```
library(MicrobiotaProcess)

colors <- c("#00A087FF", "#3C5488FF")

otuda <- read.table("./IBD_data/ibd_asv_table.txt", header=T, check.names=F,
                     comment.char="", row.names=1, sep="\t")
otuda <- data.frame(t(otuda), check.names=F)
sampleda <- read.csv("./IBD_data/ibd_meta.csv", row.names=1, comment.char="")
taxda <- read.table("./IBD_data/ibd_taxa.txt", header=T,
                    row.names=1, check.names=F, comment.char="")
mpse <- mp_import_dada2(seqtab=otuda, taxatab=taxda, sampleda=sampleda)
mpse
```

```
## # A MPSE-tibble (MPSE object) abstraction: 67,295 x 11
## # OTU=1565 | Samples=43 | Assays=Abundance | Taxanomy=Kingdom, Phylum, Class, Order, Family, Genus, Species
##      OTU      Sample Abundance Group Kingdom Phylum Class Order Family Genus Species
##    <chr>     <chr>      <int> <chr> <chr>   <chr>  <chr> <chr> <chr>  <chr> <chr>
##  1 OTU_1497 S2067~         0 CD    k__Bac~ p__[T~ c__D~ o__D~ f__De~ g__D~ s__un_~
##  2 OTU_1005 S2067~         0 CD    k__Bac~ p__[T~ c__D~ o__T~ f__Th~ g__T~ s__un_~
##  3 OTU_1333 S2067~         0 CD    k__Bac~ p__[T~ c__D~ o__T~ f__Th~ g__T~ s__un_~
##  4 OTU_1663 S2067~         0 CD    k__Bac~ p__[T~ c__D~ o__T~ f__Th~ g__T~ s__un_~
##  5 OTU_1665 S2067~         0 CD    k__Bac~ p__[T~ c__D~ o__T~ f__Th~ g__T~ s__un_~
##  6 OTU_1190 S2067~         0 CD    k__Bac~ p__Ac~ c__[~ o__D~ f__un~ g__u~ s__un_~
##  7 OTU_1449 S2067~         0 CD    k__Bac~ p__Ac~ c__[~ o__R~ f__El~ g__u~ s__un_~
##  8 OTU_1583 S2067~         0 CD    k__Bac~ p__Ac~ c__A~ o__i~ f__mb~ g__u~ s__un_~
##  9 OTU_1443 S2067~         0 CD    k__Bac~ p__Ac~ c__A~ o__i~ f__un~ g__u~ s__un_~
## 10 OTU_1035 S2067~         0 CD    k__Bac~ p__Ac~ c__A~ o__A~ f__C1~ g__u~ s__un_~
## # ... with 67,285 more rows
```

```
tree <- treeio::read.iqtree("./IBD_data/iqtree/training_otus.treefile")
tree
```

```
## 'treedata' S4 object that stored information of
## './IBD_data/iqtree/training_otus.treefile'.
##
## ...@ phylo:
##
## Phylogenetic tree with 1681 tips and 1679 internal nodes.
##
## Tip labels:
##    OTU_878, OTU_728, OTU_417, OTU_749, OTU_996, OTU_1551, ...
## Node labels:
##    , 0/89, 63.2/88, 0/37, 72.3/76, 72.3/36, ...
##
## Unrooted; includes branch lengths.
##
## with the following features available:
##   'SH_aLRT',  'UFboot'.
##
## # The associated data tibble abstraction: 3,360 x 5
## # The 'node', 'label' and 'isTip' are from the phylo tree.
##      node label     isTip SH_aLRT UFboot
##     <int> <chr>      <lgl>   <dbl>  <dbl>
## 1       1 OTU_878    TRUE       NA     NA
## 2       2 OTU_728    TRUE       NA     NA
## 3       3 OTU_417    TRUE       NA     NA
## 4       4 OTU_749    TRUE       NA     NA
## 5       5 OTU_996    TRUE       NA     NA
## 6       6 OTU_1551   TRUE       NA     NA
## 7       7 OTU_1122   TRUE       NA     NA
## 8       8 OTU_1021   TRUE       NA     NA
```

```
## 9      9 OTU_764  TRUE         NA      NA
## 10    10 OTU_717  TRUE         NA      NA
## # ... with 3,350 more rows
```

```r
otutree(mpse) <- tree
mpse %<>% mp_rrarefy()
mpse
```

```
## # A MPSE-tibble (MPSE object) abstraction: 67,295 x 12
## # OTU=1565 | Samples=43 | Assays=Abundance, RareAbundance | Taxanomy=Kingdom, Phylum, Class, Order, Family,
## #    OTU   Sample Abundance RareAbundance Group Kingdom Phylum Class Order Family
## #    <chr> <chr>      <int>         <int> <chr> <chr>   <chr>  <chr> <chr> <chr>
## #  1 OTU_1~ S2067~        0             0 CD    k__Bac~ p__[T~ c__D~ o__D~ f__De~
## #  2 OTU_1~ S2067~        0             0 CD    k__Bac~ p__[T~ c__D~ o__T~ f__Th~
## #  3 OTU_1~ S2067~        0             0 CD    k__Bac~ p__[T~ c__D~ o__T~ f__Th~
## #  4 OTU_1~ S2067~        0             0 CD    k__Bac~ p__[T~ c__D~ o__T~ f__Th~
## #  5 OTU_1~ S2067~        0             0 CD    k__Bac~ p__[T~ c__D~ o__T~ f__Th~
## #  6 OTU_1~ S2067~        0             0 CD    k__Bac~ p__Ac~ c__[~ o__D~ f__un~
## #  7 OTU_1~ S2067~        0             0 CD    k__Bac~ p__Ac~ c__[~ o__R~ f__El~
## #  8 OTU_1~ S2067~        0             0 CD    k__Bac~ p__Ac~ c__A~ o__i~ f__mb~
## #  9 OTU_1~ S2067~        0             0 CD    k__Bac~ p__Ac~ c__A~ o__i~ f__un~
## # 10 OTU_1~ S2067~        0             0 CD    k__Bac~ p__Ac~ c__A~ o__A~ f__C1~
## # ... with 67,285 more rows, and 2 more variables: Genus <chr>, Species <chr>
```

```r
mpse %<>% mp_cal_rarecurve(.abundance=RareAbundance, action="add", chunk=400)

p1 <- mpse %>%
    mp_plot_rarecurve(
        .rare = RareAbundanceRarecurve,
        .alpha = Observe,
        show.legend = FALSE
    )

p2 <- mpse %>%
    mp_plot_rarecurve(
        .rare = RareAbundanceRarecurve,
        .alpha = Observe,
        .group = Group
    ) +
    scale_color_manual(values=colors) +
    scale_fill_manual(values=colors) +
    theme(legend.position=c(0.2, 0.8),
          legend.background=element_blank())
p3 <- mpse %>%
    mp_plot_rarecurve(
        .rare = RareAbundanceRarecurve,
        .alpha = Observe,
        .group = Group,
        plot.group = TRUE
    ) +
    scale_color_manual(values=colors) +
    scale_fill_manual(values=colors) +
    theme(legend.position=c(0.2, 0.8),
          legend.background=element_blank()
    )
p1 + p2 + p3
```
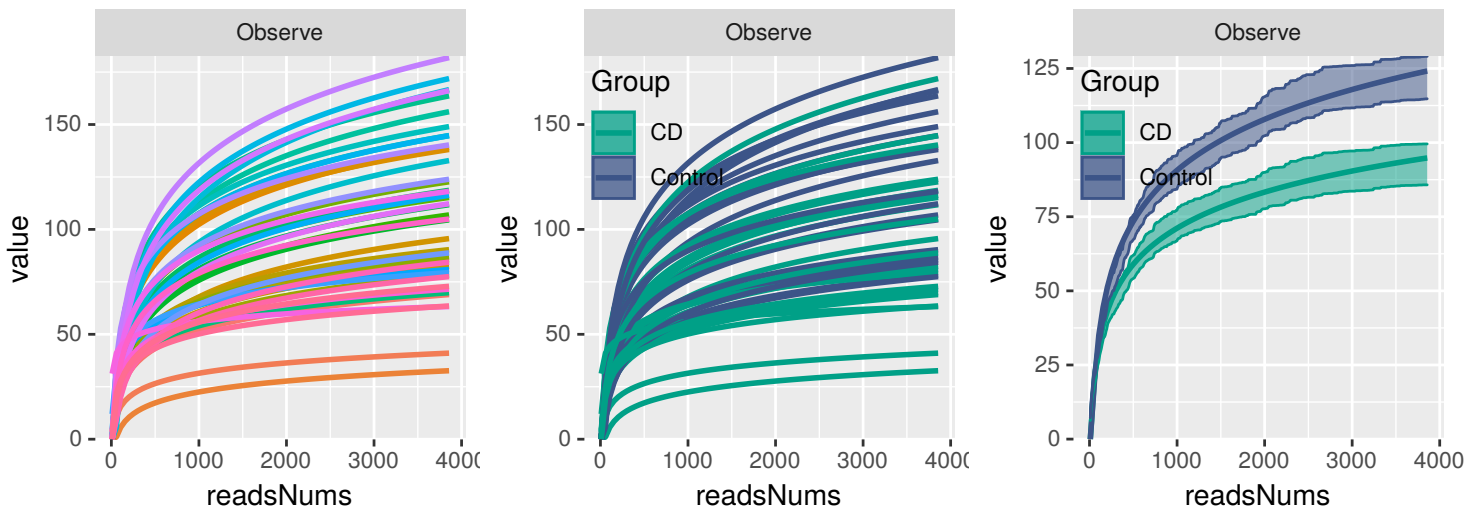
```
mpse %<>% mp_cal_alpha(.abundance=RareAbundance, action = "add")
mpse
```

```
## # A MPSE-tibble (MPSE object) abstraction: 67,295 x 19
## # OTU=1565 | Samples=43 | Assays=Abundance, RareAbundance | Taxanomy=Kingdom, Phylum, Class, Order, Family,
##    OTU      Sample  Abundance RareAbundance Group RareAbundanceRa~ Observe Chao1
##    <chr>    <chr>       <int>         <int> <chr> <I<list>>          <dbl> <dbl>
## 1  OTU_1497 S206700         0             0 CD    <tibble [2,574 ~      71  81.1
## 2  OTU_1005 S206700         0             0 CD    <tibble [2,574 ~      71  81.1
## 3  OTU_1333 S206700         0             0 CD    <tibble [2,574 ~      71  81.1
## 4  OTU_1663 S206700         0             0 CD    <tibble [2,574 ~      71  81.1
## 5  OTU_1665 S206700         0             0 CD    <tibble [2,574 ~      71  81.1
## 6  OTU_1190 S206700         0             0 CD    <tibble [2,574 ~      71  81.1
## 7  OTU_1449 S206700         0             0 CD    <tibble [2,574 ~      71  81.1
## 8  OTU_1583 S206700         0             0 CD    <tibble [2,574 ~      71  81.1
## 9  OTU_1443 S206700         0             0 CD    <tibble [2,574 ~      71  81.1
## 10 OTU_1035 S206700         0             0 CD    <tibble [2,574 ~      71  81.1
## # ... with 67,285 more rows, and 11 more variables: ACE <dbl>, Shannon <dbl>, Simpson <dbl>, J <dbl>,
## #   Kingdom <chr>, Phylum <chr>, Class <chr>, Order <chr>, Family <chr>,
## #   Genus <chr>, Species <chr>
```

```
p1 <- mpse %>%
     mp_plot_alpha(
       .group = Group,
       .alpha = c(Observe, ACE, Chao1, Shannon, Simpson, J),
     ) +
     scale_fill_manual(values=colors, guide="none") +
     scale_color_manual(values=colors, guide="none")
p1
```
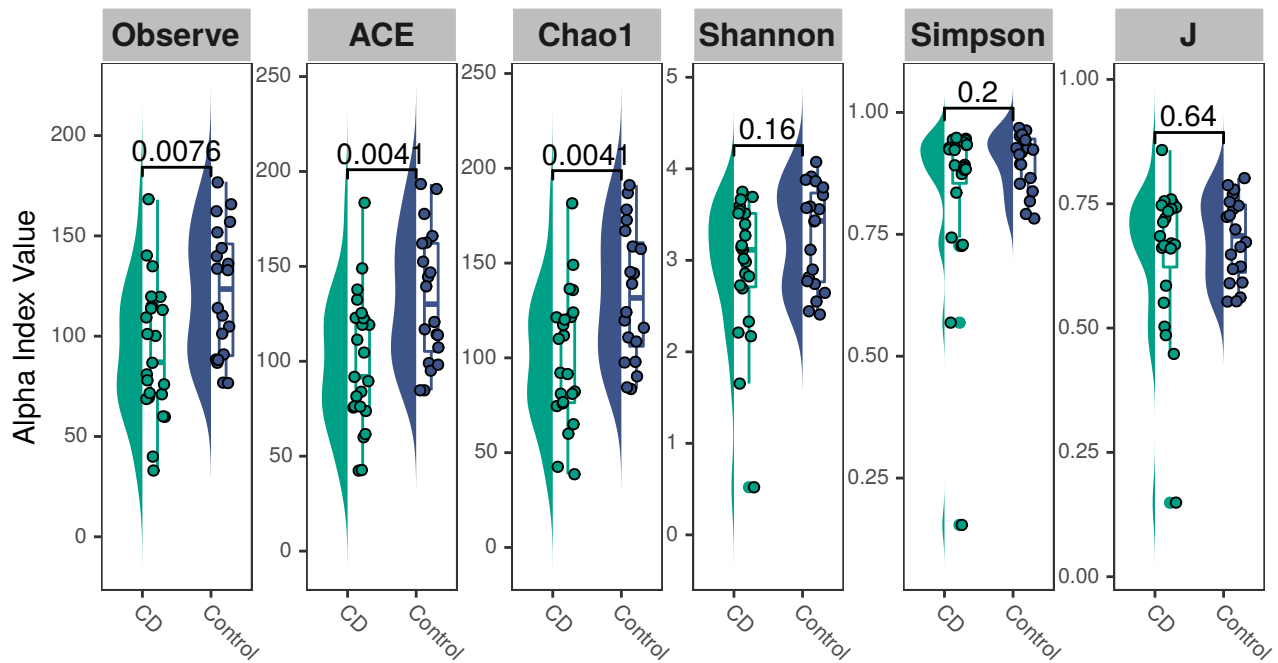
```
# The result can also be extracted, processed and visualized manually.
tbl.alpha <-
    mpse %>%
    mp_extract_sample()
tbl.alpha
```

```
## # A tibble: 43 x 9
##    Sample  Group   RareAbundanceRarec~ Observe Chao1   ACE Shannon Simpson     J
##    <chr>   <chr>   <I<list>>             <dbl> <dbl> <dbl>   <dbl>   <dbl> <dbl>
##  1 S206700 CD      <tibble [2,574 x 4~      71  81.1  84.0   2.86   0.892 0.670
##  2 S206701 CD      <tibble [2,574 x 4~      40  42.5  42.4   1.65   0.569 0.448
##  3 S206702 CD      <tibble [2,574 x 4~      33  38.6  42.7   0.520  0.154 0.149
##  4 S206703 Control <tibble [2,574 x 4~     136 145.  144.    3.55   0.918 0.723
##  5 S206704 Control <tibble [2,574 x 4~     133 139   139.    3.92   0.963 0.802
##  6 S206708 CD      <tibble [2,574 x 4~     100 112.  111.    2.69   0.835 0.585
##  7 S206709 CD      <tibble [2,574 x 4~      81  92.1  89.5   2.21   0.726 0.503
##  8 S206710 Control <tibble [2,574 x 4~      91  97.5  99.0   2.81   0.866 0.624
##  9 S206711 Control <tibble [2,574 x 4~      88 120.  114.    2.90   0.894 0.648
## 10 S206718 Control <tibble [2,574 x 4~      77  83.6  84.7   2.44   0.783 0.561
## # ... with 33 more rows
```

```
tbl.alpha %<>% dplyr::select(-c(RareAbundanceRarecurve))
tbl.alpha %<>%
  tidyr::pivot_longer(cols=!c("Sample", "Group"), names_to="measure", values_to="alpha") %>%
  dplyr::mutate(measure=forcats::fct_relevel(measure, c("Observe", "Chao1", "ACE", "Simpson", "Shannon", "J")))
```

```
library(ggplot2)
library(ggsignif)
library(gghalves)
p2 <- ggplot(data=tbl.alpha, aes(x=Group, y=alpha, fill=Group)) +
    geom_half_violin(color=NA, side="l", trim=FALSE) +
    geom_boxplot(aes(color=Group), fill=NA, position=position_nudge(x=.22), width=0.2) +
    geom_half_point(side="r", shape=21) +
    geom_signif(comparisons=list(c("CD", "Control")), test="wilcox.test", textsize=2, margin_top=0.1) +
    facet_wrap(facet=vars(measure), scales="free_y", nrow=1) +
    scale_fill_manual(values=colors, guide="none") +
    scale_color_manual(values=colors, guide="none")
p2
```
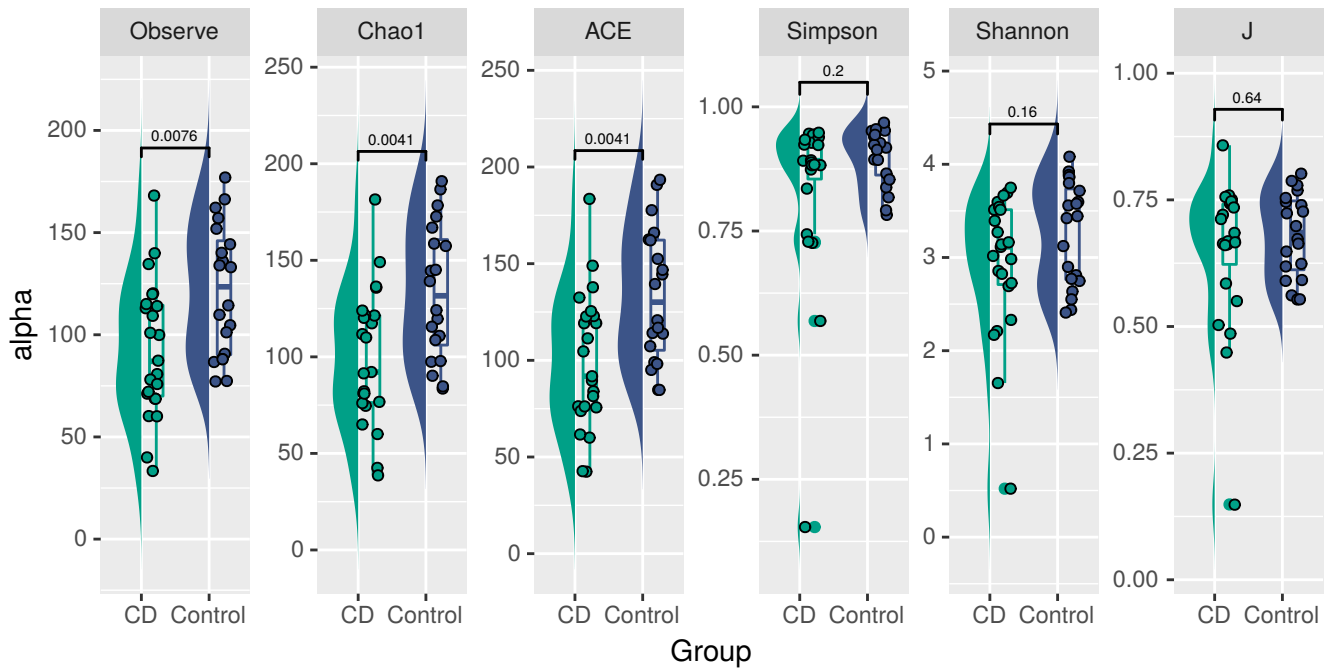
```r
library(ggupset)
library(ggVennDiagram)
mpse %<>%
      mp_cal_venn(.abundance=RareAbundance, .group=Group) %>%
      mp_cal_upset(.abundance=RareAbundance, .group=Group)

p1 <- mpse %>%
      mp_plot_venn(
        .group=Group,
        .venn=vennOfGroup
      ) +
      scale_color_manual(values=colors) +
      scale_fill_viridis_b(guide="none", option = "plasma")

p2 <- mpse %>%
      mp_plot_upset(
        .group=Group,
        .upset=ggupsetOfGroup
      )

dat <- tibble::tibble(x=1, y=0.8, label=list(p1))

f <- p2 +
     ggpp::geom_plot_npc(
        data = dat,
        mapping = aes(label=label, npcx=x, npcy=y),
        vp.width = 0.6,
        vp.height = 0.3,
        hjust=0.8
     )
f
```
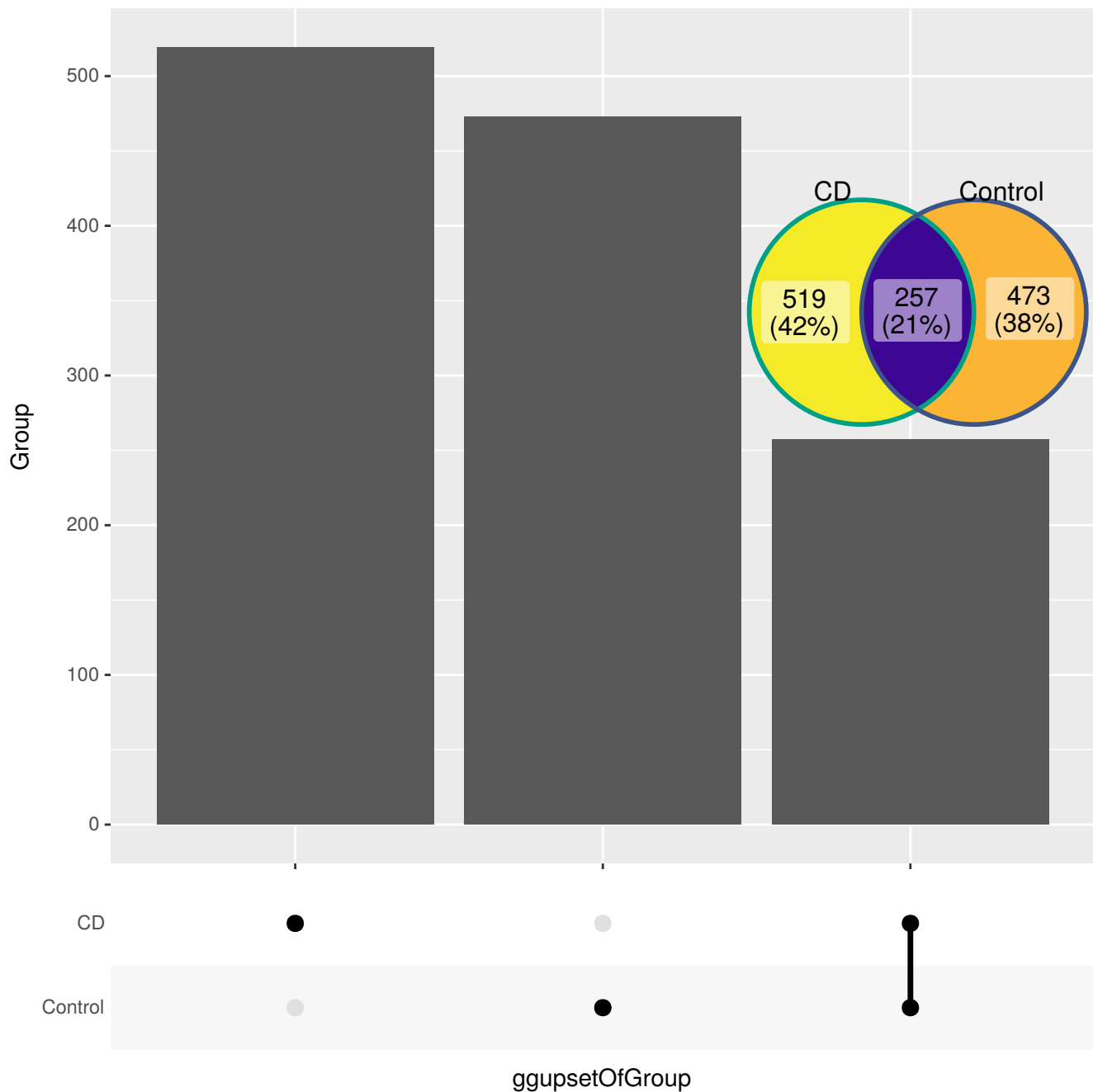
```
# visualized the result manually.
tbl.upset <- mpse %>% mp_extract_feature
p1 <- tbl.upset %>%
        dplyr::filter(vapply(ggupsetOfGroup, length, numeric(1))>0) %>%
        ggplot(aes(x=ggupsetOfGroup)) +
        geom_bar() +
        ggupset::scale_x_upset() +
        ggupset::theme_combmatrix(combmatrix.label.extra_spacing=40)

tbl.venn <- mpse %>%
        mp_extract_sample() %>%
        dplyr::select(c(Group, vennOfGroup)) %>%
        dplyr::distinct()
tbl.venn

## # A tibble: 2 x 2
##   Group   vennOfGroup
##   <chr>   <I<list>>
## 1 CD      <chr [776]>
```
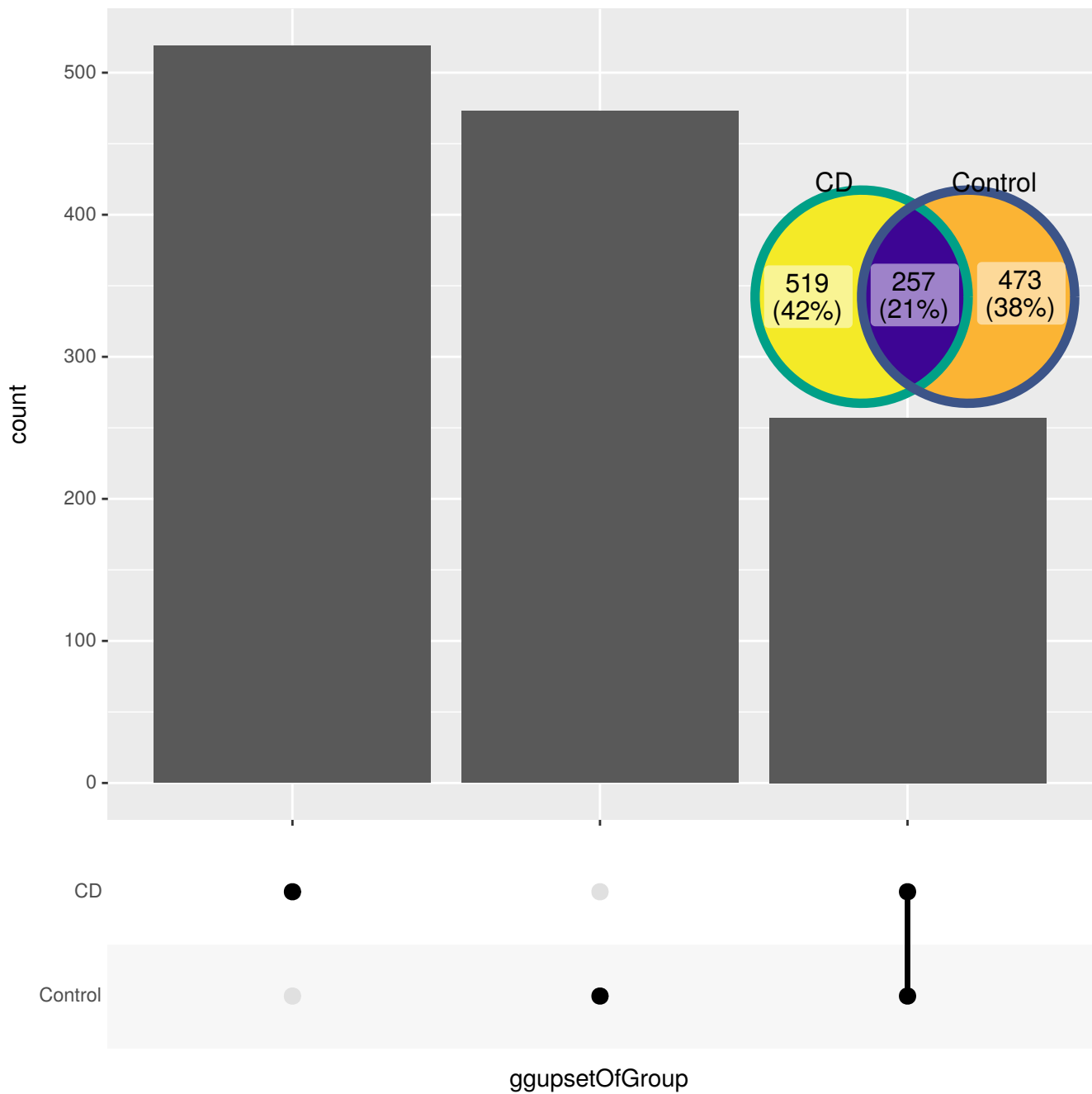
```
## 2 Control <chr [730]>
```

```r
p2 <- tbl.venn %>%
      dplyr::pull(var=vennOfGroup, name=Group) %>%
      ggVennDiagram(edge_size=2) +
      scale_color_manual(values=colors) +
      scale_fill_viridis_b(guide="none", option = "plasma")

dat <- tibble::tibble(x=1, y=0.8, label=list(p2))
dat
```

```
## # A tibble: 1 x 3
##       x     y label
##   <dbl> <dbl> <list>
## 1     1   0.8 <gg>
```

```r
p <- p1 + ggpp::geom_plot_npc(data = dat,
                              mapping = aes(label=label, npcx=x, npcy=y),
                              vp.width = 0.6,
                              vp.height = 0.3,
                              hjust=0.8)

p
```

```
mpse %<>%
    mp_decostand(
      .abundance = Abundance,
      method = 'hellinger'
    )
mpse
```

```
## # A MPSE-tibble (MPSE object) abstraction: 67,295 x 22
## # OTU=1565 | Samples=43 | Assays=Abundance, RareAbundance, hellinger | Taxanomy=Kingdom, Phylum, Class, Orde
##    OTU   Sample Abundance RareAbundance hellinger Group RareAbundanceRa~ Observe
##    <chr> <chr>      <int>         <int>     <dbl> <chr> <I<list>>           <dbl>
## 1  OTU_~ S2067~         0             0         0 CD    <tibble [2,574 ~       71
## 2  OTU_~ S2067~         0             0         0 CD    <tibble [2,574 ~       71
## 3  OTU_~ S2067~         0             0         0 CD    <tibble [2,574 ~       71
## 4  OTU_~ S2067~         0             0         0 CD    <tibble [2,574 ~       71
## 5  OTU_~ S2067~         0             0         0 CD    <tibble [2,574 ~       71
## 6  OTU_~ S2067~         0             0         0 CD    <tibble [2,574 ~       71
## 7  OTU_~ S2067~         0             0         0 CD    <tibble [2,574 ~       71
## 8  OTU_~ S2067~         0             0         0 CD    <tibble [2,574 ~       71
```

```
##  9 OTU_~ S2067~          0             0          0 CD     <tibble [2,574 ~       71
## 10 OTU_~ S2067~          0             0          0 CD     <tibble [2,574 ~       71
## # ... with 67,285 more rows, and 14 more variables: Chao1 <dbl>, ACE <dbl>, Shannon <dbl>,
## #   Simpson <dbl>, J <dbl>, vennOfGroup <I<list>>, ggupsetOfGroup <I<list>>,
## #   Kingdom <chr>, Phylum <chr>, Class <chr>, Order <chr>, Family <chr>,
## #   Genus <chr>, Species <chr>
```

```r
mpse %<>% mp_cal_dist(.abundance=hellinger, distmethod="bray")
mpse
```

```
## # A MPSE-tibble (MPSE object) abstraction: 67,295 x 23
## # OTU=1565 | Samples=43 | Assays=Abundance, RareAbundance, hellinger | Taxanomy=Kingdom, Phylum, Class, Orde
##      OTU   Sample Abundance RareAbundance hellinger Group RareAbundanceRa~ Observe
##      <chr> <chr>      <int>         <int>     <dbl> <chr> <I<list>>           <dbl>
## 1  OTU_~ S2067~          0             0          0 CD     <tibble [2,574 ~       71
## 2  OTU_~ S2067~          0             0          0 CD     <tibble [2,574 ~       71
## 3  OTU_~ S2067~          0             0          0 CD     <tibble [2,574 ~       71
## 4  OTU_~ S2067~          0             0          0 CD     <tibble [2,574 ~       71
## 5  OTU_~ S2067~          0             0          0 CD     <tibble [2,574 ~       71
## 6  OTU_~ S2067~          0             0          0 CD     <tibble [2,574 ~       71
## 7  OTU_~ S2067~          0             0          0 CD     <tibble [2,574 ~       71
## 8  OTU_~ S2067~          0             0          0 CD     <tibble [2,574 ~       71
## 9  OTU_~ S2067~          0             0          0 CD     <tibble [2,574 ~       71
## 10 OTU_~ S2067~          0             0          0 CD     <tibble [2,574 ~       71
## # ... with 67,285 more rows, and 15 more variables: Chao1 <dbl>, ACE <dbl>, Shannon <dbl>,
## #   Simpson <dbl>, J <dbl>, vennOfGroup <I<list>>, bray <I<list>>,
## #   ggupsetOfGroup <I<list>>, Kingdom <chr>, Phylum <chr>, Class <chr>,
## #   Order <chr>, Family <chr>, Genus <chr>, Species <chr>
```

```r
p1 <- mpse %>%
     mp_plot_dist(
       .distmethod = bray,
       .group = Group
     )

p2 <- mpse %>%
     mp_plot_dist(
       .distmethod = bray,
       .group = Group,
       group.test = TRUE
     ) +
     scale_color_brewer(palette="Dark2",guide="none")

aplot::plot_list(p1, p2, widths=c(3, 1))
```
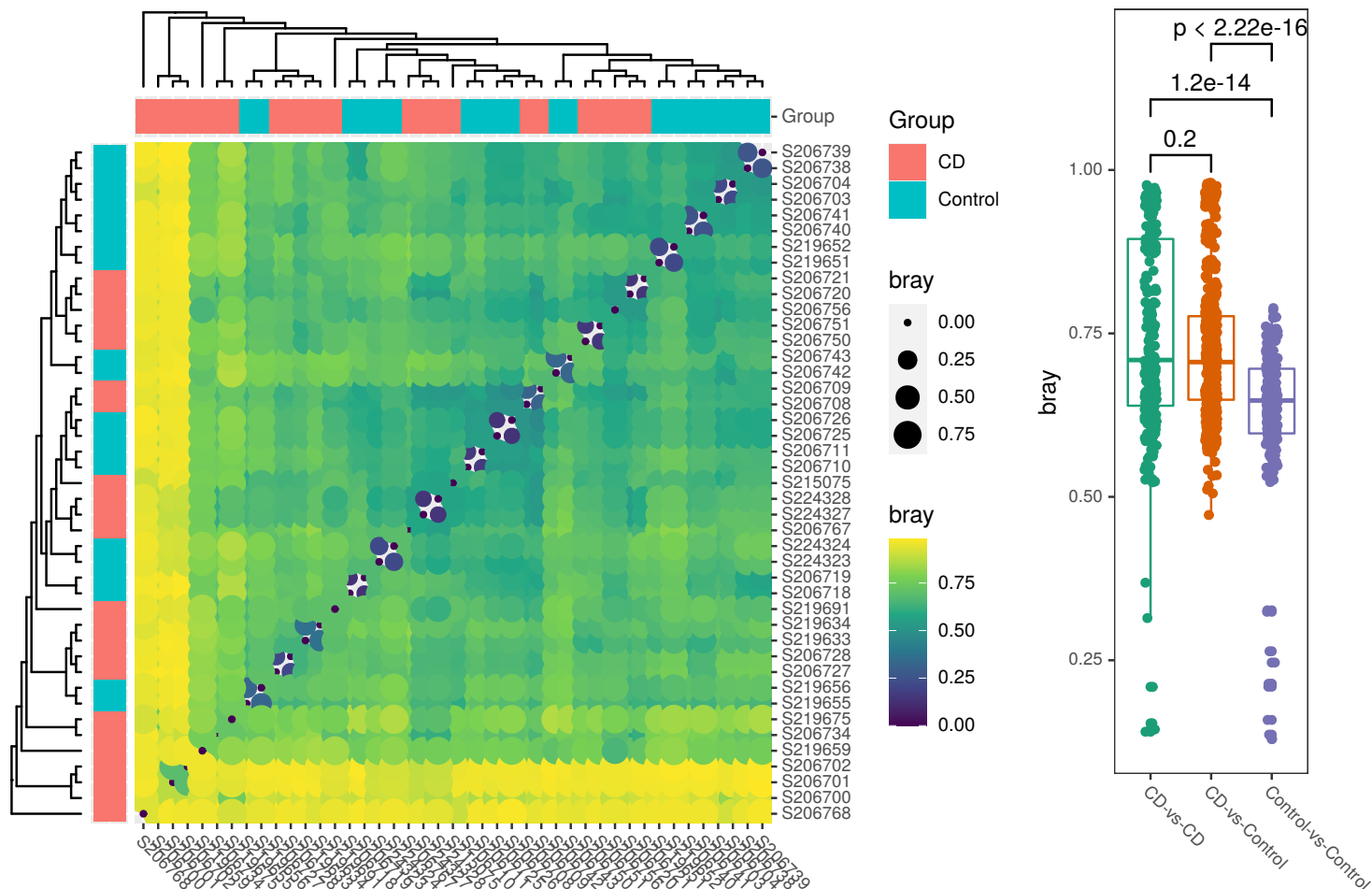
```
# visualized the result manually
tbl <- mpse %>%
        mp_extract_dist(distmethod="bray", .group=Group)
tbl
```

```
## # A tibble: 903 x 4
## # Rowwise:
##    Sample   bray GroupsComparison braySampley
##    <chr>   <dbl> <chr>            <chr>
##  1 S206700 0.874 CD-vs-CD         S206701
##  2 S206700 0.873 CD-vs-CD         S206702
##  3 S206700 0.790 CD-vs-Control    S206703
##  4 S206700 0.817 CD-vs-Control    S206704
##  5 S206700 0.875 CD-vs-CD         S206708
##  6 S206700 0.879 CD-vs-CD         S206709
##  7 S206700 0.849 CD-vs-Control    S206710
##  8 S206700 0.856 CD-vs-Control    S206711
##  9 S206700 0.891 CD-vs-Control    S206718
## 10 S206700 0.889 CD-vs-Control    S206719
## # ... with 893 more rows
```

```
library(ggridges)
p <- tbl %>%
    ggplot(aes(x = bray,
               y = GroupsComparison,
               fill = GroupsComparison)
    ) +
    geom_density_ridges(jittered_points = TRUE,
                        position = "points_sina",
                        alpha = 0.8,
                        point_shape = 23,
```

```
                            point_size = 1,
                            point_color = "black",
                            color = "grey50"
      ) +
      geom_signif(comparisons=utils::combn(unique(tbl$GroupsComparison), 2) %>%
                            apply(2, list) %>%
                            unlist(recursive = FALSE),
                orientation="y",
                margin_top = 0.2,
                step_increase = 0.1,
      ) +
      labs(x="bray distance", y=NULL) +
      scale_fill_brewer(palette="Dark2",guide="none")
p
```

Picking joint bandwidth of 0.0318