

MetaDAG experiment example

Version beta 0.2

Invalid Date

Table of contents

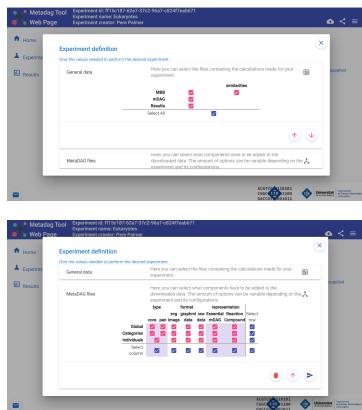
MetaDag Eukaryotes experiment

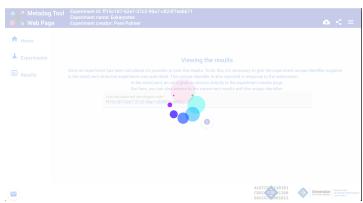
This is an example of an experiment of [metaDag](#) data with results.

Hash: ff15c187-62e7-37c2-96a7-c824f7eab671

URL: <http://bioinfo.uib.es/metadag/results?uuid=ff15c187-62e7-37c2-96a7-c824f7eab671>

```
library(tidyverse)
library(igraph)
library(ComplexHeatmap)
library(viridis)
library(circlize)
library(plotly)
library(randomcoloR)
library(factoextra)
library(RColorBrewer)
library(kableExtra)
library(igraph)
library(GGally)
```





Load data MetaDag experiment

```
experiment=
"results_ff15c187-62e7-37c2-96a7-c824f7eab671"
path_exp=paste0("data/",experiment,"/data/")
knitr::kable(data.frame(Directory_files_and_folders=dir(path_exp),
                        Type=c(rep("Data file",2),
                               rep("Directory",2),
                               rep("Data file",6),
                               "Directory")))
```

Directory_files_and_folders	Type
Different_MBB.csv	Data file
Different_mDAG.csv	Data file
Global	Directory
Individuals	Directory
Results.csv	Data file
Similarities_MBB_MSAMethod.csv	Data file
Similarities_MBB_MunkresMethod.csv	Data file
Similarities_mDAG_MSAMethod.csv	Data file
Similarities_mDAG_MunkresMethod.csv	Data file
Similarities_mDAGOnReaction.csv	Data file
TaxonomyLevels	Directory

```
MBB=read_csv(paste0(path_exp,"Different_MBB.csv"),show_col_types = FALSE)
_mDAG=read_csv(paste0(path_exp,"Different_mDAG.csv"),show_col_types = FALSE)
Results=read_csv(paste0(path_exp,"Results.csv"),show_col_types = FALSE)
save(MBB,mDAG,Results,file="MBB_mDAG_Results.Rdata")
```

MBB

In this experiment MBB is a table with 5112 rows and 4106 columns.

```
knitr::kable(MBB[1:20,1:30]) %>% scroll_box(width = "100%", height = "200px")
```

MBB Id	natural	#pathways	Protists	Fungi	Plants	Animals	Alveolates	Amoebozoa	Annelids
0	0	0	0	0	0	0	0	0	0
0.0	0	0	0	0	0	0	0	0	0
0.0.0	1	1	1	0	0	0	0	0	0
0.0.0.0	1	1	1	0	0	0	0	0	0
0.0.1	1	1	1	0	0	0	0	0	0
0.0.1.0	1	86	2	77	1	6	1	0	0
0.0.1.1	1	86	2	77	1	6	1	0	0
0.0.1.2	1	15	15	0	0	0	12	0	0
0.0.1.3	1	5	5	0	0	0	0	4	0
0.0.1.4	0	0	0	0	0	0	0	0	0
0.0.1.4.0	1	609	14	128	128	339	1	3	1
0.0.1.4.1	1	609	14	128	128	339	1	3	1
0.0.1.4.2	1	322	6	92	133	91	1	0	0
0.0.1.4.3	1	322	6	92	133	91	1	0	0
0.0.1.4.4	1	1	0	0	1	0	0	0	0
0.0.1.4.5	1	1	1	0	0	0	0	1	0
0.0.1.5	0	0	0	0	0	0	0	0	0
0.0.1.5.0	1	5	5	0	0	0	0	4	0
0.0.2	1	1	1	0	0	0	0	0	0
0.0.3	1	1	1	0	0	0	0	1	0

mDAG

Abstract/unique mDAG's in this experiment

```
dim(mDAG)
```

```
[1] 884 5224
```

```
kable(mDAG[1:20,1:30]) %>% scroll_box(width = "100%", height = "200px")
```

mDAG Id	#Categories	Animals	Plants	Fungi	Protists	Alveolates	Amoebozoa	Annelids	Arth
0001	3	1	0	0	0	0	0	0	
0002	2	0	0	1	0	0	0	0	
0003	2	1	0	0	0	0	0	0	
0004	3	1	0	0	0	0	0	0	
0005	3	1	0	0	0	0	0	0	
0006	3	0	1	0	0	0	0	0	
0007	2	0	1	0	0	0	0	0	
0008	3	0	1	0	0	0	0	0	
0009	3	0	1	0	0	0	0	0	
0010	3	1	0	0	0	0	0	0	
0011	3	1	0	0	0	0	0	0	
0012	3	0	0	0	1	0	0	0	
0013	3	1	0	0	0	0	0	0	
0014	3	0	0	0	1	1	0	0	
0015	2	0	0	1	0	0	0	0	
0016	3	0	0	0	1	0	1	0	
0017	3	1	0	0	0	0	0	0	
0018	3	1	0	0	0	0	0	0	
0019	3	1	0	0	0	0	0	0	
0020	3	1	0	0	0	0	0	0	

```
dim(mDAG)
```

```
[1] 884 5224
```

```
names(mDAG)[1:6]
```

```
[1] "mDAG_Id"      "#Categories" "Animals"       "Plants"        "Fungi"
[6] "Protists"
```

```
head(names(mDAG)[7:(dim(mDAG)[2]-1150)])# 28 to 1213 code MBB: 1 if MBB in mDAG 0
```

```
[1] "Alveolates"      "Amoebozoa"      "Annelids"
[4] "Arthropods"       "Ascomycetes"    "Basal_angiosperms"
```

Results

Tabular data Results for this experiment

```

kable(Results[1:20,1:10])%>%
  row_spec(0, angle = 0) %>%
  scroll_box(width = "300%", height = "1000px")

```

organism	Categories	mDAG Id	Full Name
aaf	Protists Stramenopiles Pelagophytes	0036	Aureococcus anophagefferens
aag	Animals Arthropods Insects	0035	Aedes aegypti (yellow fever mosquito)
aalb	Animals Arthropods Insects	0276	Aedes albopictus (Asian tiger mosquito)
aali	Animals Arthropods Insects	0267	Anopheles albimanus
aalt	Fungi Ascomycetes Dothideomycetes	0240	Alternaria alternata
aam	Animals Vertebrates Birds	0040	Apteryx mantelli mantelli (North Island brown kiwi)
aamp	Animals Vertebrates Mammals	0313	Arvicola amphibius (Eurasian water vole)
aang	Animals Vertebrates Fishes	0317	Anguilla anguilla (European eel)
aara	Animals Arthropods Insects	0362	Anopheles arabiensis
abe	Fungi Ascomycetes Eurotiomycetes	0060	Trichophyton benhamiae
abp	Fungi Basidiomycetes	0068	Agaricus bisporus var. burnettii JB137-S8
abv	Fungi Basidiomycetes	0073	Agaricus bisporus var. bisporus H97
acan	Protists Amoebozoa Acanthamoeba	0873	Acanthamoeba castellanii
acar	Animals Vertebrates Birds	0884	Antrostomus carolinensis (chuck-will's-widow)
acep	Animals Arthropods Insects	0054	Atta cephalotes (leaf cutting ant)
acer	Animals Arthropods Insects	0057	Apis cerana (Asiatic honeybee)
achc	Animals Vertebrates Birds	0103	Aquila chrysaetos chrysaetos (golden eagle)
ache	Fungi Ascomycetes Eurotiomycetes	0106	Aspergillus chevalieri
achl	Animals Vertebrates Birds	0081	Acanthisitta chloris (rifleman)
acoz	Animals Arthropods Insects	0255	Anopheles coluzzii

```

dim(Results)

```

```

[1] 884 3997

```

```

names(Results)[1]# organisms  kegg id  class representant of mDAG

```

```

[1] "organism"

```

```

names(Results)[2]# taxonomy separate by |

```

```

[1] "Categories"

```

```

names(Results)[3]# mDAG_Id

[1] "mDAG_Id"

names(Results)[4]# Full name representant

[1] "Full Name"

names(Results)[6:36]# columns 6 to 2726 Variable Reactions name id: This reaction belongs

[1] "R00710_rev(1.2.1.3)"   "R00711(1.2.1.5)"      "R00711_rev(1.2.1.5)"
[4] "R00755(4.1.1.1)"       "R00746(1.1.1.2)"      "R00746_rev(1.1.1.2)"
[7] "R00754(1.1.1.1)"       "R00754_rev(1.1.1.1)"    "R00014(4.1.1.1)"
[10] "R00014(1.2.4.1)"      "R03270(1.2.4.1)"      "R02569(2.3.1.12)"
[13] "R02569_rev(2.3.1.12)"  "R00703(1.1.1.27)"    "R00703_rev(1.1.1.27)"
[16] "R00200(2.7.1.40)"      "R00658(4.2.1.11)"      "R00658_rev(4.2.1.11)"
[19] "R01518(5.4.2.11)"      "R01518_rev(5.4.2.11)"  "R01061(1.2.1.12)"
[22] "R01061_rev(1.2.1.12)"  "R01015(5.3.1.1)"      "R01015_rev(5.3.1.1)"
[25] "R01070(4.1.2.13)"      "R01070_rev(4.1.2.13)"  "R04779(2.7.1.11)"
[28] "R04780(3.1.3.11)"      "R02740(5.3.1.9)"      "R02740_rev(5.3.1.9)"
[31] "R00959(5.4.2.2)"

reactions=names(Results)[-c(1:5)]
reverse_reactions=stringr::str_detect(reactions,"rev")
reverse_reactions=table(reverse_reactions)
dimnames(reverse_reactions)$reverse_reactions=c("One direction","Reverse reaction")
reverse_reactions

reverse_reactions
One direction Reverse reaction
3398           594

```

Graphs

Read graphs in graphml format

Individuals graph for each organism

Read individual graph for Homo sapiens (Keg id: hsa) at directory(Individuals/hsa)

```
files_hsa=dir(paste0(path_exp,"Individuals/hsa"))
files_hsa

[1] "hsa_mDAG.graphml"      "hsa_mDAG.pdf"          "hsa_mDAG.svg"
[4] "hsa_mDAG_adj.csv"     "hsa_mDAG_biggerDAG.pdf" "hsa_mDAG_biggerDAG.svg"
[7] "hsa_mDAG_essential.pdf" "hsa_mDAG_essential.svg" "hsa_mDAG_nl.csv"
[10] "hsa_mDAG_structure.csv" "hsa_R_adj.csv"        "hsa_R_nl.csv"
[13] "hsa_RC.graphml"       "hsa_RC.pdf"           "hsa_RC.svg"
```

files_Individual_hsa	Description
hsa_mDAG.graphml	graph mDAG graphml format
hsa_mDAG.pdf	plot mDAG pdf format
hsa_mDAG.svg	plot mDAG svg format
hsa_mDAG_adj.csv	csv graph adjacency matrix of mDAG
hsa_mDAG_biggerDAG.pdf	plot bigger connected component mDAG pdf format
hsa_mDAG_biggerDAG.svg	plot bigger connected component mDAG svg format
hsa_mDAG_essential.pdf	plot essential MBB mDAG pdf format
hsa_mDAG_essential.svg	plot essential MBB mDAG svg format
hsa_mDAG_nl.csv	csv names labels MBB vertex of mDAG
hsa_mDAG_structure.csv	csv matrix membership of reactions of each MBB of mDAG ordered by decre
hsa_R_adj.csv	csv graph adjacency matrix of RC graph
hsa_R_nl.csv	csv names labels MBB vertex of RC
hsa_RC.graphml	graph RC graphml format
hsa_RC.pdf	plot RC pdf format
hsa_RC.svg	plot RC svg format

```

graph_mDAG=read.graph(paste0(path_exp,"Individuals/hsa/hsa_mDAG.graphml"),format = "graphml")
summary(graph_mDAG)

IGRAPH c38fa60 D--- 1026 1086 --
+ attr: color (v/c), label (v/c), id (v/c), id (e/c)

barplot(table(igraph::degree(graph_mDAG,mode="all")),
        ylim=c(0,350),col="blue",
        main="Frequency of nodes degree",
        ylab="Frequency",xlab="degree")




| degree | Frequency |
|--------|-----------|
| 0      | ~100      |
| 1      | ~350      |
| 2      | ~350      |
| 3      | ~150      |
| 4      | ~50       |
| 5      | ~10       |
| 6      | ~5        |
| 7      | ~2        |
| 8      | ~1        |
| 9      | ~1        |
| 11     | ~1        |
| 15     | ~1        |


```

mDag Global core for eukaryotes

Note that the mDAG core is empty as it does not contain any reactions.

Core mDAG

```

graph_core_mDAG=read.graph(
  paste0(path_exp,"Global/core/core_mDAG.graphml"), format = "graphml")
summary(graph_core_mDAG)

IGRAPH c47f642 D--- 0 0 --
+ attr: color (v/c), label (v/c), id (v/c)

knitr::include_graphics(
  paste0(path_exp,"Global/core/core_mDAG.pdf"))

```

Figure 1: Core mDAG is empty

The graph core mDAG have 0 vertex and 0, is an empty graph.

Core reaction graph (RC)

```
graph_core_RC=read.graph(
  paste0(path_exp,"Global/core/core_RC.graphml"), format = "graphml")
summary(graph_core_mDAG)

IGRAPH c47f642 D--- 0 0 --
+ attr: color (v/c), label (v/c), id (v/c)

knitr::include_graphics(
  paste0(path_exp,"Global/core/core_RC.pdf"))
```

Figure 2: Core mDAG is empty

The graph core reaction graph have 0 vertex and 0, is an empty graph.

mDag Global pan for eukaryotes

Pan mDAG

```
graph_pan_mDAG=read.graph(
  paste0(path_exp,"TaxonomyLevels/Kingdom/Animals/pan/Animals_pan_mDAG.graphml"), format =
  "graph TD"; summary(graph_pan_mDAG)
```

```
IGRAPH c49f6d2 D--- 1184 1261 --
+ attr: color (v/c), label (v/c), id (v/c), id (e/c)
```

The graph pan mDAG have 1184 vertex and 1261.

Pan Reaction graph (RC)

```
graph_pan_RC=read.graph(
  paste0(path_exp,"TaxonomyLevels/Kingdom/Animals/pan/Animals_pan_RC.graphml"), format =
  "graph TD"; summary(graph_pan_RC)
```

```
IGRAPH c4abfa6 D--- 4556 5798 --
+ attr: color (v/c), label (v/c), id (v/c), id (e/c)
```

The graph pan reaction graph have 4556 vertex and `igraph::ecount(graph_pan_RC)`:

```
compo=components(graph_mDAG, mode = "weak")
str(compo)
```

```
List of 3
$ membership: num [1:1026] 1 1 1 1 1 1 1 1 1 ...
$ csizes     : num [1:167] 589 1 1 1 1 1 4 3 4 3 ...
$ no         : int 167
```

```
compo$csizes
```

```

[1] 589   1   1   1   1   1   1   4   3   4   3   2   3   3   1   1   1   1   2   6
[19]   3   1   3   6   1   1   1   1   1   3   1   6   2   1   1   1   1   1   1   2   1
[37]   1  14   1  16   1   6   2   2   4   1   1   1   1   1   1   1   1   1   1   1
[55]  13   1   1   1   1   2   6   5   5   2   2   10  1   1   1   1   1   2   2   1
[73]   1   1  62   6   2   1   2   1   1   1   2   1   2   14  3   1   1   1   1   1
[91]   1   1   1   1   1   1   3   6   1   3   1   3   2   1   1   2   3   4   1   1   2
[109]   3   1   1   2   5   1   1   2   3   2   1   1   2   3   4   1   1   1   1   2
[127]   1   1   2   1   1   1   3   1   2   2   1   6   1   1   1   2   1   3   1
[145]   1   1   1   1   1   2   7   1  15  3   1   1   1   1   2   1   1   1
[163]   1   1   1   1   2

```

```

k=which.max(compo$csizes==max(compo$csizes))
k

```

```
[1] 1
```

```
table(compo$membership)
```

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
589	1	1	1	1	1	4	3	4	3	2	3	3	1	1	1	1	2	6	3	1
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	
3	6	1	1	1	1	1	3	1	6	2	1	1	1	2	1	1	14	1	16	
41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	
1	6	2	2	4	1	1	1	1	1	1	1	1	1	13	1	1	1	1	2	
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	
6	5	5	2	2	10	1	1	1	2	2	1	1	1	62	6	2	1	2	1	
81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	
1	1	2	1	2	14	3	1	1	1	1	1	1	1	1	1	1	3	6	1	3
101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	
1	3	2	1	1	1	2	2	3	1	1	2	5	1	1	2	3	2	1	1	
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	
2	3	4	1	1	2	1	1	2	1	1	1	1	1	3	1	2	2	1	6	
141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	
1	1	1	2	1	1	1	1	1	2	7	1	15	3	1	1	1	1	2	1	
161	162	163	164	165	166	167														
3	1	1	1	1	1	1	2													

```

vertex=which(compo$membership==k)
length(vertex)

[1] 589

Big_Component=induced_subgraph(graph_mDAG, vids=vertex)
igraph::vcount(Big_Component)

[1] 589

igraph::ecount(Big_Component)

[1] 774

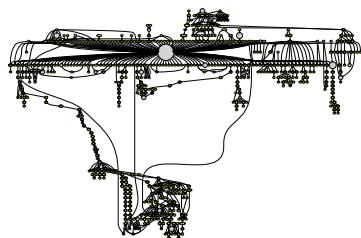
```

The curated plot of the bigger component of hsa mDAG

```

knitr::include_graphics(paste0(path_exp,
                               "Individuals/hsa/hsa_mDAG_biggerDAG.pdf"))

```



Similarities and metadata for an experiment

We will first load the metadata and adapt them to the structure of the similarities to facilitate the creation of the graphs and statistics.

Remember de path of the experiment:

```
path_exp
```

```
[1] "data/results_ff15c187-62e7-37c2-96a7-c824f7eab671/data/"
```

Load meta data from eukariotes experimet

Meta data mDa_Id and taxonomy sort by Kingdom,Filum,Class,mDAG_Id

```
path_exp
```

```
[1] "data/results_ff15c187-62e7-37c2-96a7-c824f7eab671/data/"
```

```
Results=read_csv(paste0(path_exp,"Results.csv"))
names(Results)[c(1,3,4)]=c("Organism","mDAG_Id","Full_Name")
#code=Results %>% select(Organism:mDAG_Id)
taxo=Results %>% select(Organism:Full_Name)
index=is.na(taxo$Categories)
taxo=taxo %>% separate(Categories,into=c("Kingdom","Phylum","Class"))
taxo$Class[index]=paste(taxo$Kingdom[index],taxo$Phylum[index])
meta_taxo=taxo %>% arrange(Kingdom,Phylum)
index=which(is.na(meta_taxo$Class))
meta_taxo$Class[index]=meta_taxo$Phylum[index]
aux=table(meta_taxo$Class)
freq=tibble(Class=names(aux),Freq_Class=aux)
names(freq)=c("Class","Freq_Class")
meta_taxo =meta_taxo %>% left_join(freq)%>%
  arrange(Kingdom,Phylum,desc(Freq_Class))
```

```

head(meta_taxo)

# A tibble: 6 x 7
  Organism Kingdom Phylum      Class    mDAG_Id Full_Name          Freq_Class
  <chr>     <chr>   <chr>      <chr>    <chr>    <chr>                <table[1d>
1 hro       Animals Annelids  Annelids  0576    Helobdella robusta    1
2 aag       Animals Arthropods Insects  0035    Aedes aegypti (yellow~ 130
3 aalb      Animals Arthropods Insects  0276    Aedes albopictus (Asi~ 130
4 aali      Animals Arthropods Insects  0267    Anopheles albimanus   130
5 aara      Animals Arthropods Insects  0362    Anopheles arabiensis  130
6 acep      Animals Arthropods Insects  0054    Atta cephalotes (leaf~ 130

table(meta_taxo$Kingdom) %>% kable %>%
  kable_styling("striped", full_width = F, position="left")%>%
  scroll_box(width = "400px", height = "200px")

```

Var1	Freq
Animals	535
Fungi	154
Plants	139
Protists	56

```

table(meta_taxo$Phylum,meta_taxo$Kingdom) %>% kable %>%
  kable_styling("striped", full_width = F, position="left")%>%
  scroll_box(width = "500px", height = "500px")

```

	Animals	Fungi	Plants	Protists
Alveolates	0	0	0	25
Amoebozoa	0	0	0	7
Annelids	1	0	0	0
Arthropods	158	0	0	0
Ascomycetes	0	113	0	0
Basal	0	0	2	0
Basidiomycetes	0	36	0	0
Brachiopoda	1	0	0	0
Cephalochordates	2	0	0	0
Choanoflagellates	0	0	0	2
Cnidarians	10	0	0	0
Cryptomonads	0	0	0	1
Echinoderms	3	0	0	0
Eudicots	0	0	98	0
Euglenozoa	0	0	0	9
Ferns	0	0	1	0
Flatworms	4	0	0	0
Green	0	0	11	0
Haptophyta	0	0	0	1
Hemichordates	1	0	0	0
Heterolobosea	0	0	0	1
Metamonada	0	0	0	2
Microsporidians	0	5	0	0
Mollusks	14	0	0	0
Monocots	0	0	23	0
Mosses	0	0	1	0
Nematodes	6	0	0	0
Placozoans	1	0	0	0
Poriferans	1	0	0	0
Red	0	0	3	0
Stramenopiles	0	0	0	8
Tunicates	2	0	0	0
Vertebrates	331	0	0	0

Similarities MSA,Munkres methods

In this section we will show the similarities between mDAG's using different methods.

The experiment data set consists of 884 eukaryotes from the animal, plant, fungus, and protist kingdoms.

Kingdom	Abs. Freq.
Animals	535
Fungi	154
Plants	139
Protists	56

```
list_Sim=dir(path_exp,pattern="^Similarities")
list_Sim

[1] "Similarities_MBB_MSAMethod.csv"           "Similarities_MBB_MunkresMethod.csv"
[3] "Similarities_mDAG_MSAMethod.csv"          "Similarities_mDAG_MunkresMethod.csv"
[5] "Similarities_mDAGOnReaction.csv"
```

Load MDAG similarities

```
Sim_MSA_mDAG=read_csv(paste0(path_exp,
                               "Similarities_mDAG_MSAMethod.csv"))
Sim_MSA_mDAG=as.matrix(Sim_MSA_mDAG[,-1])
rownames(Sim_MSA_mDAG)=colnames(Sim_MSA_mDAG)
Sim_MSA_mDAG=Sim_MSA_mDAG[meta_taxo$mDAG_Id,meta_taxo$mDAG_Id]

Sim_Mun_mDAG=read_csv(paste0(path_exp,
                               "Similarities_mDAG_MunkresMethod.csv"))
Sim_Mun_mDAG=as.matrix(Sim_Mun_mDAG[,-1])
rownames(Sim_Mun_mDAG)=colnames(Sim_Mun_mDAG)
Sim_Mun_mDAG=Sim_Mun_mDAG[meta_taxo$mDAG_Id,meta_taxo$mDAG_Id]
```

Heatmaps

Heatmap Similarity MSA method

```
dff<-meta_taxo %>% select(Kingdom) %>% as.data.frame()
#str(dff)

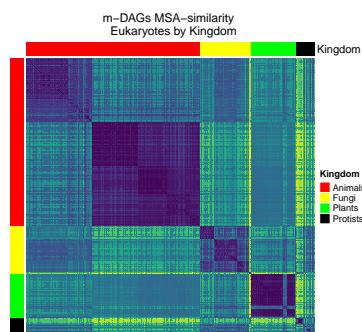
colorsK <- list(Kingdom= c("Animals"="red",
                           "Plants"="green",
                           "Fungi"="yellow",
                           "Protists"="black"))

anot <- HeatmapAnnotation(df=dff, col = colorsK)

#S=Sim_MSA_mDAG

MSA_heat_1 <- Heatmap(matrix = Sim_MSA_mDAG,
                       column_title="m-DAGs MSA-similarity \n Eukaryotes by Kingdom",
                       name = "Kingdom",
                       heatmap_legend_param = list(
at = c(0.4,0.5,0.6,0.7,0.8,0.9,1)),
col=rev(viridis(256)),
cluster_rows = FALSE,
cluster_columns = FALSE,
top_annotation = annot,
show_column_names = FALSE,
show_row_names = FALSE,
left_annotation = rowAnnotation(df = dff,
                                 col=c

draw(MSA_heat_1, merge_legend = TRUE)
```



```

meta_animals= meta_taxo %>% filter(Kingdom=="Animals")
nombres=unique(meta_animals$Phylum)
aux_order=table(meta_animals$Phylum)
dff<-meta_taxo %>% filter(Kingdom=="Animals") %>% select(Phylum) %>% as.data.frame()
#str(dff)
nombres=unique(dff$Phylum)
col=rainbow(length(nombres))
colorsP=list(Phylum=col)
names(colorsP$Phylum)=nombres

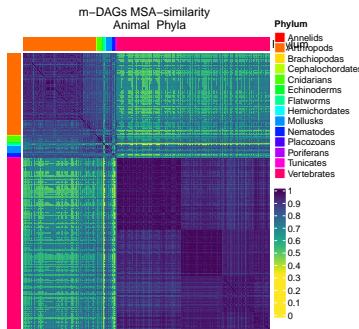
anotation_MSA_H2 <- HeatmapAnnotation(df=dff, col = colorsP,show_legend = TRUE)

#orderP_freq=sort(table(dff$Phylum),decreasing = TRUE)
#orderP_freq=tibble(Phylum=names(orderP_freq),Freq=orderP_freq)

MSA_heat_2 <- Heatmap(matrix = Sim_MSA_mDAG[1:535,1:535],
                       column_title="m-DAGs MSA-similarity \n Animal Phyla",
                       name = " ",
                       heatmap_legend_param = list(at = seq(0,1,by=0.1)),
                       col=rev(viridis(256)),
                       cluster_rows = FALSE,
                       cluster_columns = FALSE,
                       top_annotation = anotation_MSA_H2,
                       show_column_names = FALSE,
                       show_row_names = FALSE,
                       left_annotation = rowAnnotation(df = dff,
                                                       col=c

```

```
draw(MSA_heat_2, merge_legend = TRUE)
```



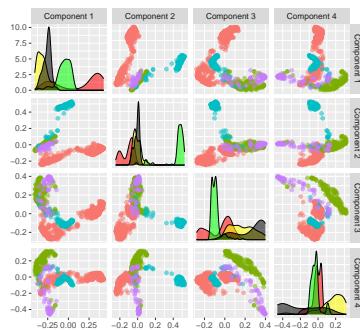
MDS (Multidimensional Scaling)

```
## Metric multidimensional scaling (mMDS)
mds7 <- cmdscale(sqrt(1-Sim_MSA_mDAG^2), k=7, eig=TRUE)
#pairs(mds7$points[,1:4])
mds7$GOF
```

```
[1] 0.4406568 0.5541042
```

```
mds <- mds7$points %>% as_tibble()
colnames(mds) <- paste0("Dim.", 1:dim(mds7$points)[2])
```

```
cooordinates=as_tibble(mds7$points)
colnames(cooordinates)=paste("Component", 1:7)
ggpairs(coordinates, columns=1:4, aes(color=meta_taxo$Kingdom, alpha=0.5, title="MDS 4 dimensions"))
```



```
mds <- mds %>%
  mutate(groups = as.factor(meta_taxo$Kingdom))
```

```
#,text= ~paste("Age:", groups, '<br>Name:')
length(unique(meta_taxo$Phylum))
```

```
[1] 33
```

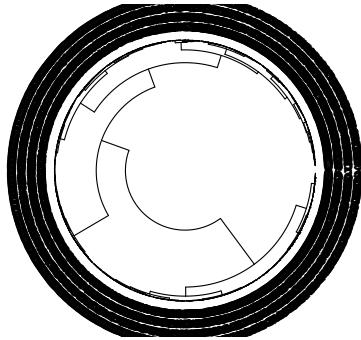
```
col_mds=rainbow(33)

#col_mds=c("purple","green","yellow","coral")
#mcol_mds=bremer.pal(7,"Greens")

# fig <-
# plot_ly(
#   mds, x = ~Dim.1, y = ~Dim.2,
#   color = ~groups,
#   colors= colors,
#   type="scatter",
#   mode="markers") %>%
#   layout(
#     xaxis = list(autorange=2,
#                 range=c(-0.8,0.8)), yaxis = list(autorange=2,
#                 range=c(-0.8,0.8)))
# #
# jpeg(filename="figures/fig1.jpeg")
# fig
# dev.off
```


Hierarchical cluster

```
library(dendextend)
D=as.dist(sqrt(1-Sim_MSA_mDAG^2))
hc=hclust(as.dist(D),method ="ward.D")
library(circlize)
circlize_dendrogram(as.dendrogram(hc))
```



```
clust4=cutree(hc,4)
table(clust4,meta_taxo$Kingdom)
```

clust4	Animals	Fungi	Plants	Protists
1	195	0	0	0
2	9	154	14	56
3	331	0	0	0
4	0	0	125	0

Similarities Munkres method

Heatmap Similarity Mun method

```
dff<-meta_taxo %>% select(Kingdom) %>% as.data.frame()
#str(dff)

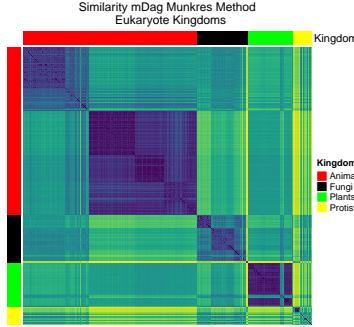
colorsK <- list(Kingdom= c("Animals"="red",
                           "Plants"="green",
                           "Fungi"="black",
                           "Protists"="yellow"))

anotation_heat <- HeatmapAnnotation(df=dff, col = colorsK)
dim(dff)

[1] 884    1

Mun_heat_1<- Heatmap(matrix = Sim_Mun_mDAG,
                      column_title="Similarity mDag Munkres Method\n Eukaryote Kingdoms",
                      name = "Kingdom",
                      heatmap_legend_param = list(
at = c(0.4,0.5,0.6,0.7,0.8,0.9,1)),
                      col=rev(viridis(256)),
                      cluster_rows = FALSE,
                      cluster_columns = FALSE,
                      top_annotation = anotation_heat,
                      show_column_names = FALSE,
                      show_row_names = FALSE,
                      left_annotation = rowAnnotation(df = dff, col = colorsK,show_annotation_name=FALSE))

draw(Mun_heat_1, merge_legend = TRUE)
```



```

meta_animals= meta_taxo %>% filter(Kingdom=="Animals")
nombres=unique(meta_animals$Phylum)
aux_order=table(meta_animals$Phylum)
dff<-meta_taxo %>% filter(Kingdom=="Animals") %>% select(Phylum) %>% as.data.frame()
names_phylum=unique(dff$Phylum)
names_phylum

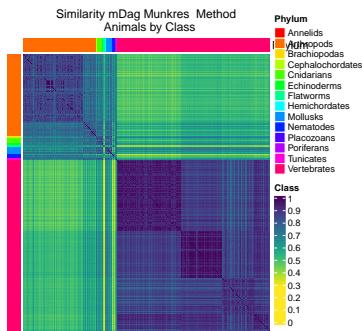
[1] "Annelids"          "Arthropods"        "Brachiopodas"      "Cephalochordates"
[5] "Cnidarians"        "Echinoderms"       "Flatworms"         "Hemichordates"
[9] "Mollusks"          "Nematodes"         "Placozoans"        "Poriferans"
[13] "Tunicates"         "Vertebrates"

col=rainbow(length(nombres))
colorsP=list(Phylum=col)
names(colorsP$Phylum)=names_phylum
anotacion <- HeatmapAnnotation(df=dff, col = colorsP)

Mun_heat_2 <- Heatmap(matrix = Sim_Mun_mDAG[1:535,1:535],
                      column_title="Similarity mDag Munkres Method\n Animals by Class",
                      name = "Class",
                      heatmap_legend_param = list(
at = seq(0,1,by=0.1)),
                      col=rev(viridis(256)),
                      cluster_rows = FALSE,
                      cluster_columns = FALSE,
                      top_annotation = anotacion,
                      show_column_names = FALSE,
                      show_row_names = FALSE,
                      left_annotation = rowAnnotation(df = dff,
                                         col = colorsP,

```

```
draw(Mun_heat_2, merge_legend = TRUE)
```



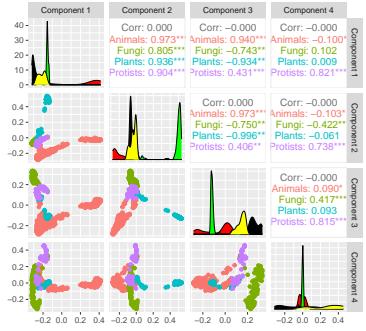
MDS (Multidimensional Scaling)

```
## Metric multidimensional scaling
mds7 <- cmdscale(sqrt(1-Sim_Mun_mDAG^2), k=7, eig=TRUE)
mds7$GOF
```

```
[1] 0.5599891 0.5796237
```

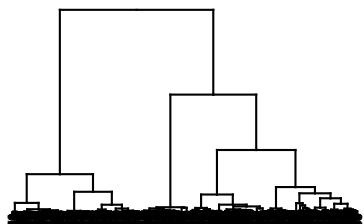
```
mds <- mds7$points %>% as_tibble()
colnames(mds) <- paste0("Dim.", 1:dim(mds7$points)[2])

cooordinates=as_tibble(mds7$points)
colnames(cooordinates)=paste("Component", 1:7)
ggpairs(cooordinates, columns=1:4,
        aes(color=meta_taxo$Kingdom,
            title="MDS 4 dimensions projection", legend=1),
        lower=list(continuous="points")) +
        scale_fill_manual(values = colorsK$Kingdom) +
        theme(legend.position = "left")
```



Hierarchical cluster

```
D=as.dist(sqrt(1-Sim_Mun_mDAG^2))
hc=hclust(as.dist(D),method ="ward.D")
ggplot(as.ggdend(as.dendrogram(hc)))
```



```
clust4=cutree(hc,4)
table(clust4,meta_taxo$Kingdom)
```

	Animals	Fungi	Plants	Protists
1	197	0	0	0
2	7	154	14	56
3	331	0	0	0
4	0	0	125	0

Similarity comparisons Eukaryotes

Comparison of two similarities

Load the similarities for pairs and comparison

```
n=length(meta_taxo$mDAG_Id)
n

[1] 884

dim(Sim_MSA_mDAG)

[1] 884 884

aux1=base::rep(x=1:n,each=c(n:1))

aux=as_tibble(Sim_MSA_mDAG)
aux$mDag=names(aux)
aux=aux %>% pivot_longer(cols=`0576`:`0286`,
                           names_to="mDag_2",values_to="Sim_MSA")
aux_2= aux %>% mutate(i=pmax(as.integer(mDag),as.integer(mDag_2)),
                      j=pmin(as.integer(mDag),as.integer(mDag_2)))%>%
  unite("ij",i:j) %>% filter(duplicated(ij))

aux=as_tibble(Sim_Mun_mDAG)
aux$mDag=names(aux)
aux=aux %>% pivot_longer(cols=`0576`:`0286`,names_to="mDag_2",values_to="Sim_Mun")
aux_2 = aux_2 %>% left_join(aux)

Sim_comp=aux_2
rm(aux,aux_2)
```

Spearman and Pearson correlations

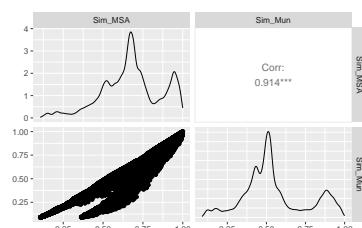
```
cor(Sim_comp[,c(3,5)],method="spearman")
```

```
Sim_MSA   Sim_Mun  
Sim_MSA 1.0000000 0.8902412  
Sim_Mun 0.8902412 1.0000000
```

```
cor(Sim_comp[,c(3,5)],method="pearson")
```

```
Sim_MSA   Sim_Mun  
Sim_MSA 1.000000 0.913906  
Sim_Mun 0.913906 1.000000
```

```
ggpairs(Sim_comp[,c(3,5)])
```



```
boxplot(Sim_comp[,c(3,5)])
```

