

Reproducible Analysis: The Two-Wing admixed structure of environmental microbial communities

Yushi Tang

2022-08-25

```
library(ggplot2)      # for generating plots
library(latex2exp)    # for plot text latex
library(cowplot)      # for merging plots
library(gridExtra)    # for gridding plots
library(grid)         # for gridding plots
library(stringr)      # for uppercase first letter
```

Preface

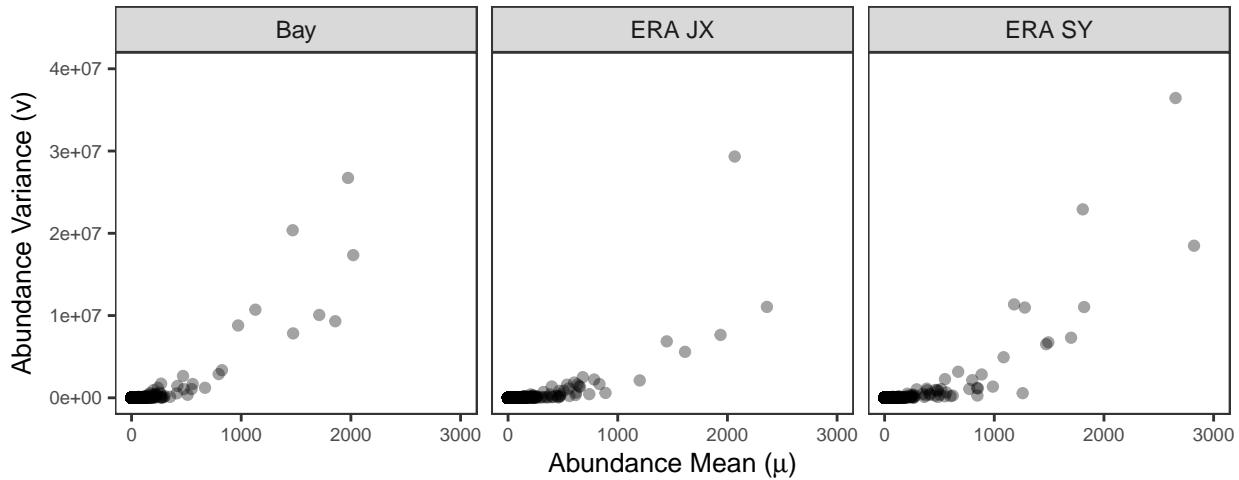
This document provides reproducible research records for our manuscript about the Two-Wing admixed structure of environmental microbial communities. We provide step-by-step instructions for main results in both the original article and the supplemental materials.

Main Figures

Figure 1. Observed abundance distribution

The over-dispersion pattern of microbial communities

```
load('../output/HZ/hz_mean_var.RData')
dat_plt$Region <- factor(dat_plt$Region, levels=c('Bay','ERA JX','ERA SY'))
ggplot(dat_plt, aes(x=miseqmean, y=miseqvar)) +
  geom_point(alpha=0.36, size=1.5) +
  xlab(TeX('Abundance Mean ($\mu$)')) +
  ylab(TeX('Abundance Variance ($v$)')) +
  xlim(0,3000) +
  ylim(0,4e7) +
  theme_bw() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.title.x = element_text(size = 10),
        axis.title.y = element_text(size = 10),
        axis.text.x = element_text(size = 7),
        axis.text.y = element_text(size = 7)) +
  theme(aspect.ratio=1) +
  facet_grid(~Region)
```



```

load('../output/AO_ASV/ao_mean_var.RData')
dat_plt$Industry <- factor(dat_plt$Industry, levels=c('Dye','Pharmaceutical','Pesticide'))
dat_plt$Process <- factor(dat_plt$Process, levels=c('Influent','Anoxic','Oxic','Effluent'))
ggplot(dat_plt, aes(x=miseqmean, y=miseqvar)) +
  geom_point(alpha=0.36, size=1.5) +
  xlab(TeX('Abundance Mean ($\mu$)')) +
  ylab(TeX('Abundance Variance ($v$)')) +
  xlim(0,1000) +
  ylim(0,4e6) +
  theme_bw() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.title.x = element_text(size = 10),
        axis.title.y = element_text(size = 10),
        axis.text.x = element_text(size = 7),
        axis.text.y = element_text(size = 7)) +
  theme(aspect.ratio=1) +
  facet_grid(Industry~Process)

```

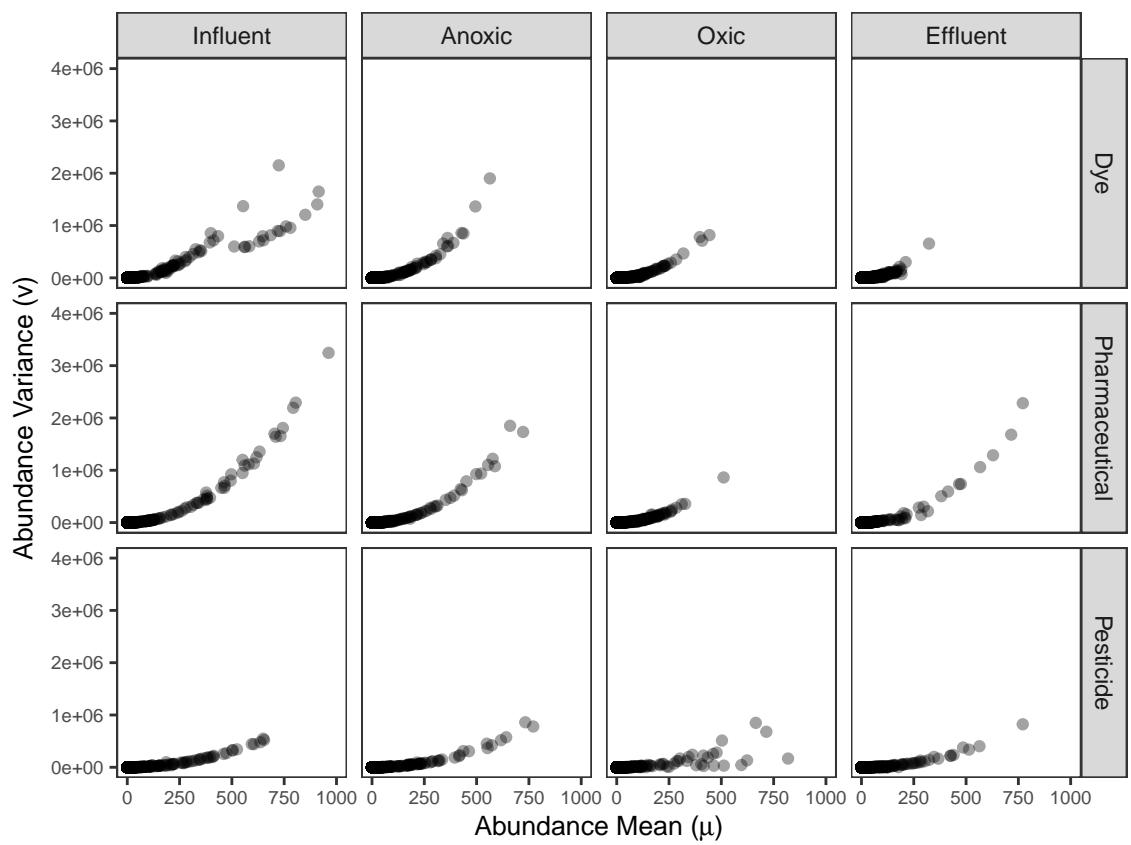
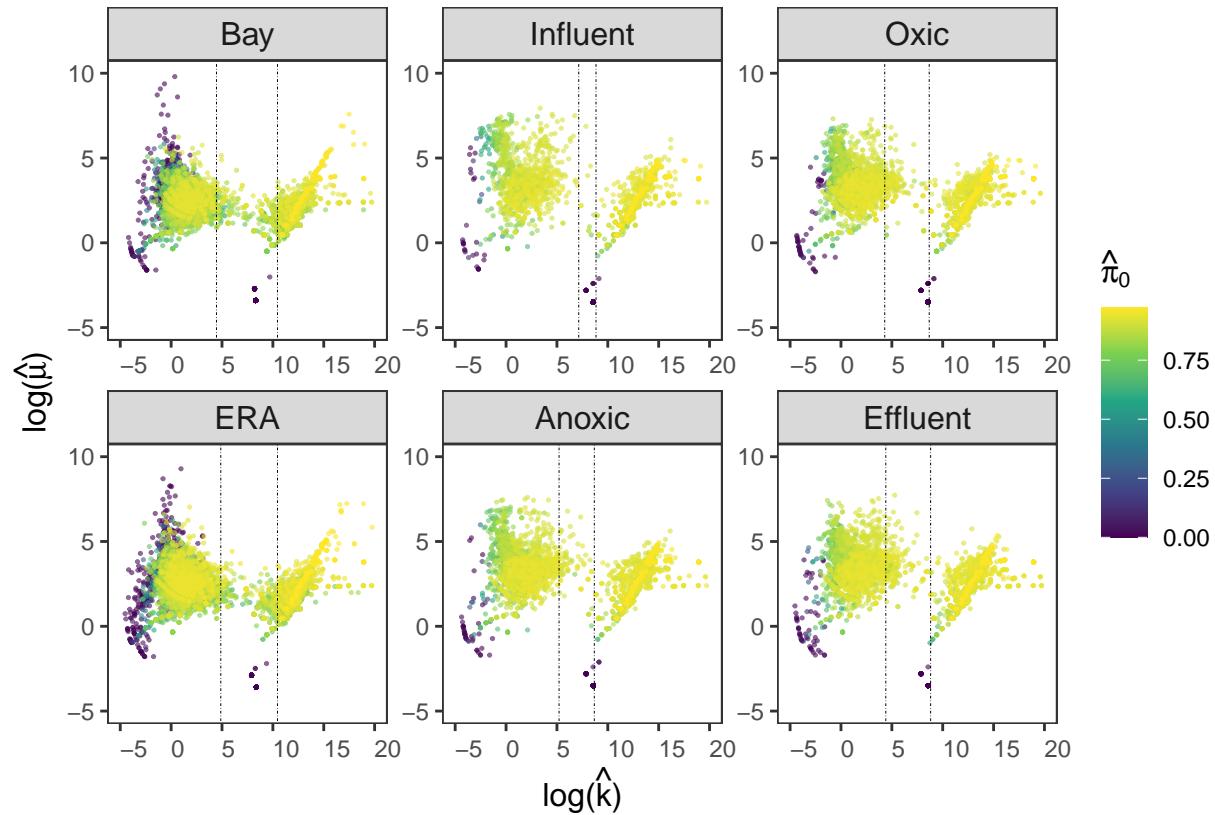


Figure 2. The Two-Wing admixed structure

Before boundary refining (Figure 2A)



After boundary refining (Figure 2B)

```
# load parameters mle
vec_process <- c('influent','anoxic','oxic','effluent')
param_plt <- data.frame()
for(item in vec_process){
  load(paste0('../output/AO_ASV/param_trio_asv_',item,'.RData'))
  load(paste0('../output/AO_ASV/mic_od_',item,'.RData'))
  param_trio$ID <- rownames(param_trio)
  # extract gamma-poisson distributed microbes
  param_gpm <- param_trio[param_trio$k!=Inf,]
  # annotate procedure
  param_gpm$Procedure <- str_to_title(item)
  # annotate taxa category
  load(paste0('../output/AO_ASV/mic_id_mu_k_',item,'.RData'))
  param_gpm$Wing <- ifelse(param_gpm$ID %in% id_mu, 'mu', 'Other')
  param_gpm$Wing <- ifelse(param_gpm$ID %in% id_k, 'k', param_gpm$Wing)
  param_plt <- rbind(param_plt, param_gpm)
}

vec_process <- c('bay','era')
```

```

for(item in vec_process){
  load(paste0('../output/HZ/param_trio_',item,'.RData'))
  load(paste0('../output/HZ/mic_od_',item,'.RData'))
  param_trio$ID <- rownames(param_trio)
  # extract gamma-poisson distributed microbes
  param_gpm <- param_trio[param_trio$k!=Inf,]
  # annotate procedure
  param_gpm$Procedure <- str_to_title(item)
  # annotate taxa category
  load(paste0('../output/HZ/mic_id_mu_k_',item,'.RData'))
  param_gpm$Wing <- ifelse(param_gpm$ID %in% id_mu, 'mu', 'Other')
  param_gpm$Wing <- ifelse(param_gpm$ID %in% id_k, 'k', param_gpm$Wing)

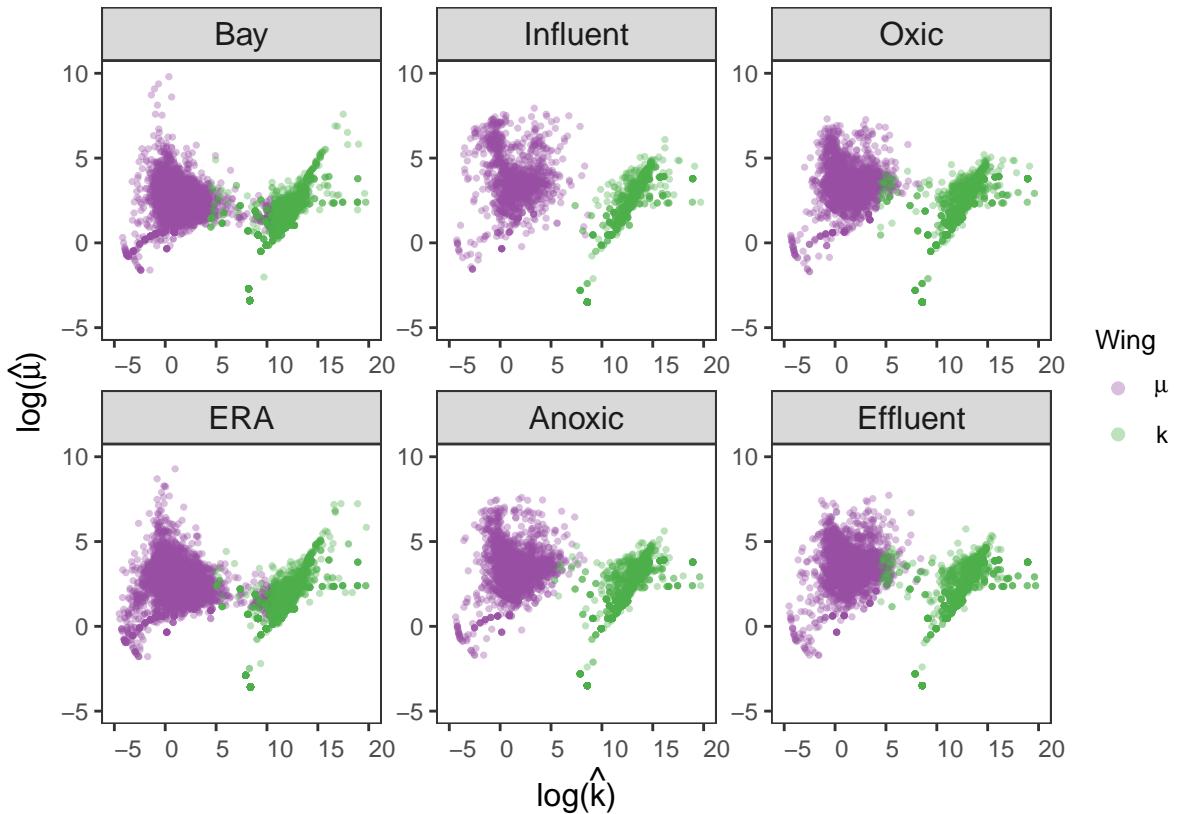
  param_plt <- rbind(param_plt, param_gpm)
}

rownames(param_plt) <- c(1:nrow(param_plt))
param_plt$Procedure <- ifelse(param_plt$Procedure=='Era','ERA',param_plt$Procedure)

param_plt$Procedure <- factor(param_plt$Procedure,
                               levels=c('Bay','Influent','Oxic',
                                       'ERA','Anoxic','Effluent'))
param_plt$Wing <- factor(param_plt$Wing,
                           levels=c('mu','k', 'Other'))

ggplot() +
  geom_point(data=param_plt,
             aes(x=log(k), y=log(mu), color=Wing), alpha=0.36, size=0.6) +
  ylab(TeX('\\log{(\hat{\mu})}')) +
  xlab(TeX('\\log{k}')) +
  ggtitle(TeX('')) +
  xlim(-5,20) +
  ylim(-5,10) +
  theme_bw() +
  theme(aspect.ratio=1) +
  facet_wrap(Procedure~, scale='free') +
  scale_colour_manual(values = c('#984ea3','#4daf4a'),
                      labels = unname(TeX(c('$\\mu$', '$k$')))) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.title.x = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        axis.text.x = element_text(size = 9, vjust = 0.5, hjust = 0),
        axis.text.y = element_text(size=9),
        strip.text = element_text(size = 12),
        legend.title = element_text(color = "black", size = 10),
        legend.text = element_text(color = "black", size = 9)) +
  guides(color = guide_legend(override.aes = list(size = 1.8)))

```



```
sessionInfo()
```

```
## R version 4.2.0 (2022-04-22)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Catalina 10.15.3
##
## Matrix products: default
## BLAS:    /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRblas.0.dylib
## LAPACK:  /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] grid      stats     graphics   grDevices  utils      datasets   methods
## [8] base
##
## other attached packages:
## [1] stringr_1.4.0   gridExtra_2.3    cowplot_1.1.1    latex2exp_0.9.4
## [5] ggplot2_3.3.6
##
## loaded via a namespace (and not attached):
## [1] highr_0.9        pillar_1.8.0      compiler_4.2.0    tools_4.2.0
## [5] digest_0.6.29    viridisLite_0.4.0  evaluate_0.15    lifecycle_1.0.1
## [9] tibble_3.1.8      gtable_0.3.0      pkgconfig_2.0.3   rlang_1.0.4
## [13] cli_3.3.0        rstudioapi_0.13   yaml_2.3.5       xfun_0.31
## [17] fastmap_1.1.0    withr_2.5.0       dplyr_1.0.9      knitr_1.39
## [21] generics_0.1.3    vctrs_0.4.1       tidyselect_1.1.2  glue_1.6.2
```

```
## [25] R6_2.5.1           fansi_1.0.3        rmarkdown_2.14      farver_2.1.1
## [29] purrrr_0.3.4        magrittr_2.0.3       scales_1.2.0       htmltools_0.5.3
## [33] colorspace_2.0-3    labeling_0.4.2      utf8_1.2.2        stringi_1.7.8
## [37] munsell_0.5.0
```