Mouse Gut Microbiota - Mental Health Database

Cao Zeyuan, Jia Liu, Mikhail Kouzminov

Description:

The goal is to put all the data utilized for the paper into an online database. The user will be able to choose interesting data and automatically generate visualizations, including some of the ones used for the paper, including stack bar visualizations and heatmaps. We can further expand the function of the database by encouraging users to input their own data and make it a free public Gut Microbiota Database. This would require moderation of some sort to prevent users inputting bad data, but we will allow them to use our database as a tool to visualize their own similar data. Further, if we have enough time, we can add more data analysis functions into the database.

Data:

All data are related to gut microbiota and mental health, such as species distribution in gut, or changes in mental behaviors (as shown by results in a maze) and brain IL-6 levels under different experimental conditions (ie given probiotics or not post-surgery). Most of the data sets are from supplements of paper, others are relatively unpublished data from mgh lab (the format of unpublished data is csv) . There are also images of Western blot experiments for data gathered for this experiment.

Main Lab contact (through former internship of Cao):

Zhongcong Xie

zxie@hms.harvard.edu

617-724-9308

Anesthesia, Critical Care and Pain Medicine

55 Fruit Street

Boston, MA 02114-2696

User Accessibility:

The database will be made available to the public - while the data is new and gathered specifically for the paper, none of it is sensitive.

Some examples of questions that will be answered by MMDB (Mouse Gut Microbiota - Mental Health Database):

a. How do particular species of gut microbiota affect AD in mice?

b. Which families of gut microbiota are relatively more expressed?

c. Which classes of gut microbiota are relatively more expressed?

d. What experimental dietary conditions are linked to good mental health in mice (Would need more data, this experiment only accounted for probiotics)?

e. How is Alzheimer's Disorder in mice linked to gut microbiota?

A specific list of tasks that you expect to accomplish.

A. Create tables for dataset

B. According to the assumed frequency of use and retrieval objects, the index of the table is optimized to improve the retrieval speed.

C. Link webpage to database

D. Create visualizations

E. Design webpage

ER model (diagram) including key and participation constraints.

In the following tables, underlines stand for primary key, *italics* stand for foreign key. For each table, set the primary key as a clustered index. For value table(c-value, f-value, g-value, p-value, s-value, o-value tables), all set spid as unclustered index.

Class: cid, name, *pid (refer to Phylum table)*

Family: fid, name, *oid (refer to Order table)*

Genus: <u>gid</u>, name, *fid (refer to Class table)*

Phylum: <u>pid</u>, name

Species: <u>sid</u>, name, *gid (refer to Genus table)*

Order: <u>oid</u>, name, *cid (refer to Class table)*

Samples: <u>spid</u>, name

c-value: cvid, value, *cid, spid (refer to Class table)*
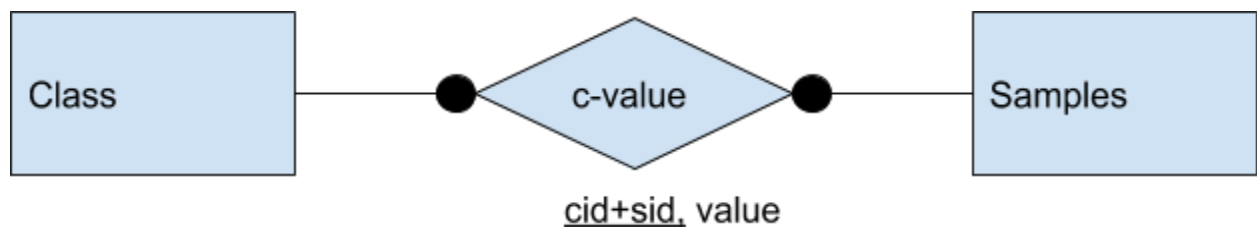
f-value: fvid, value, *fid, spid (refer to Class table)*
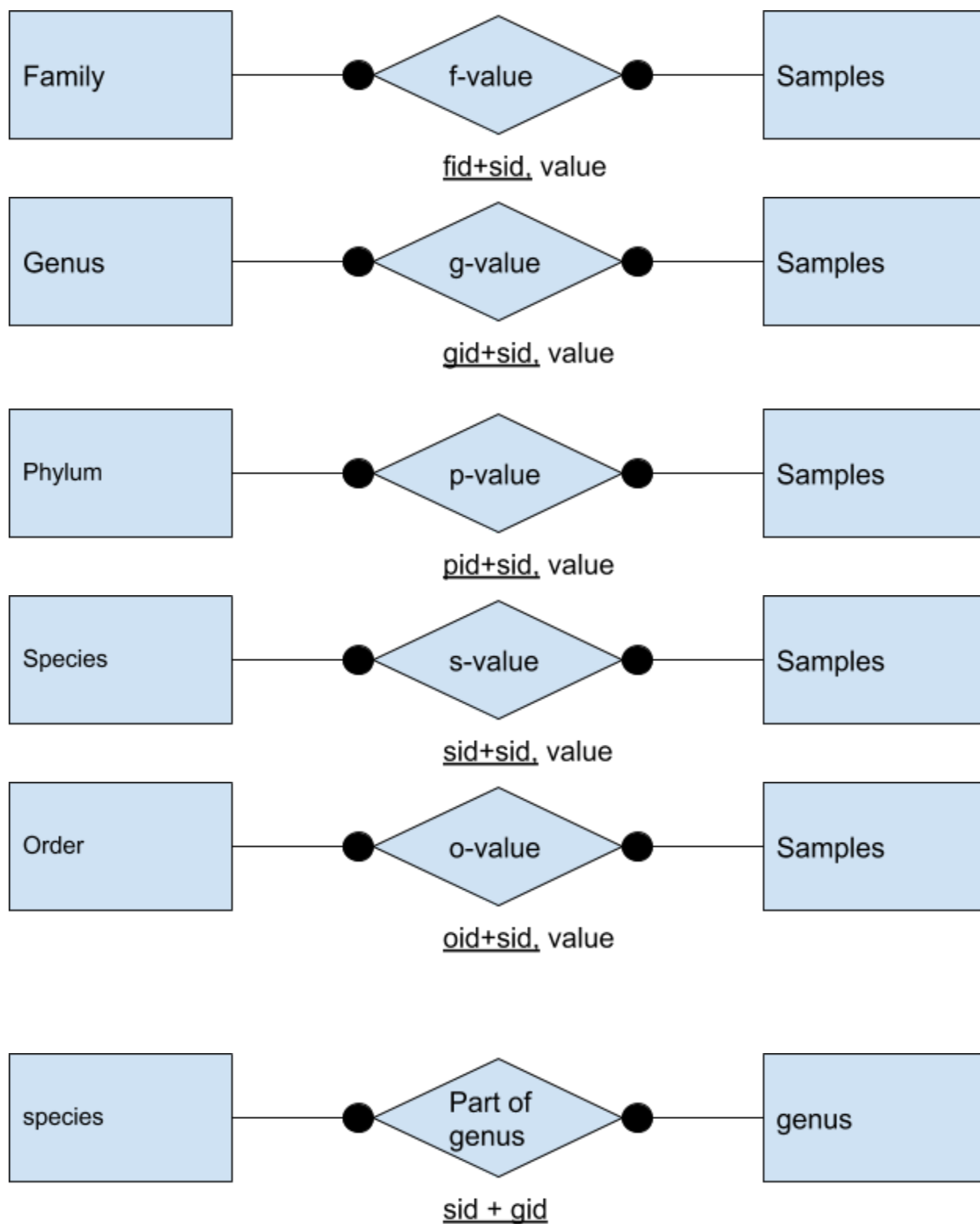
g-value: gvid, value, *gid, spid (refer to Genus table)*
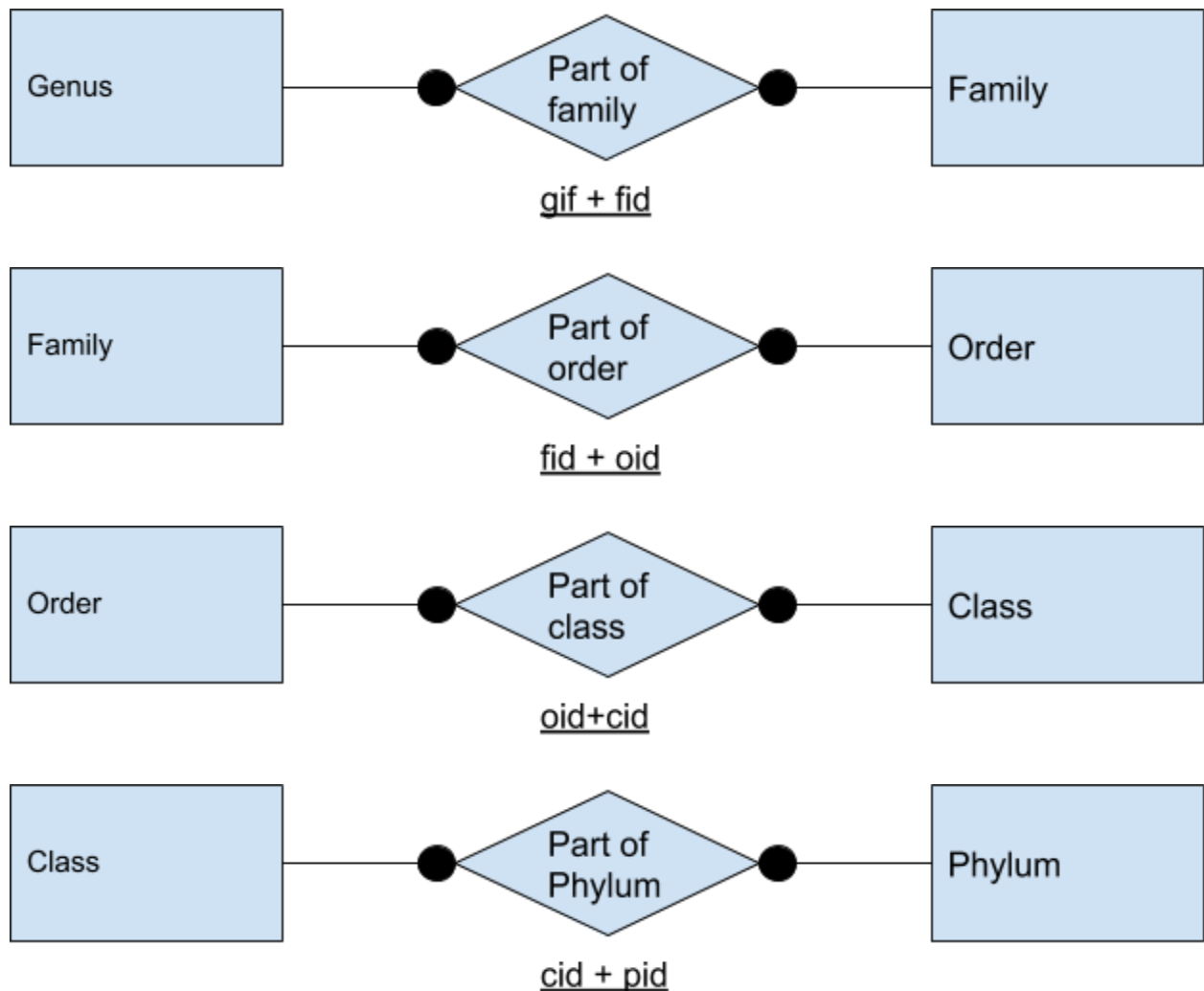
p-value: pvid, value, *pid, spid (refer to Phylum table)*

s-value: svid, value, *sid, spid (refer to Species table)*

o-value: ovid, value, *oid, spid (refer to Order table)*

In the Species table, foreign key gid refers to table Genus (gid) indicating which genus it belongs to. In the Genus table, foreign key fid refers to table Family (fid) indicating which family it belongs to. In the Family table, foreign key oid refers to table Order (oid) indicating which order it belongs to. In the Order table, foreign key cid refers to table Class (cid) indicating which class it belongs to. In the Class table, foreign key pid refers to table Phylum (pid) indicating which phylum it belongs to.

cid+sid, value

| Family | ●— f-value —● | Samples |

fid+sid, value

| Genus | ●— g-value —● | Samples |

gid+sid, value

| Phylum | ●— p-value —● | Samples |

pid+sid, value

| Species | ●— s-value —● | Samples |

sid+sid, value

| Order | ●— o-value —● | Samples |

oid+sid, value

| species | ●— Part of genus —● | genus |

sid + gid

Genus — Part of family — Family
gif + fid

Family — Part of order — Order
fid + oid

Order — Part of class — Class
oid+cid

Class — Part of Phylum — Phylum
cid + pid

Description of tables:

Tables show relative abundance of classes, families, genuses, and species of microbiota in various samples. Samples are columns, classes, families, genuses, and species are represented by rows of different tables.

Three sample SQL queries for common functions of the database.

- Select Samples.name, value from Species join s-value using(sid) join Sample using(spid) where power(10,value)>1;

search for all samples in the Species table and their abundance which have relative abundance greater than 10% in at least 1 of the samples.

- Select Phylum.name from Phylum join Class using(pid) where Class.name regexp 'Bacilli';

  #search for the name of the phylum that species 'Bacilli' belongs to.

- Select power(10,c-value.value) from Samples join Class using(spid) where Samples.name regexp 'LZX10C';

  #search for class percent relative abundance of sample 'LZX10C'

Description of data processing, scheduled or performed by the user interface program, but which is external to the database (e.g. BLAST searches, statistical analysis, etc.) if any.

None planned

List of other databases to which this database provides links.

Pubmed

Description of graphical output of the database. Include examples.

Stacked Bar plot of relative abundance of bacterial subtypes (Done on multiple levels, including species, genus, etc…) in mice with Alzheimer's disease vs in mice without. Heatmaps of average probiotic vs saline bacterial presence. Heatmaps of average AD vs Wild Type bacterial presence.

Description of the data download function of the database (i.e., what data comes out and in what format).

Download Excel files showing relative abundance for various genes - being able to select for Wild Type vs Alzheimer's Disease samples as well as select for classification level to download data as.