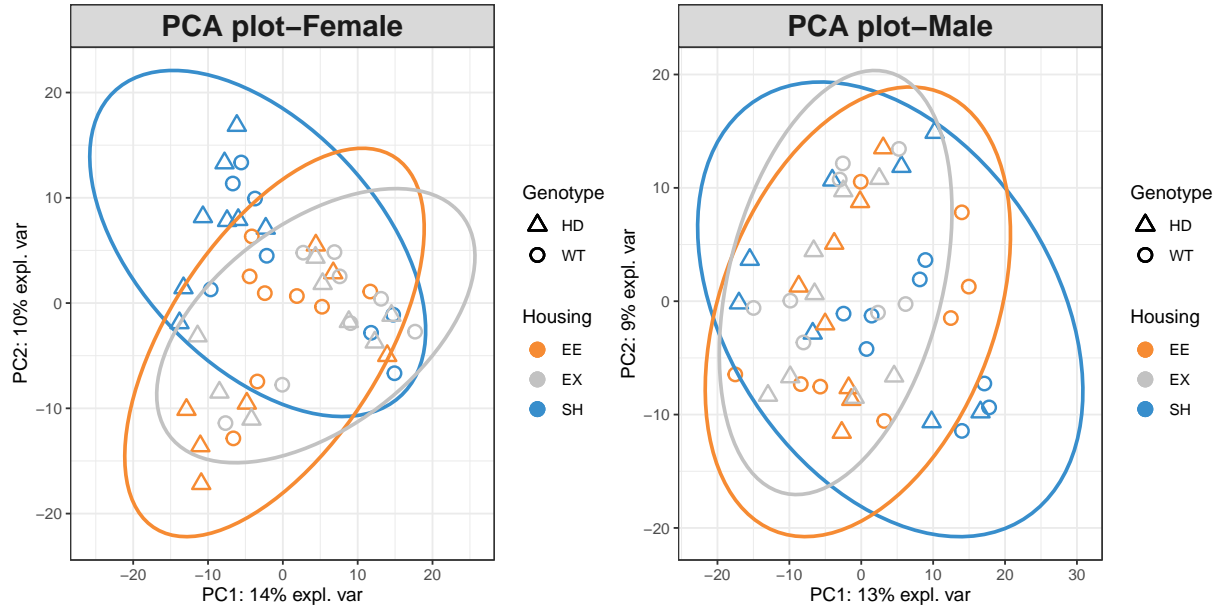


Figure 8: PCA plot grouped by sex

6 sPLS-DA - Identifying signatures that discriminate Housing and Genotype (with 5 components)

6.1 Female

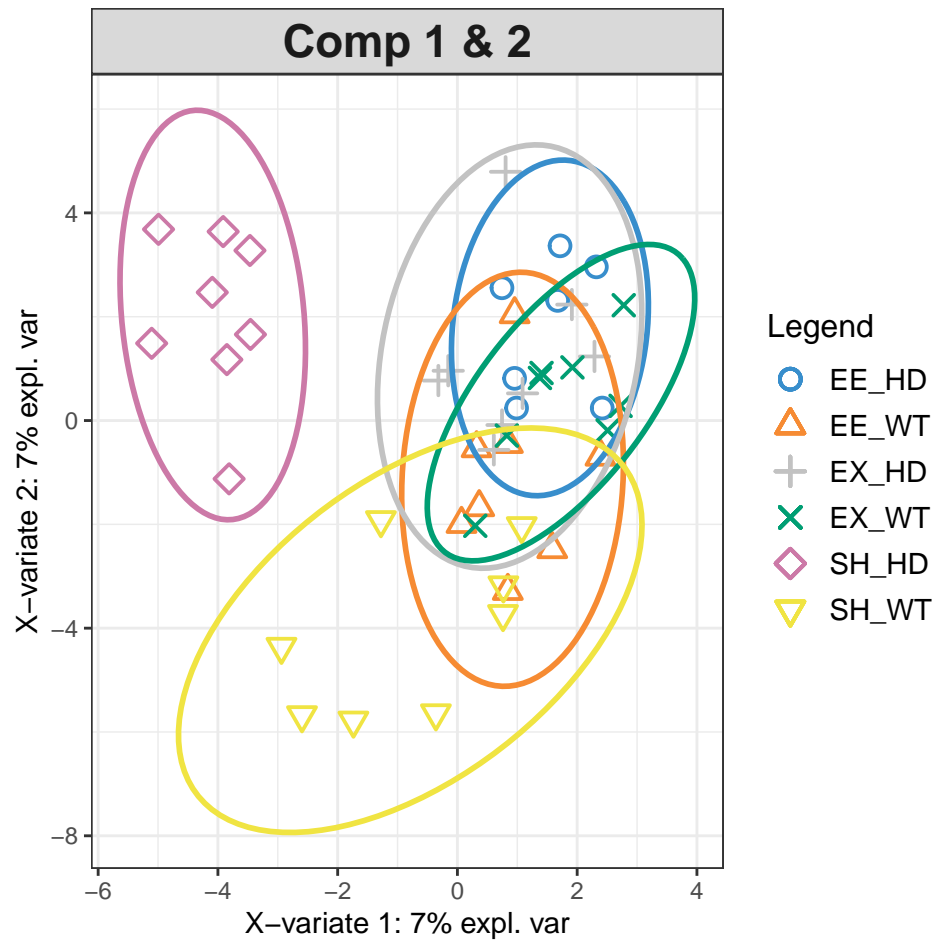


Figure 9: Identification of OTU signatures associated with Housing and Genotype in Females using sPLS-DA

6.2 Male

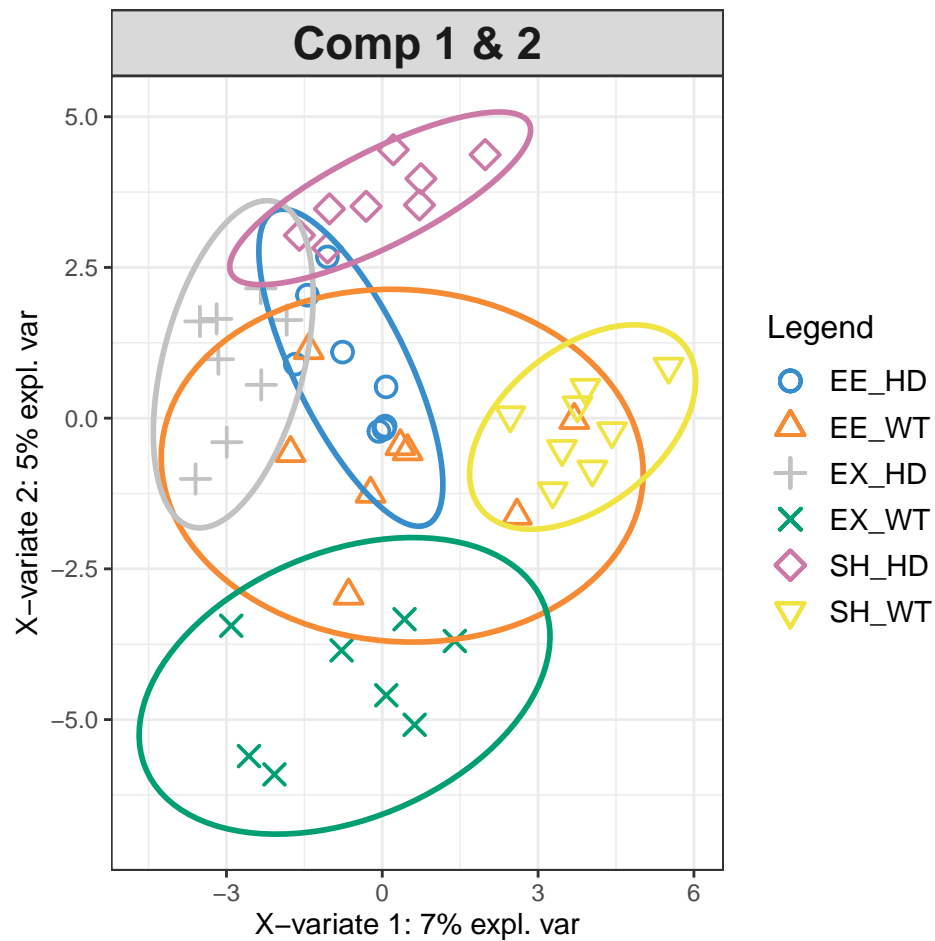


Figure 10: Identification of OTU signatures associated with Housing and Genotype in Males using sPLS-DA

Overall Error Rate (Mahalanobis distance)

	Error	SD
Female	0.5000000	0.0503995
Male	0.4729167	0.0644376

- There is no clear separation in Housing and genotype in both males and females. Thus, Sparse PLS discriminant analysis (sPLS-DA) from the 'mixOmics' package in R was used to identify signature of discriminative OTUs associated with Housing among different sex and genotype.

7 sPLS-DA - Identifying signatures that discriminate Housing (with 2 components)

7.1 Female- HD

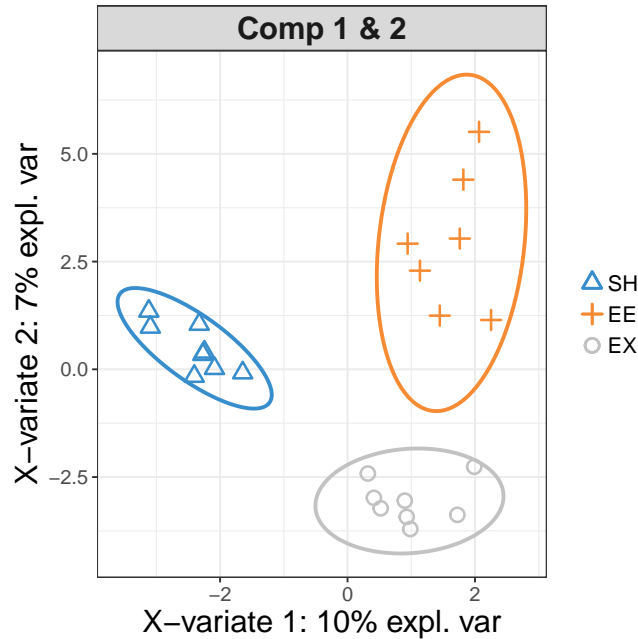


Figure 11: Identification of OTU signatures associated with Housing type in Female HD using sPLS-DA

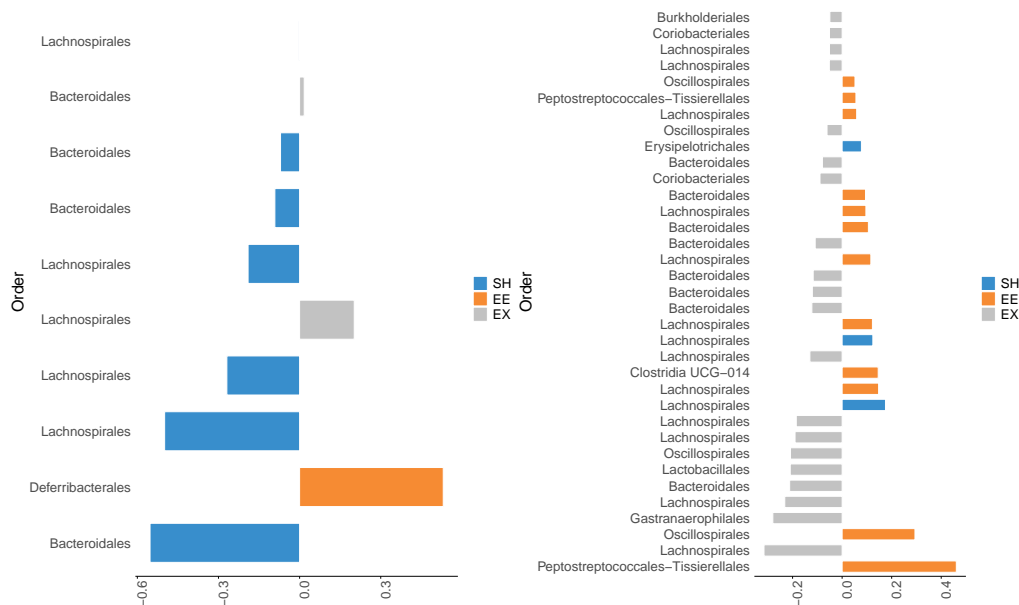


Figure 12: Loadings of OTU signatures associated with Housing type in Female HD

- Figure 12, displays the loading weights for component 1 and 2, where colors indicate the housing condition for which the selected OTU has a maximal median value.

7.2 Female- WT

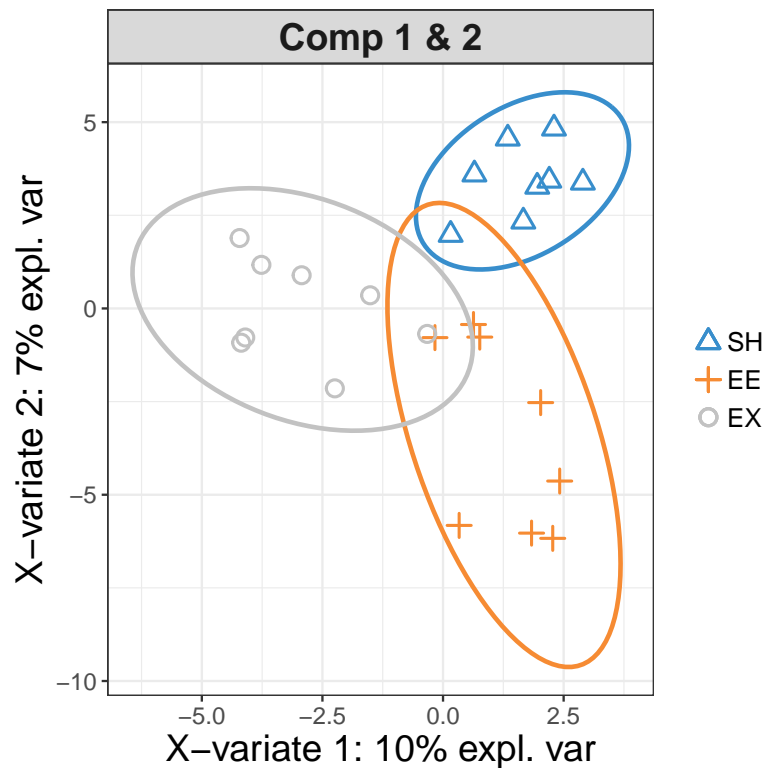


Figure 13: Identification of OTU signatures associated with Housing type in Female WT using sPLS-DA

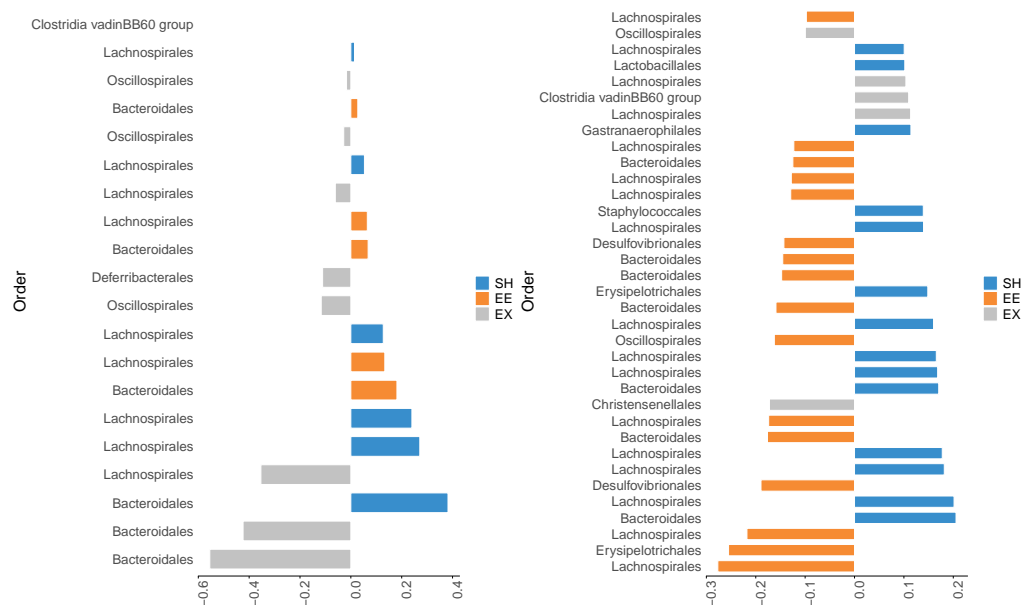


Figure 14: Loadings of OTU signatures associated with Housing type in Female WT

7.3 Male- HD

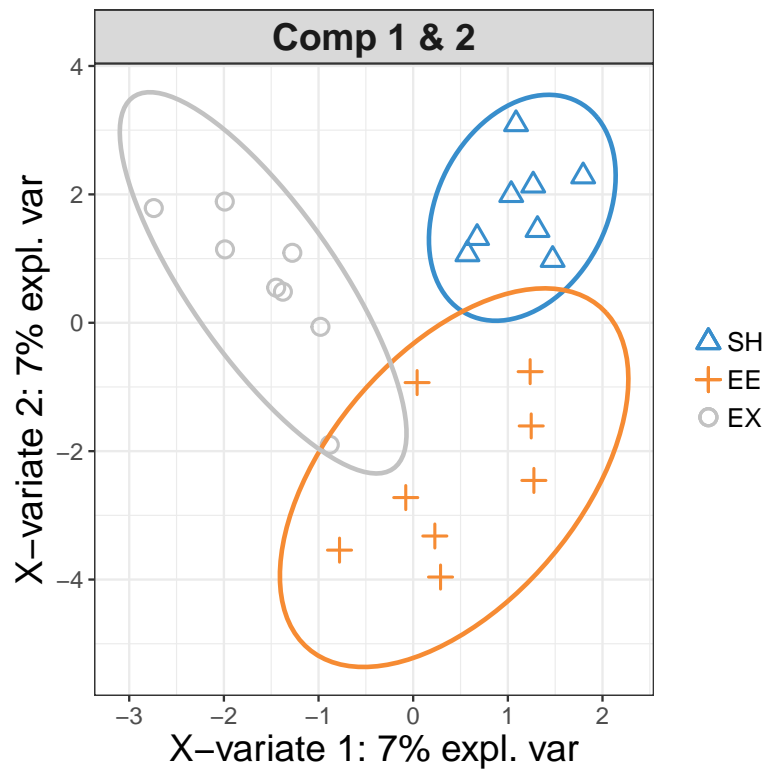


Figure 15: Identification of OTU signatures associated with Housing type in Male HD using sPLS-DA

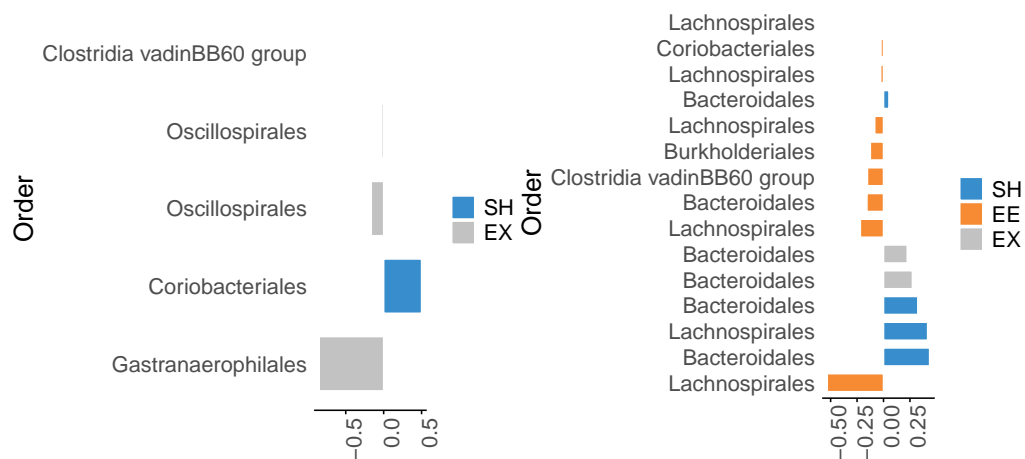


Figure 16: Loadings of OTU signatures associated with Housing type in Male HD

7.4 Male- WT

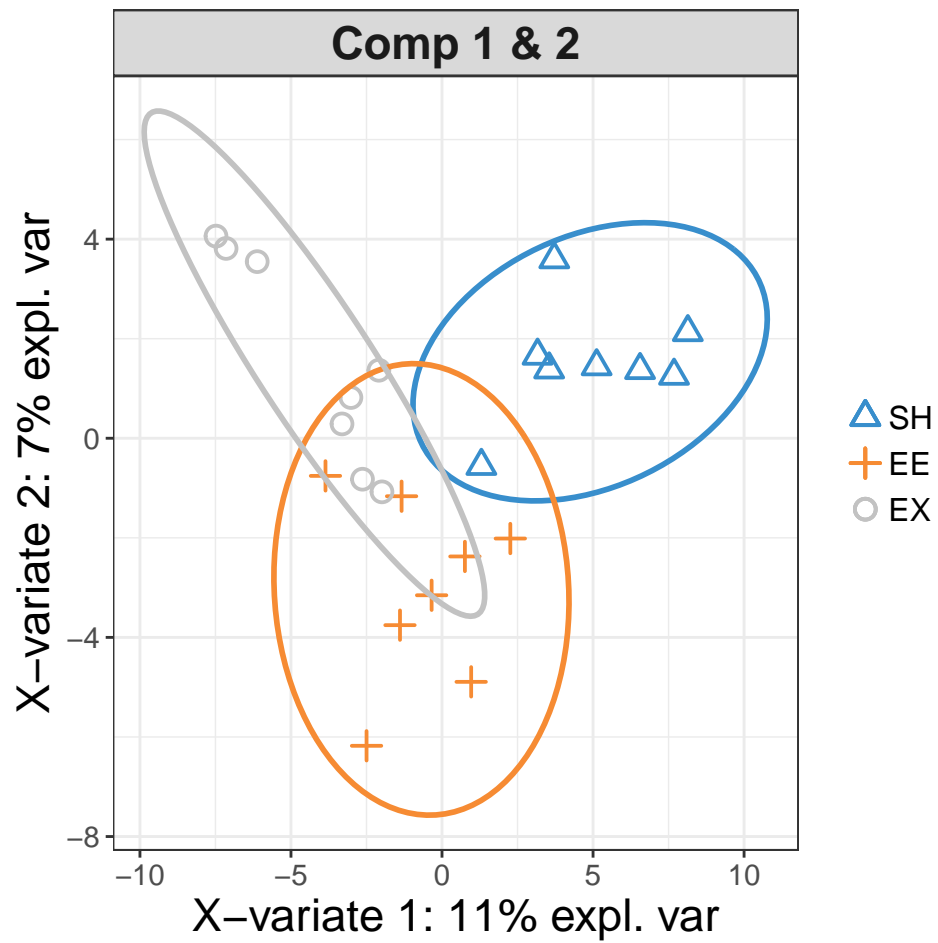


Figure 17: Identification of OTU signatures associated with Housing type in Male WT using sPLS-DA

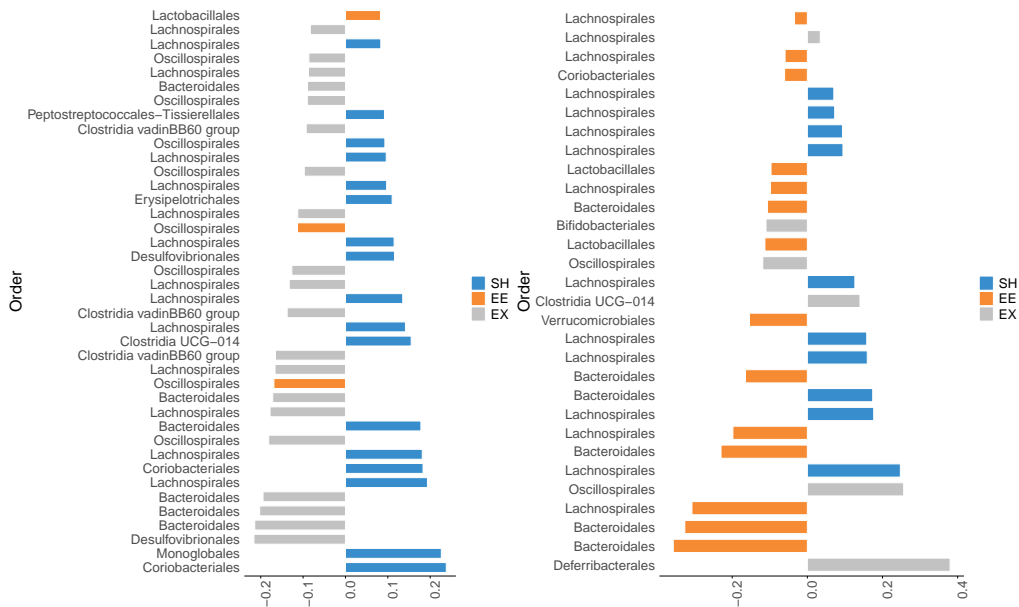


Figure 18: Loadings of OTU signatures associated with Housing type in Male WT

Overall Error Rate (Mahalanobis distance)

	Error	SD
Female_HD	0.3260870	0.0966787
Female_WT	0.3375000	0.0664638
Male_HD	0.2625000	0.0922331
Male_WT	0.4208333	0.0746649

Error Rate by class (Mahalanobis distance)

	SH	EE	EX
Female_HD	0.1500	0.4857143	0.3625
Female_WT	0.2750	0.4750000	0.2625
Male_HD	0.2000	0.3625000	0.2250
Male_WT	0.1875	0.6500000	0.4125

8 LMM for OTUs among Females and Males

- Results of these univariate analysis are saved in the “Results.xlsx” file.

9 Information about the used R session

```
sessionInfo()
```

```
## R version 4.1.0 (2021-05-18)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
```

```
## BLAS: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_AU.UTF-8/en_AU.UTF-8/en_AU.UTF-8/C/en_AU.UTF-8/en_AU.UTF-8
##
## attached base packages:
## [1] stats4      parallel  stats      graphics  grDevices  utils      datasets
## [8] methods    base
##
## other attached packages:
## [1] patchwork_1.1.1      ape_5.5          DECIPHER_2.20.0
## [4] RSQLite_2.2.8        Biostrings_2.60.2 GenomeInfoDb_1.28.4
## [7] XVector_0.32.0       IRanges_2.26.0   S4Vectors_0.30.2
## [10] BiocGenerics_0.38.0  vegan_2.5-7      permute_0.9-5
## [13] stargazer_5.2.2      xlsx_0.6.5       nlme_3.1-153
## [16] gtools_3.9.2         gghighlight_0.3.2 kableExtra_1.3.4
## [19] ggfortify_0.4.12     mixOmics_6.16.3  lattice_0.20-45
## [22] MASS_7.3-54          ggpubr_0.4.0     ggplot2_3.3.5
## [25] phyloseq_1.36.0      dada2_1.20.0     Rcpp_1.0.7
## [28] dplyr_1.0.7          biomformat_1.20.0 knitr_1.36
##
## loaded via a namespace (and not attached):
## [1] readxl_1.3.1          backports_1.2.1
## [3] systemfonts_1.0.2     plyr_1.8.6
## [5] igraph_1.2.6          splines_4.1.0
## [7] BiocParallel_1.26.2   digest_0.6.28
## [9] foreach_1.5.1         htmltools_0.5.2
## [11] fansi_0.5.0           memoise_2.0.0
## [13] magrittr_2.0.1        cluster_2.1.2
## [15] openxlsx_4.2.4        RcppParallel_5.1.4
## [17] matrixStats_0.61.0    rARPACK_0.11-0
## [19] svglite_2.0.0         jpeg_0.1-9
## [21] colorspace_2.0-2      blob_1.2.2
## [23] rvest_1.0.1           ggrepel_0.9.1
## [25] haven_2.4.3           xfun_0.26
## [27] crayon_1.4.1          RCurl_1.98-1.5
## [29] jsonlite_1.7.2        survival_3.2-13
## [31] iterators_1.0.13      glue_1.4.2
## [33] gtable_0.3.0          zlibbioc_1.38.0
## [35] webshot_0.5.2         DelayedArray_0.18.0
## [37] car_3.0-11            Rhdf5lib_1.14.2
## [39] abind_1.4-5           scales_1.1.1
## [41] DBI_1.1.1             rstatix_0.7.0
## [43] viridisLite_0.4.0     foreign_0.8-81
## [45] bit_4.0.4            httr_1.4.2
## [47] RColorBrewer_1.1-2    ellipsis_0.3.2
## [49] farver_2.1.0          pkgconfig_2.0.3
## [51] rJava_1.0-5           utf8_1.2.2
## [53] labeling_0.4.2        tidyselect_1.1.1
## [55] rlang_0.4.11          reshape2_1.4.4
## [57] cachem_1.0.6          munsell_0.5.0
## [59] cellranger_1.1.0      tools_4.1.0
## [61] generics_0.1.0        ade4_1.7-18
```

```
## [63] broom_0.7.9          evaluate_0.14
## [65] stringr_1.4.0        fastmap_1.1.0
## [67] yaml_2.2.1           bit64_4.0.5
## [69] zip_2.2.0            purrr_0.3.4
## [71] xml2_1.3.2           compiler_4.1.0
## [73] rstudioapi_0.13      curl_4.3.2
## [75] png_0.1-7            ggsignif_0.6.3
## [77] tibble_3.1.5         stringi_1.7.5
## [79] RSpectra_0.16-0      forcats_0.5.1
## [81] Matrix_1.3-4         multtest_2.48.0
## [83] vctrs_0.3.8          pillar_1.6.3
## [85] lifecycle_1.0.1      rhdf5filters_1.4.0
## [87] cowplot_1.1.1        data.table_1.14.2
## [89] bitops_1.0-7         corpcor_1.6.10
## [91] GenomicRanges_1.44.0 R6_2.5.1
## [93] latticeExtra_0.6-29  hwriter_1.3.2
## [95] ShortRead_1.50.0     gridExtra_2.3
## [97] rio_0.5.27           codetools_0.2-18
## [99] assertthat_0.2.1     rhdf5_2.36.0
## [101] SummarizedExperiment_1.22.0 xlsxjars_0.6.1
## [103] withr_2.4.2          GenomicAlignments_1.28.0
## [105] Rsamtools_2.8.0      GenomeInfoDbData_1.2.6
## [107] mgcv_1.8-37          hms_1.1.1
## [109] grid_4.1.0           tidyr_1.1.4
## [111] rmarkdown_2.11       MatrixGenerics_1.4.3
## [113] carData_3.0-4        Biobase_2.52.0
## [115] ellipse_0.4.2
```