annotate

2022-10-17

Load VDJdb

```
database <- read_tsv("vdjdb.slim.txt")
```

```
## Rows: 68061 Columns: 16
## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## chr (12): gene, cdr3, species, antigen.epitope, antigen.gene, antigen.specie...
## dbl  (3): v.end, j.start, vdjdb.score
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Function to build 1mm graph

```
make_graph_1mm <- function(cdrs, no_singletons = T) {
  set.seed(42)

  # hamming = 1 graph
  mm <- stringdistmatrix(cdrs,
                         method = "hamming",
                         useNames = "strings") %>%
    as.matrix()
  mm[mm != 1] <- 0
  gg <- graph_from_adjacency_matrix(mm, mode = "undirected") %>%
    simplify()

  # connected components
  cc <- clusters(gg)
  result <- tibble(cdr3 = names(cc$membership),
                   cdr_cluster = paste0("C", cc$membership)) %>%
    group_by(cdr_cluster) %>%
    mutate(cdr_cluster_sz = n()) %>%
    ungroup()

  # layout components with 2+ nodes
  if (no_singletons) {
    gg <- delete.vertices(gg, which(degree(gg) == 0))
  }

  coords <- gg %>%
    layout_with_graphopt(niter = 3000, charge = 0.005)

  # put together
  result %>%
    left_join(tibble(
```

```
      cdr3 = names(V(gg)),
      cdr_graph_x = coords[, 1],
      cdr_graph_y = coords[, 2]
    ))
}
```

Build CDR3 homology graphs for mouse alpha and beta chains

```
cdrs.graph <- database %>%
  filter(species == "MusMusculus") %>%
  group_by(gene, len = nchar(cdr3)) %>%
  group_modify(~ make_graph_1mm(.x$cdr3 %>% unique, F))
```

```
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
## Joining, by = "cdr3"
```

Plot graph

```
database.plt <- database %>%
  merge(cdrs.graph) %>%
  filter(species == "MusMusculus") %>%
  group_by(gene, antigen.epitope) %>%
  mutate(antigen.epitope = ifelse(length(unique(cdr3)) >= 100,
                                  antigen.epitope, NA))

database.plt %>%
  filter(!is.na(antigen.epitope)) %>%
  ggplot(aes(x = cdr_graph_x,
             y = cdr_graph_y)) +
  geom_point(data = database.plt %>%
               filter(is.na(antigen.epitope)), color = "grey25") +
  geom_point(aes(color = antigen.epitope), alpha = 0.95) +
  scale_color_brewer(palette = "Set3") +
  facet_wrap(~ gene) +
```

```
  theme_void() +
  theme(aspect = 1, legend.position = "bottom")
```