

Bias_simulation

Alexander Piper

03/10/2019

```
library(tidyverse)
```

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

```
## v ggplot2 3.2.1    v purrr  0.3.3
## v tibble  2.1.3    v dplyr  0.8.3
## v tidyr   1.0.0    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.4.0
```

```
## -- Conflicts -----
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(metacal)
```

Multiplicative model of PCR Bias

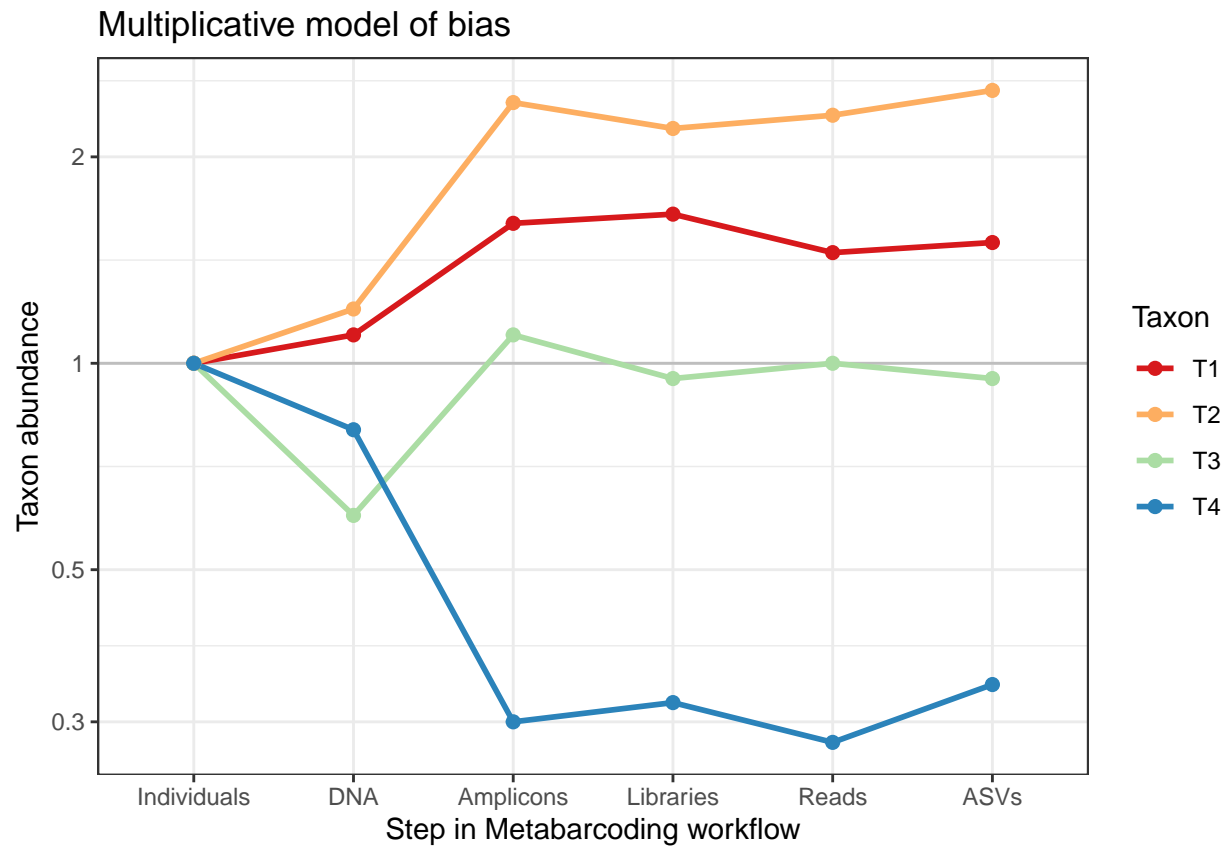
If bias is multiplicative - we can try estimate this multiplier and fit correction factors

```
tb2 <- tibble(
  Taxon = c("T1", "T2", "T3", "T4"),
  Individuals = c(1, 1, 1, 1),
  DNA = c(1.1, 1.2, 0.6, 0.8),
  Amplicons = c(1.6, 2.4, 1.1, 0.3),
  Libraries = c(1.65, 2.2, 0.95, 0.32),
  Reads = c(1.45, 2.3, 1, 0.28),
  ASVs = c(1.5, 2.5, 0.95, 0.34)
) %>%
  gather("Type", "Abundance", -Taxon)

gg.multi <- tb2 %>%
  ggplot(aes(x=Type, y=Abundance, color = Taxon)) +
  geom_hline(yintercept=1,
             aes(group = Abundance), color = "grey") +
  geom_path(aes(group = Taxon), size=1) +
  geom_point(size=2) +
  scale_y_log10(breaks = c(0.1, 0.3, 0.5, 1, 2, 3), labels=c(0.1, 0.3, 0.5, 1, 2, 3)) +
  scale_x_discrete(limits = c("Individuals", "DNA", "Amplicons", "Libraries", "Reads", "ASVs")) +
  scale_color_brewer(palette="Spectral") +
  ggtitle("Multiplicative model of bias") +
  xlab("Step in Metabarcoding workflow") +
  ylab("Taxon abundance") +
  theme_bw()
```

```
## Warning in geom_hline(yintercept = 1, aes(group = Abundance), color = "grey"): Using both `yintercept`
```

```
print(gg.multi)
```



Problem of compositional data

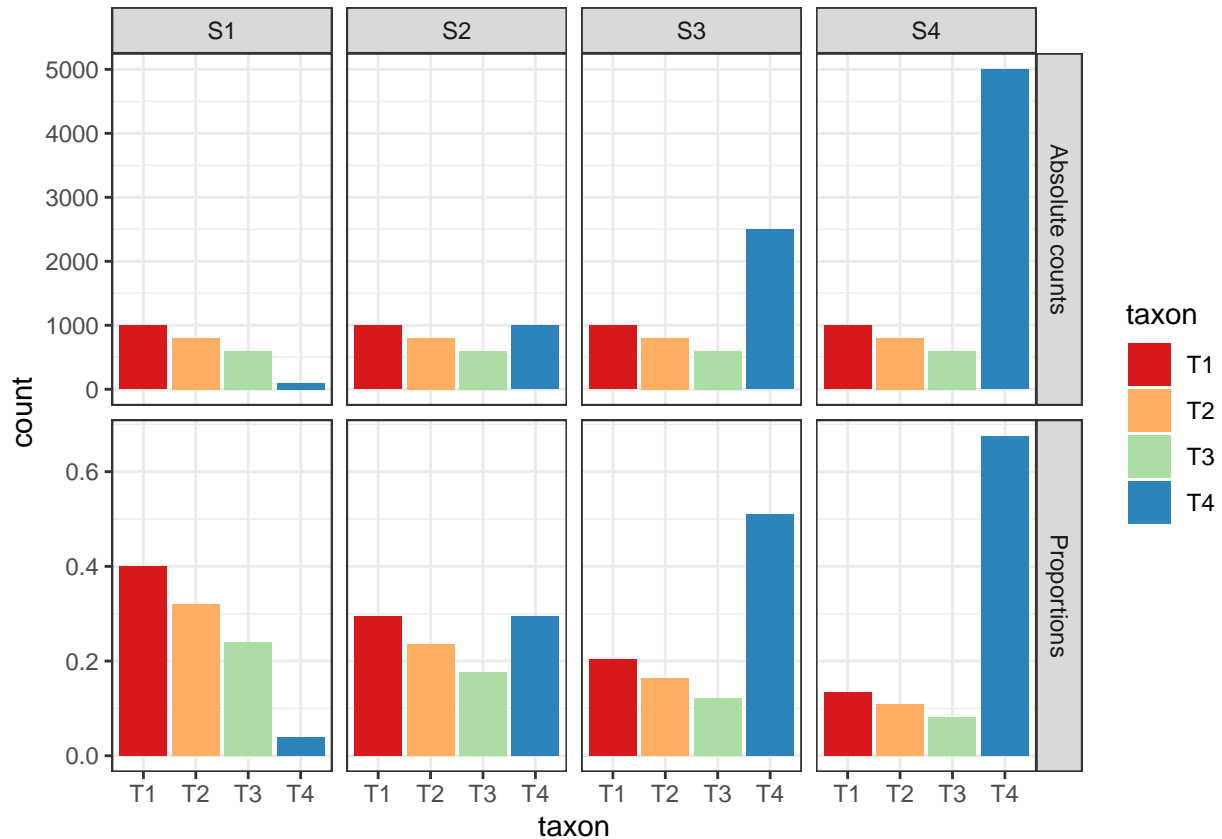
```
counts <- as.data.frame(t(data.frame( T1 = c(1000,1000,1000,1000),
                                      T2 = c(800,800,800,800),
                                      T3 = c(600,600,600,600),
                                      T4 = c(100,1000,2500,5000) )))

prop <- as_tibble(sapply(counts,function(x)(x / sum(x)))) %>%
  rename_all(~ c("S1", "S2", "S3", "S4")) %>%
  mutate(taxon=c("T1","T2","T3","T4")) %>%
  mutate(type="Proportions")

sim <- as_tibble(counts,rownames="taxon") %>%
  rename_all(~ c("taxon","S1", "S2", "S3", "S4")) %>%
  mutate(type="Absolute counts") %>%
  bind_rows(prop) %>%
  gather(sample,count,-taxon,-type)

ggplot(data = sim, aes(x=taxon,y=count, fill=taxon)) +
  geom_bar(stat="identity") +
```

```
facet_grid(type~sample,scales = "free_y") +
scale_fill_brewer(palette="Spectral") +
theme_bw()
```



Taxon proportions vs taxon ratios

```
#devtools::install_github("mikemc/metacal")

library(metacal)

#Create new tibble - by setting T1 to 1 across all, we are looking at the ratio to it
tb <- tibble(
  Taxon = c("T1", "T2", "T3", "T4"),
  Bias = c(1, 18, 6, 4),
  S1 = c(1, 1, 1, 1),
  S2 = c(1, 1/15, 4/15, 2/15),
  S3 = c(1, 4/15, 1/15, 2/15),
)

tb <- tb %>%
  gather("Sample", "Actual", S1:S3) %>%
  mutate(Observed = Actual * Bias) %>%
```

```
gather("Type", "Abundance", Actual, Observed) %>%
mutate_by(c(Sample, Type), Proportion = close_elts(Abundance)) %>%
select(Sample, Taxon, Bias, everything())
tb
```

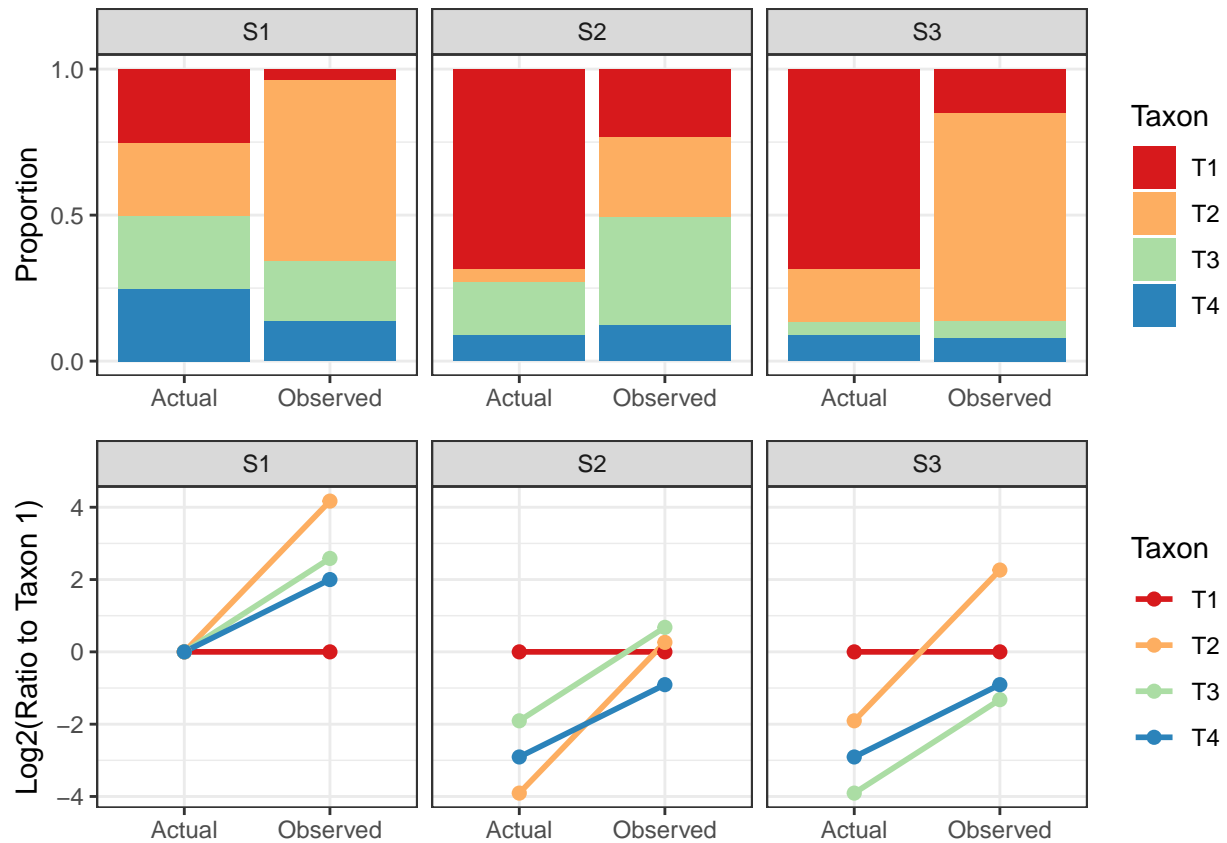
```
## # A tibble: 24 x 6
##   Sample Taxon Bias Type   Abundance Proportion
##   <chr>  <chr> <dbl> <chr>     <dbl>     <dbl>
## 1 S1     T1      1 Actual     1         0.25
## 2 S1     T2     18 Actual     1         0.25
## 3 S1     T3      6 Actual     1         0.25
## 4 S1     T4      4 Actual     1         0.25
## 5 S2     T1      1 Actual     1         0.682
## 6 S2     T2     18 Actual    0.0667     0.0455
## 7 S2     T3      6 Actual    0.267     0.182
## 8 S2     T4      4 Actual    0.133     0.0909
## 9 S3     T1      1 Actual     1         0.682
## 10 S3    T2     18 Actual    0.267     0.182
## # ... with 14 more rows
```

#We can view the measurement error across the three samples with bar plots, as done in main text Figure

```
p.props <- ggplot(tb, aes(x = Type, y = Proportion, fill = Taxon)) +
  geom_bar(stat = "identity") +
  scale_fill_brewer(palette = "Spectral") +
  facet_wrap(~Sample) +
  scale_y_continuous(breaks = c(0, 0.5, 1))+
  theme_bw() +
  theme(strip.text = element_text(size = 9), axis.title.x = element_blank(),
        legend.position = "right")

p.ratios <- ggplot(tb, aes(x = Type, y = log2(Abundance), color = Taxon)) +
  geom_path(aes(group = Taxon),size=1) +
  geom_point(size=2) +
  facet_wrap(~Sample) +
  scale_color_brewer(palette="Spectral") +
  labs(y = "Log2(Ratio to Taxon 1)") +
  theme_bw() +
  theme(strip.text = element_text(size = 9), axis.title.x = element_blank(),
        legend.position = "right")

library(patchwork)
p.props / p.ratios
```

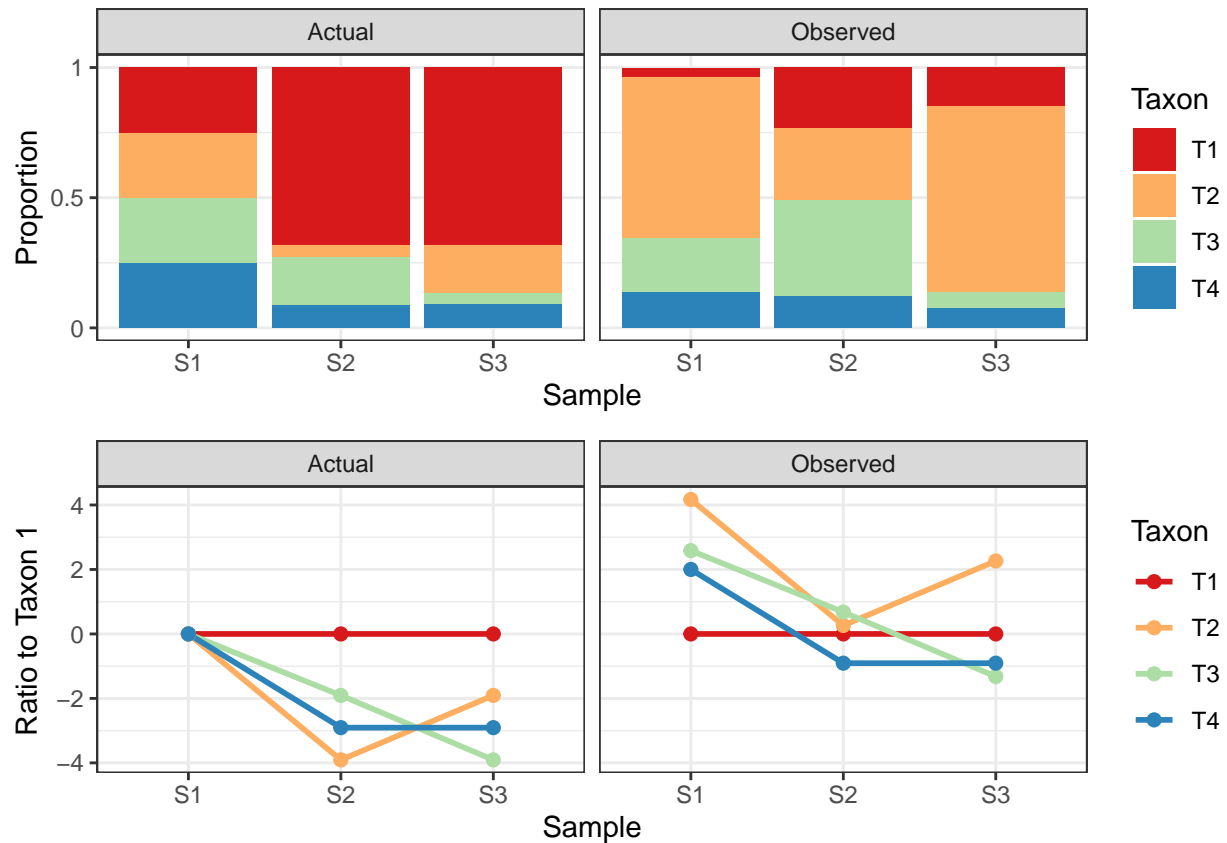


#Compare actual vs observed whole communities

```
p2.props <- ggplot(tb, aes(x = Sample, y = Proportion, fill = Taxon)) +
  geom_bar(stat = "identity") +
  facet_wrap(~Type) +
  scale_y_continuous(breaks = c(0, 0.5, 1), labels = c(0, 0.5, 1)) +
  scale_fill_brewer(palette="Spectral")+
  theme_bw() +
  theme(strip.text = element_text(size = 9))

p2.ratios <- ggplot(tb, aes(x = Sample, y = log2(Abundance), color = Taxon)) +
  geom_path(aes(group = Taxon),size=1) +
  geom_point(size=2) +
  facet_wrap(~Type) +
  scale_color_brewer(palette = "Spectral") +
  labs(y = "Ratio to Taxon 1") +
  theme_bw() +
  theme(strip.text = element_text(size = 9))

p2.props / p2.ratios
```



Compare distances between samples

```
B <- c(1, 18, 6)
A1 <- c(1, 1, 1)
A2 <- c(1, 1/15, 4/15)
A3 <- c(1, 4/15, 1/15)
O1 <- close_elts(A1 * B)
O2 <- close_elts(A2 * B)
O3 <- close_elts(A3 * B)

#Bray curtis similartiy
# Actual
xydist(A1, A2, method = "bray")
```

```
## [1] 0.4166667
```

```
xydist(A1, A3, method = "bray")
```

```
## [1] 0.4166667
```

```
xydist(A2, A3, method = "bray")
```

```
## [1] 0.15
```

```
# Observed  
xydist(O1, O2, method = "bray")
```

```
## [1] 0.4042105
```

```
xydist(O1, O3, method = "bray")
```

```
## [1] 0.1754839
```

```
xydist(O2, O3, method = "bray")
```

```
## [1] 0.4584041
```

```
#Aitchinson distance - invariant to bias  
# Actual  
xydist(A1, A2, method = "aitchison")
```

```
## [1] 1.915062
```

```
xydist(A1, A3, method = "aitchison")
```

```
## [1] 1.915062
```

```
xydist(A2, A3, method = "aitchison")
```

```
## [1] 1.960516
```

```
# Observed  
xydist(O1, O2, method = "aitchison")
```

```
## [1] 1.915062
```

```
xydist(O1, O3, method = "aitchison")
```

```
## [1] 1.915062
```

```
xydist(O2, O3, method = "aitchison")
```

```
## [1] 1.960516
```

Sessioninfo

```
sessionInfo()
```

```
## R version 3.6.1 (2019-07-05)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 17763)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_Australia.1252 LC_CTYPE=English_Australia.1252
## [3] LC_MONETARY=English_Australia.1252 LC_NUMERIC=C
## [5] LC_TIME=English_Australia.1252
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] patchwork_0.0.1 metacal_0.1.0   forcats_0.4.0   stringr_1.4.0
## [5] dplyr_0.8.3     purrr_0.3.3     readr_1.3.1     tidyr_1.0.0
## [9] tibble_2.1.3    ggplot2_3.2.1   tidyverse_1.2.1
##
## loaded via a namespace (and not attached):
## [1] tidymodels_0.2.5  xfun_0.10       reshape2_1.4.3
## [4] haven_2.1.1       lattice_0.20-38 colorspace_1.4-1
## [7] vctrs_0.2.0       generics_0.0.2  htmltools_0.4.0
## [10] yaml_2.2.0        utf8_1.1.4      rlang_0.4.1
## [13] pillar_1.4.2      glue_1.3.1      withr_2.1.2
## [16] RColorBrewer_1.1-2 modelr_0.1.5     readxl_1.3.1
## [19] lifecycle_0.1.0   plyr_1.8.4      munsell_0.5.0
## [22] gtable_0.3.0      cellranger_1.1.0 rvest_0.3.4
## [25] evaluate_0.14     labeling_0.3     knitr_1.25
## [28] fansi_0.4.0       broom_0.5.2     Rcpp_1.0.2
## [31] scales_1.0.0      backports_1.1.5 jsonlite_1.6
## [34] useful_1.2.6      hms_0.5.2       digest_0.6.22
## [37] stringi_1.4.3     grid_3.6.1      cli_1.1.0
## [40] tools_3.6.1       magrittr_1.5     lazyeval_0.2.2
## [43] crayon_1.3.4      pkgconfig_2.0.3 zeallot_0.1.0
## [46] ellipsis_0.3.0    xml2_1.2.2      lubridate_1.7.4
## [49] assertthat_0.2.1  rmarkdown_1.16  httr_1.4.1
## [52] rstudioapi_0.10   R6_2.4.0        nlme_3.1-141
## [55] compiler_3.6.1
```