mouse_microbiome

John Guittar

May 18, 2018

```r
# set wd and load packages
wd <- 'C:\\Users\\John\\Documents\\msu\\mouse_microbiome\\'
setwd(wd)
source('custom_functions.R')
```

```
##                                          Custom_Functions
## 1               loadpax: install+load multiple packages
## 2 grid_arrange_shared_legend: Multiple plots, one legend
## 3                             multiplot: multiple plots
```

```r
loadpax(pkg = c('tidyverse','knitr','data.table'))
```

```
## Loading required package: tidyverse
```

```
## -- Attaching packages ------------------------------------------------------------
```

```
## v ggplot2 2.2.1      v purrr   0.2.4
## v tibble  1.4.2      v dplyr   0.7.4
## v tidyr   0.8.0      v stringr 1.3.1
## v readr   1.1.1      v forcats 0.3.0
```

```
## -- Conflicts --------------------------------------------------------------------- tid
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
## Loading required package: knitr
```

```
## Loading required package: data.table
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
## The following object is masked from 'package:purrr':
##
##     transpose
```

```
##  tidyverse      knitr data.table
##      TRUE       TRUE       TRUE
```

```r
# load data
x_raw <- fread('Data\\2018_06_12_run2_otus.csv')
tax <- fread('Data\\otus_table_joined_taxa.txt')

#Organize tax data into traditional classifications
tax <- tax %>%
  rename(otu = `#OTU ID`, tax = taxonomy) %>%
  mutate(tax = gsub("\\s|.__|\\[|\\]|Other", "", tax)) %>%
  separate(tax, sep = ';', c("Domain", "Phylum", "Class", "Order", "Family", "Genus", "Species"), fill =
tax <- as.data.frame(apply(tax, 2, function(x) ifelse(x == '', 'unclassified', x)), stringsAsFactors = 1
```

```r
tax$otu <- as.numeric(tax$otu)

# Note that, oddly, x[130,4719] is 'NA'. I'm assuming it is zero...?
is.na(x_raw[130,4719])
```

```
##      174320
## [1,]    TRUE
```

```r
x_raw[130,4719] <- 0

#remove samples with fewer than 5000 reads
# remove mouse 310 because it dies after hour 120
# create narrow table
# note that there are two samples for Cage 322 at Hour 168. They are merged below
x <- x_raw %>%
  rename(Cage = `Cage#`) %>%
  filter(Reads > 5000) %>%
  filter(Cage != 310) %>%
  select(-Day, -Run, -Reads, -SampleID) %>%
  gather(otu, abun, -Cage, -Abx, -Hour, -Cohoused, -Dose, -Pair) %>%
  group_by(Cage, Abx, Hour, Cohoused, Dose, Pair, otu) %>%
  summarise(abun = sum(abun)) %>%
  group_by(Cage, Hour) %>%
  mutate(abun = abun / sum(abun),
         otu = as.numeric(otu))

#Set plotting theme
th <- theme_bw() +
  theme(
    panel.grid.minor = element_blank(),
    panel.grid.major = element_blank(),
    panel.background = element_blank())
```

Percent of total phyla present within single mice, grouped by taxonomic level. Need to add OTU as a final category.

```r
j <- tax %>%
  gather(tax_level, name, -otu) %>%
  full_join(x, by = 'otu') %>%
  filter(tax_level != 'Domain') %>%
  filter(abun > 0)

j <- j %>%
  mutate(tax_level = factor(tax_level, levels = c('Phylum','Class','Order','Family','Genus','Species'))
  group_by(tax_level) %>%
  mutate(total_rich = n_distinct(name)) %>%
  group_by(Cage, tax_level) %>%
  mutate(rich = n_distinct(name), val = rich / total_rich)

j %>%
  ggplot(aes(x = tax_level, y = val, group = Cage)) +
  geom_line() +
```

```
expand_limits(y = c(0,1)) +
th
```
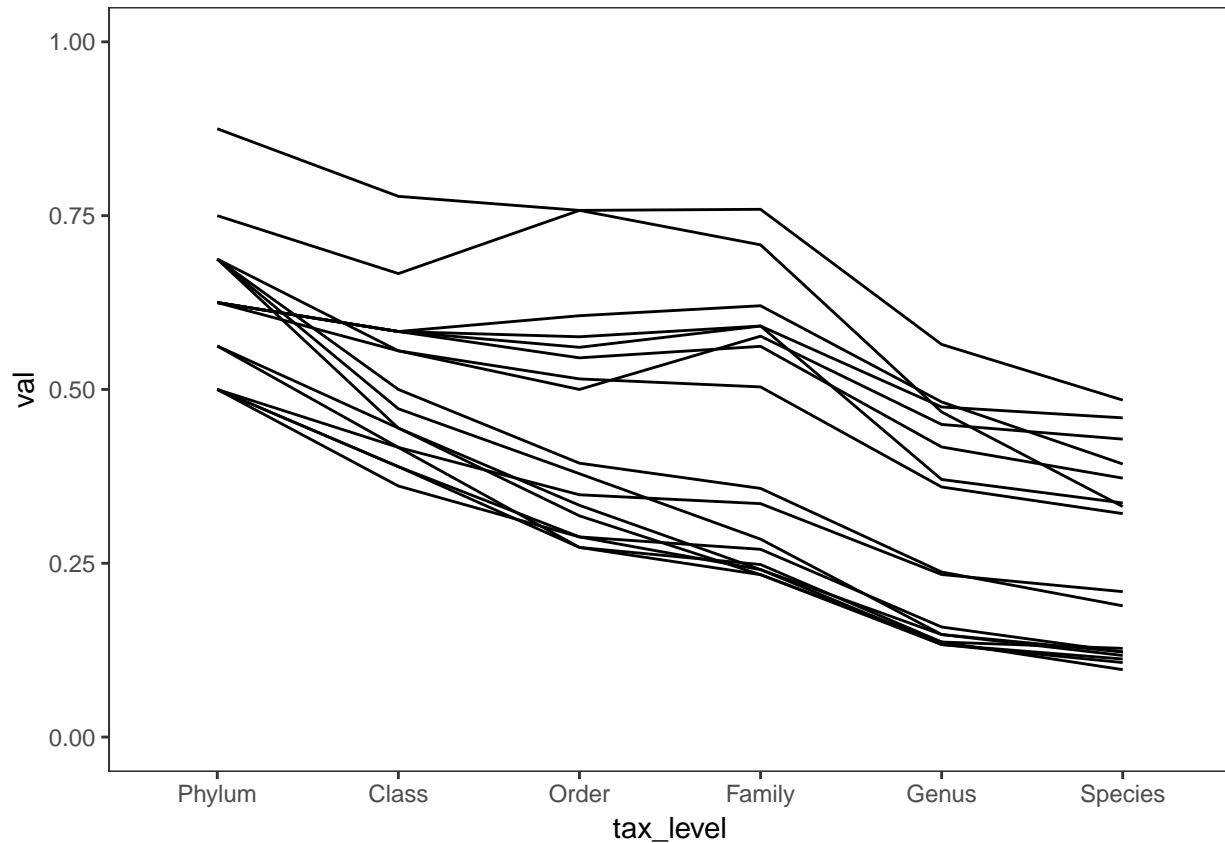

First, I categorize each OTU in each antibiotic treated mouse into *Absent* (never was there), *Extinct* (was there at the end of antibiotic treatment but then disappeared), *Persistent* (was present throughout the experiment), and *Immigrant* (Was absent at the end of AB treatment, but colonized during recover).

The first plot shows the Hour at which immigrants arrive, plotted by the abundance of that immigrant in the control mouse.

```
#I am assuming that OTUs that are absent from a given mouse in the last two samples collected during an

j <- x %>%
  group_by(Cage) %>%
  filter(Hour >= 63 | Abx == 'N') %>%
  group_by(Cage, otu) %>%
  mutate(
    abun0 = sum(abun[Hour <= 97]),
    status0 = if (sum(abun[Hour <= 97]) > 0) 'present' else 'absent',
    status1 = if (sum(abun[Hour > 97]) > 0) 'present' else 'absent',
    status = ifelse(status0 == 'absent' & status1 == 'absent', 'absent',
               ifelse(status0 == 'absent' & status1 == 'present', 'immigrant',
                 ifelse(status0 == 'present' & status1 == 'absent', 'extinct',
                   'persistent'))),
    arrival_time = if (status[1] == 'immigrant') min(Hour[Hour > 97 & abun > 0]) else NA) %>%
  ungroup()
```

```r
tmp <- j %>%
  filter(!is.na(Pair)) %>%
  group_by(Pair, otu) %>%
  mutate(control_abun = ifelse(Abx == 'Y', unique(abun0[Abx == 'N']), 0))

tmp %>%
  filter(status == 'immigrant') %>%
  #filter(control_abun > 0) %>%
  mutate(
    time_to_arrival = arrival_time - 120,
    control_abun = control_abun * 100 + 0.01) %>%
  ggplot(aes(x = control_abun, y = time_to_arrival)) +
    geom_jitter(alpha = 0.05, color = 'red') +
    scale_x_log10() +
    stat_smooth(color = 'black', method = 'lm') +
    stat_smooth(color = 'black', lty = 3) +
    labs(x = "OTU abundance in cohoused control mouse + 0.1 %", y = 'Hours before OTU arrival in Abx tre
    th
```
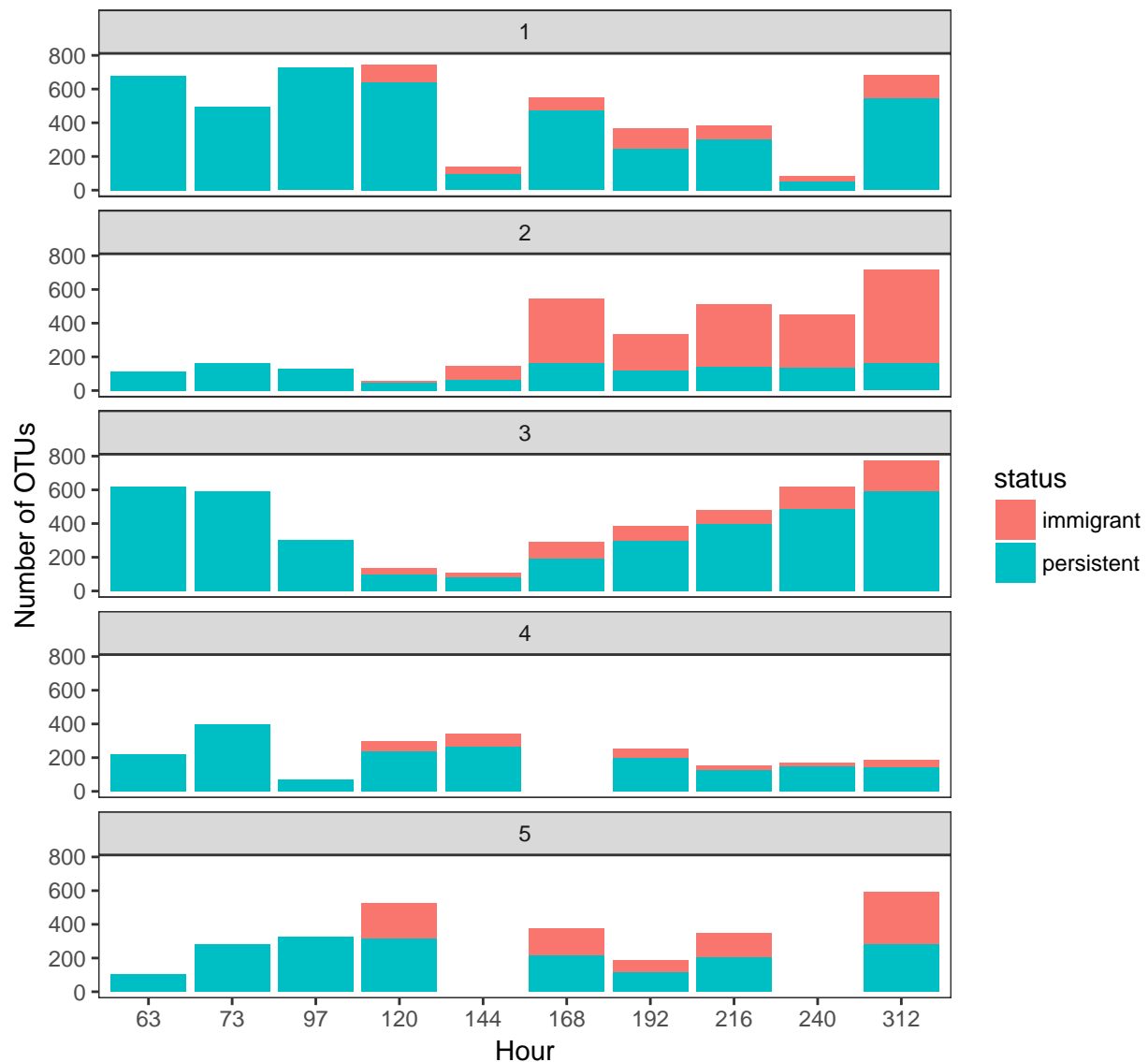
`geom_smooth()` using method = 'gam'
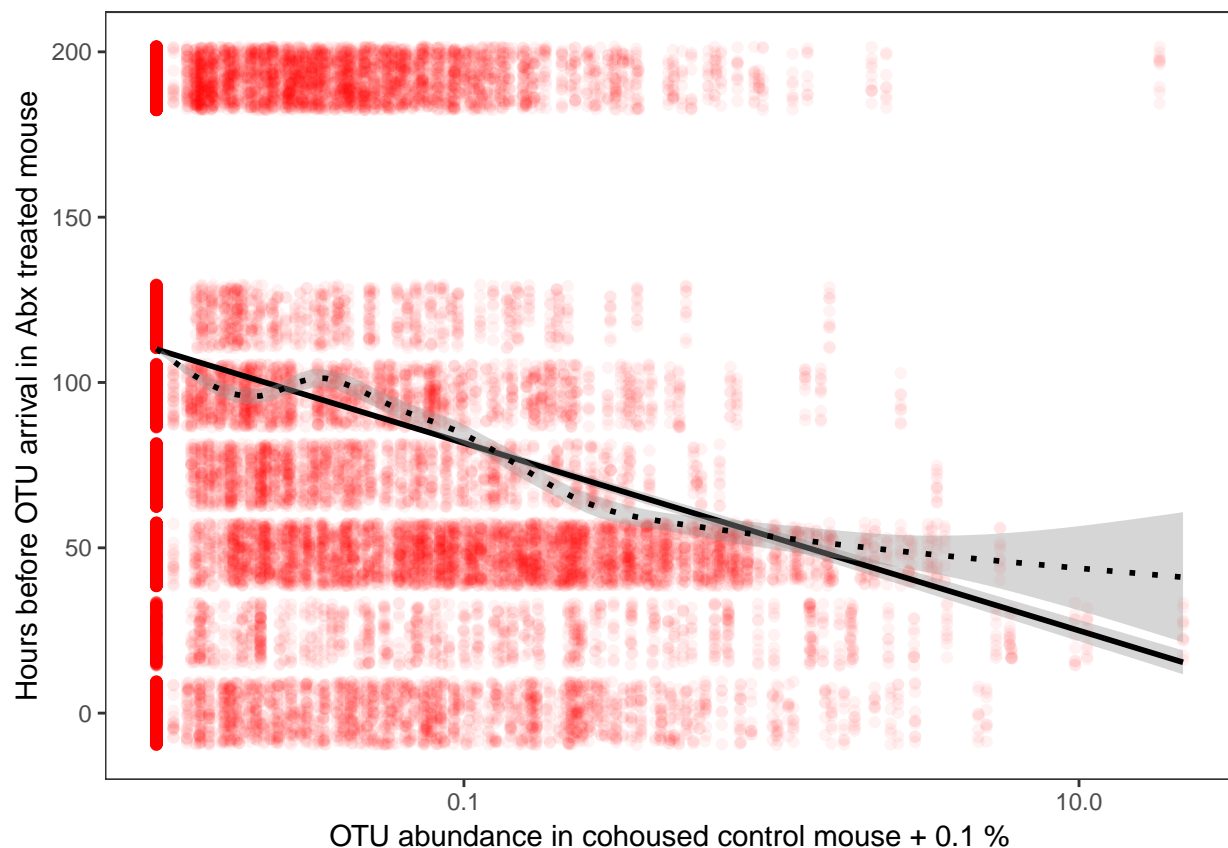

```r
#numbers of OTUs, colored by local status
tmp %>%
  filter(Abx == 'Y' & status %in% c('immigrant','persistent') & abun > 0) %>%
  mutate(status = as.character(status)) %>%
  ungroup() %>%
```

```
  arrange(status) %>%
  ggplot(aes(x = factor(Hour), fill = status)) +
    geom_bar() +
    facet_wrap(~Pair, ncol = 1) +
    labs(x = "Hour", y = "Number of OTUs") +
    th
```



```
#Relative contribution of OTUs, grouped by local status
tmp %>%
  filter(Abx == 'Y' & status %in% c('immigrant','persistent')) %>%
  mutate(status = as.character(status)) %>%
  ungroup() %>%
  arrange(status) %>%
  ggplot(aes(x = factor(Hour), y = abun, fill = status)) +
    geom_bar(stat = 'identity') +
    facet_wrap(~Pair, ncol = 1) +
    labs(x = "Hour", y = "Relative abundances") +
```