nature portfolio

https://doi.org/10.1038/s41591-021-01383-w

Supplementary information

Integrated genomic, epidemiologic investigation of *Candida auris* skin colonization in a skilled nursing facility

In the format provided by the authors and unedited

Supplemental Information

Integrated genomic, epidemiologic investigation of *Candida auris* skin colonization in a skilled nursing facility

Diana M Proctor[1], Thelma Dangana[2], D. Joseph Sexton[3], Christine Fukuda[2], Rachel D Yelin[2], Mary Stanley[2], Pamela B Bell[2], Sangeetha Baskaran[2], Clay Deming[1], Qiong Chen[1], Sean Conlan[1], Morgan Park[4], NISC Comparative Sequencing Program[4], Rory M Welsh[3], Snigdha Vallabhaneni[3,5], Tom Chiller[3], Kaitlin Forsberg[3], Stephanie R. Black[6], Massimo Pacilli[6], Heidi H Kong[7], Michael Y Lin[2], Michael E Schoeny[8], Anastasia P Litvintseva[3], Julia A Segre[1,9*], Mary K Hayden[2,9*]

[1]Microbial Genomics Section, Translational and Functional Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

[2] Department of Internal Medicine, Division of Infectious Diseases, Rush University Medical Center, Chicago, IL 60612, USA.

[3]Mycotic Diseases Branch, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA.

[4]NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA.

[5]Division of Healthcare Quality Promotion, NCEZID, CDC, USA

[6]Communicable Disease Program, Chicago Department of Public Health, Chicago, IL, 60612, USA.

[7]Dermatology Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD 20892, USA

[8]College of Nursing, Rush University, Chicago, IL 60612, USA.

[9]These authors contributed equally: Julia A Segre and Mary K Hayden.
*e-mail: jsegre@nhgri.nih.gov; mhayden@rush.edu

Keywords
Candida auris, fungal pathogen, skin colonization, healthcare epidemiology, mycobiome, dynamics, medical mycology, long-term care, chlorhexidine

Supplementary Data

Supplementary Data 1

Excel workbook containing the ASV table, taxonomy table, and sample data mapping files for the merged_16s, merged_ITS, bac_match, and ITS_match datasets

Supplementary Data 2

Script needed to reproduce Figure 1 and its supporting figures

Supplementary Data 3

Script needed to reproduce Figure 2 and its supporting figures

Supplementary Data 4

Script needed to reproduce Figure 3 and its supporting figures

Supplementary Data 5

Script needed to reproduce Figure 4 and its supporting figures

47

48

49 ## Table of Contents

61

62

63

Supplementary Tables

Supplementary Table 1

Body Site	Number of samples analyzed			
	Semi-quantitative Culture[1]	Most probable number[2]	Microbiome[3]	CHG[4]
External auditory canal (Ea)	141	--	141	--
Axilla (Ax)	142	142	142	141
Perianal skin (An)	125	--	125	110
Toe Web (Tw)	139	--	139	138
Anterior Nares (N)	142	142	142	--
Buccal mucosa /tongue (Bu/To)	141	--	141	--
Tracheostomy (Tc)	102	--	102	--
Neck (Ne)	143	--	143	143
Palm/fingertips (Fg)	141	--	141	140
Inguinal Crease (Ic)	142	142	142	141
TOTAL	1,358	426	1,358	813

Supplementary Table 2

Supplementary Table 2.* MIC_{50} and MIC_{90} values for chlorhexidine gluconate and antifungal medications (n=213 isolates)

Medication Name	MIC_{50} (μg/mL)	MIC_{90} (μg/mL)
Chlorhexidine gluconate	16	32
Amphotericin B	1	1
Anidulafungin	0.12	0.25
Micafungin	0.06	0.12
Caspofungin	0.12	0.12
Fluconazole	2	4
Voriconazole	0.015	0.03

*MIC_{50} and MIC_{90} values are the concentration of antiseptic or antifungal agent needed to inhibit growth of 50% or 90%, respectively, of isolates tested in a broth microdilution testing system.

Abbreviation: MIC, minimum inhibitory concentration

Weighted (inverse of number of body sites by patient) estimates of the 50th and 90th percentiles were calculated for each antiseptic or antifungal agent.

Supplementary Data 1 – See Excel file

Supplementary Data 2

Figure 1: Calculating Sensitivity and Confidence Intervals

Diana Proctor

9/8/2020

last updated: February 24 2021

---

draft: submission to Nature Medicine October 2020, Revised submission Feb 2021

Manuscript Title: Integrated genomic, epidemiologic investigation of endemic Candida auris skin colonization

Authors: Diana M Proctor1, Thelma Dangana2, D. Joseph Sexton3, Christine Fukuda2, Rachel D Yelin2, Mary Stanley2, Pamela B Bell2, Sangeetha Baskaran2, Clay Deming1, Qiong Chen1, Sean Conlan1, Morgan Park4, NISC Comparative Sequencing Program4, Rory M Welsh3, Snigdha Vallabhaneni3,5, Tom Chiller3, Kaitlin Forsberg3, Stephanie R. Black6, Massimo Pacilli6, Heidi H Kong7, Michael Y Lin2, Michael E Schoeny8, Anastasia P Litvintseva3, Julia A Segre1+, Mary K Hayden2+

Affiliations:

- 1Microbial Genomics Section, Translational and Functional Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA
- 2 Department of Internal Medicine, Division of Infectious Diseases, Rush University Medical Center, Chicago, IL 60612, USA.
- 3Mycotic Diseases Branch, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA.
- 4NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA.
- 5Division of Healthcare Quality Promotion, NCEZID, CDC, USA
- 6Communicable Disease Program, Chicago Department of Public Health, Chicago, IL, 60612, USA.
- 7Dermatology Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD 20892, USA
- 8College of Nursing, Rush University, Chicago, IL 60612, USA.
- +Contributed equally

Description of the dataset

Here, we seek to evaluate the total body site occupancy of Candida auris on the body, as well as estimate prevalence (defined in this script as the proportion of all patients colonized using different combinations of body sites), as well as compute the sensitivity of each body site in capturing the

percentage of patients who are colonized by Candida auris, at any body site. Sensitivity is defined as the estimated prevalence / number of colonized patients, according to any body site. We compute these estimates based on culture results for each of 57 patients.

To accomplish this, we read in the following data:

1. sitecode_to_factored_sites.csv
2. Cauris_Analytic_2020-5-20.csv

Here, we render the following figures:

- Figure 1
- Supplementary Figure 3: Ridgeline Plot of CFUs
- Supplementary Figure 2: UpsetR plot of body site colonization patterns
- Supplementary Figure 4: Paired MPN Analysis

Install and Load R packages

```
 #load package method from from Dan Sprokett

# set seed
set.seed(78979)

#define packages
packages <- c("knitr",
              "tidyverse",
              "epiR",
              "viridis",
              "scales",
              "kableExtra",
              "UpSetR",
              "reshape2",
              "gridExtra",
              "phyloseq",
              "wesanderson",
              "harrypotter",
              "ggridges",
              "ggpubr",
              "dabestr",
              "ComplexHeatmap",
              "scales",
              "cowplot",
              "Hmisc")

# install packages from bioconductor
BiocManager::install(setdiff(packages,installed.packages()), update=FALSE)
n <- length(packages) - sum(sapply(packages,require,character.only=TRUE))

# print if packages loaded properly
if(n == 0){
  print("All necessary R packages loaded properly")
} else {
  print(paste0(n, " R packages did not load properly"))
}
```

```
## [1] "All necessary R packages loaded properly"
```

define knitr options

```
knitr::opts_chunk$set(echo=TRUE, warning=FALSE, message=FALSE, error = FALSE)
```

1. Let's see how to compute prevalence estimates and confidence intervals using the epi.test pacckage

We will use the epiR package to get the body site sensitivities for Candida auris colonization.

Compute the sensitivity of each body site for classification of colonization positive patients. Note that we do not have information for false negatives or false positives. We will use only information regarding true positives and true negatives in these calculations.

We define a function that takes as input 3 arguments:

- df is dataframe, having the columns (SiteID, Survey_Period);
- ref_sites is a vector of sites for which the sensitivity is being defined;
- time is the numeric survey period to analyze.

The output is a data frame with:

- the fraction of sites that were positive (TP/TP+TN)
- the lower 95% CI
- the upper 95% CI
- the ref_sites
- ncompare, which is the length of ref_sites, which is = number of sites being compared.

We use the epi.test package as before.

```
get_sensitivity_interval_notime <- function(df,  ref_sites){
  df=  df[df[["SiteID"]] %in% ref_sites,]

    #get the numbers of positives and negatives
    df = doBy::summaryBy(Cauris_Result~Unique_ptid, data=df, FUN=sum, na.rm=TRUE)
    df$Count = ifelse(df$Cauris_Result.sum > 0, 1, 0)
    TP = sum(df$Count == 1)
    TN = sum(df$Count == 0)

    dat <- as.table(matrix(c(TP, 0, TN, 0), nrow=2, byrow=TRUE))
    out = epi.tests(dat, conf.level = 0.95)
    out1 = out$rval$se
    out1$ncompare = length(ref_sites)
    rownames(out1) = toString(ref_sites)

    return(out1)

}
```

Let's look at a table provided by collaborators as a test case

First we test this package out on the data set in the paper Thurlow et al, 2013. Anatomic sites of patient colonization and environmental contamination with Klebsiella pneumoniae Carbapenemase-producing Enterobacteriaceae at Long Term Acute Care Hospitals. Infection Control & Hospital Epidemiology. 34(1). We use the data from table 2 as an example. This is the table from which data are drawn.

TABLE 2. Sensitivity of Culture of Different Anatomic Sites for *Klebsiella pneumoniae* Carbapenemase–Producing Enterobacteriaceae

	No. of positive cultures (N = 24)	Sensitivity, % (95% CI)
Skin sites		
Inguinal	19	79 (58–93)
Axillary	18	75 (53–90)
Upper back	6	25 (10–47)
Antecubital fossae	6	25 (10–47)
Nonskin sites		
Rectal[a]	21	88 (68–97)
Urine (N = 19)[b]	10	53 (29–76)
Oropharyngeal/tracheal secretions	10	42 (22–63)
Combined sites		
Rectal and inguinal	24	100 (86–100)
Rectal and axillary	23	96 (79–100)
Axillary and inguinal	22	92 (73–99)

NOTE. CI, confidence interval.
[a] Three patients had negative rectal swab cultures but positive cultures of inguinal skin.
[b] Five patients were anuric, so urine was not collected for culture.

Table 2

What does our table look like? Okay, this looks good. Our estimates and confidence intervals are the same as in the published paper. Let's move on to our data.

Replicate Table 2

	est	lower	upper
Skin sites			
Inguinal	0.79	0.58	0.93
Axillary	0.75	0.53	0.90
Upper Back	0.25	0.10	0.47
Antecubital fossa	0.25	0.10	0.47
Nonskin Sites			
Rectal	0.88	0.68	0.97
Urine	0.53	0.29	0.76
Combined sites			
Oropharyngeal	0.42	0.22	0.63
Rectal and Inguinal	1.00	0.86	1.00
Retal and Axillary	0.96	0.79	1.00
Axillary and Inguinal	0.92	0.73	0.99

Note:
CI = Confidence Interval

2.Now let's make the equivalent table for the Candida auris paper.

First read in the data

```
site_codes = read.csv("~/Desktop/candida_auris_rush/sitecode_to_factored_sites.csv")
data = read.csv("~/Desktop/candida_auris_rush/manuscript/data/Cauris_Analytic_2020-5-20.csv") %>%
  dplyr::select(., c("Unique_ptid", "Survey_Period", "Cauris_Result", "site")) %>%
  plyr::join(., site_codes)
mpn = read.csv("~/Desktop/candida_auris_rush/manuscript/data/Cauris_Analytic_2020-5-20.csv") %>%
  dplyr::select(., c("Unique_ptid", "Survey_Period", "Cauris_Result", "site", "CDC_MPN", "CFUTransform_Cauris")) %>%
  plyr::join(., site_codes)
```

how many subjects were included in each survey?

- survey 1: 56
- survey 2: 45
- survey 3: 43

```
#how many subjects are in the first survey
total=subset(data, Survey_Period==1)
totalN = length(unique(total$Unique_ptid))
totalN
```

```
## [1] 56
```

```
#how many subjects are in the seccond survey
total=subset(data, Survey_Period==2)
totalN = length(unique(total$Unique_ptid))
totalN
```

```
## [1] 45
```

```
#how many subjects are in the third survey
total=subset(data, Survey_Period==3)
totalN = length(unique(total$Unique_ptid))
totalN
```

```
## [1] 43
```

Define the working datset - newdata1

- this includes 57 subjects
- 56 were surveyed on survey 1

- 1 (subject 29) was surveyed on survey 2

```
#############define the dataset
# for the dataset we have 57 subjects
#56 were sampled on survey 1
#1 was sampled on survey 2 (subject 29)
########################
############# Let's subset on the first time point so that we don't have 2-3 entr
ies per patient
newdata1 = subset(data, Survey_Period==1)

#note that subject 29 was not sampled on day 1, just on day 2, so let's grab that
individual, subsetting on time point 2 only
subject29 = subset(data, Unique_ptid=="29" & Survey_Period=="2")
newdata1 = data.frame(rbind(newdata1, subject29))

#verify we have 57 subjects
length(unique(newdata1$Unique_ptid))
```

```
## [1] 57
```

One-way: Site-wise computation of proportion of colonized patients based on each site

Let's run the function for the various comparisons of interest

Sensitivity Analysis

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N	0.4285714	0.2971209	0.5678410	1
Fg	0.4035088	0.2756127	0.5417866	1
Tw	0.3571429	0.2335548	0.4964069	1
An	0.3529412	0.2243062	0.4993180	1
Ic	0.3392857	0.2181093	0.4781137	1
Ax	0.3035714	0.1877987	0.4409674	1
Bu/To	0.2142857	0.1159222	0.3443954	1
Tc	0.1219512	0.0408067	0.2620447	1
Ne	0.0877193	0.0290986	0.1929573	1
Ea	0.0714286	0.0198039	0.1729047	1

all two way

- we want to pick the combination of sites (1-way, 2-way, 3-way, etc.) that adds a marginal benefit to

the prior set of samples to a master table

- we sort the 2-way comparisons to identify the 2-way site comparisons that maximize our estimate of the proportion of positive patients

```
set.seed(91)
alltwosites = combn(unique(newdata1$SiteID)[1:10], 2, simplify = TRUE)


 holder <- vector("list", ncol(alltwosites))
  for(i in 1:ncol(alltwosites)) {
        twoway1  = get_sensitivity_interval_notime(df=newdata1,
                                                   ref_sites = alltwosites[,i])
     holder[[i]] =  twoway1
}

CI.twoway = do.call("rbind", holder)
colnames(CI.twoway) = c("Prevalence", "95% CI, Lower", "95% CI, Upper", "N.Site
s")
CI.twoway = CI.twoway %>%
  arrange(., desc(`Prevalence`))

kable(CI.twoway, caption = "Sensitivity Analysis") %>%
  kable_styling("basic", full_width = F, bootstrap_options ="condensed")
```

Sensitivity Analysis

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Fg	0.6140351	0.4757484	0.7400453	2
N, Tw	0.5714286	0.4321590	0.7028791	2
Fg, Tw	0.5614035	0.4236214	0.6925846	2
N, Ic	0.5535714	0.4147167	0.6865639	2
Ic, Fg	0.5438596	0.4065597	0.6764458	2
Fg, An	0.5438596	0.4065597	0.6764458	2
N, Ax	0.5357143	0.3974360	0.6700853	2
N, An	0.5178571	0.3803158	0.6534451	2
Ax, Ic	0.5000000	0.3633554	0.6366446	2
Ic, Tw	0.5000000	0.3633554	0.6366446	2
Tw, An	0.5000000	0.3633554	0.6366446	2
Fg, Bu/To	0.4912281	0.3562965	0.6271021	2

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
Ax, Tw	0.4821429	0.3465549	0.6196842	2
Ic, An	0.4821429	0.3465549	0.6196842	2
Ax, Fg	0.4736842	0.3398483	0.6103478	2
Ax, An	0.4642857	0.3299147	0.6025640	2
N, Ne	0.4561404	0.3235542	0.5934403	2
N, Bu/To	0.4464286	0.3134361	0.5852833	2
N, Ea	0.4464286	0.3134361	0.5852833	2
Ic, Bu/To	0.4464286	0.3134361	0.5852833	2
Bu/To, An	0.4464286	0.3134361	0.5852833	2
Fg, Ea	0.4385965	0.3074154	0.5763786	2
N, Tc	0.4285714	0.2971209	0.5678410	2
Ax, Bu/To	0.4285714	0.2971209	0.5678410	2
Tw, Bu/To	0.4285714	0.2971209	0.5678410	2
Fg, Ne	0.4210526	0.2914341	0.5591614	2
Fg, Tc	0.4210526	0.2914341	0.5591614	2
Tw, Ea	0.3928571	0.2649913	0.5324631	2
Ic, Ne	0.3859649	0.2599547	0.5242516	2
Tw, Ne	0.3859649	0.2599547	0.5242516	2
Ic, Ea	0.3750000	0.2491841	0.5145217	2
Ic, Tc	0.3750000	0.2491841	0.5145217	2
Tw, Tc	0.3750000	0.2491841	0.5145217	2
An, Tc	0.3703704	0.2429070	0.5125826	2
Ne, An	0.3684211	0.2444640	0.5065530	2
Ax, Ea	0.3571429	0.2335548	0.4964069	2
Ax, Tc	0.3392857	0.2181093	0.4781137	2
Ea, An	0.3392857	0.2181093	0.4781137	2
Ax, Ne	0.3333333	0.2140053	0.4706476	2
Bu/To, Ea	0.2857143	0.1729518	0.4220987	2

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
Bu/To, Ne	0.2807018	0.1697292	0.4154289	2
Bu/To, Tc	0.2321429	0.1297920	0.3641843	2
Ea, Tc	0.1607143	0.0762187	0.2832797	2
Ne, Tc	0.1578947	0.0748312	0.2786826	2
Ea, Ne	0.1228070	0.0508289	0.2367950	2

Three way

- we want to pick the combination of sites (1-way, 2-way, 3-way, etc.) that adds a marginal benefit to the prior set of samples to a master table
- we sort the 3-way comparisons to identify the 3-way site comparisons that maximize our estimate of the proportion of positive patients

```
all3 = combn(unique(newdata1$SiteID)[1:10], 3, simplify = TRUE)

 holder <- vector("list", ncol(all3))
  for(i in 1:ncol(all3)) {
        df  = get_sensitivity_interval_notime(df=newdata1, ref_sites = all3[,i])
      holder[[i]] =  df
}

CI.threeway = do.call("rbind", holder)
colnames(CI.threeway) = c("Prevalence", "95% CI, Lower", "95% CI, Upper", "N.Site
s")
CI.threeway = CI.threeway %>%
  arrange(., desc(`Prevalence`))

kable(CI.threeway, caption = "Sensitivity Analysis") %>%
  kable_styling("basic", full_width = F, bootstrap_options ="condensed")
```

Sensitivity Analysis

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Fg, Tw	0.7192982	0.5845711	0.8302708	3
N, Ic, Fg	0.7017544	0.5659736	0.8157119	3
N, Fg, An	0.6842105	0.5475702	0.8009499	3
Fg, Tw, An	0.6666667	0.5293524	0.7859947	3
N, Ax, Tw	0.6607143	0.5218863	0.7818907	3
N, Ax, Fg	0.6491228	0.5113134	0.7708544	3

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Ic	0.6428571	0.5035931	0.7664452	3
N, Fg, Bu/To	0.6315789	0.4934470	0.7555360	3
Ic, Fg, Tw	0.6315789	0.4934470	0.7555360	3
Ic, Fg, An	0.6315789	0.4934470	0.7555360	3
N, Ic, Tw	0.6250000	0.4854783	0.7508159	3
N, Ic, An	0.6250000	0.4854783	0.7508159	3
N, Tw, An	0.6250000	0.4854783	0.7508159	3
N, Fg, Ea	0.6140351	0.4757484	0.7400453	3
N, Fg, Ne	0.6140351	0.4757484	0.7400453	3
N, Fg, Tc	0.6140351	0.4757484	0.7400453	3
Ic, Fg, Bu/To	0.6140351	0.4757484	0.7400453	3
Fg, Bu/To, An	0.6140351	0.4757484	0.7400453	3
N, Ax, An	0.6071429	0.4675369	0.7350087	3
Ax, Fg, Tw	0.5964912	0.4582134	0.7243873	3
Ax, Fg, An	0.5964912	0.4582134	0.7243873	3
Fg, Tw, Bu/To	0.5964912	0.4582134	0.7243873	3
Fg, Tw, Ea	0.5964912	0.4582134	0.7243873	3
Ax, Ic, Tw	0.5892857	0.4497649	0.7190285	3
Ax, Ic, An	0.5892857	0.4497649	0.7190285	3
Ax, Tw, An	0.5892857	0.4497649	0.7190285	3
Ic, Tw, An	0.5892857	0.4497649	0.7190285	3
N, Tw, Ne	0.5789474	0.4408386	0.7085659	3
Ax, Ic, Fg	0.5789474	0.4408386	0.7085659	3
Ic, Fg, Ea	0.5789474	0.4408386	0.7085659	3
Fg, Tw, Ne	0.5789474	0.4408386	0.7085659	3
Fg, Tw, Tc	0.5789474	0.4408386	0.7085659	3
N, Ic, Bu/To	0.5714286	0.4321590	0.7028791	3
N, Tw, Bu/To	0.5714286	0.4321590	0.7028791	3

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Tw, Ea	0.5714286	0.4321590	0.7028791	3
N, Tw, Tc	0.5714286	0.4321590	0.7028791	3
Ax, Ic, Bu/To	0.5714286	0.4321590	0.7028791	3
Tw, Bu/To, An	0.5714286	0.4321590	0.7028791	3
N, Ic, Ne	0.5614035	0.4236214	0.6925846	3
Ax, Fg, Bu/To	0.5614035	0.4236214	0.6925846	3
Ic, Fg, Ne	0.5614035	0.4236214	0.6925846	3
Ic, Fg, Tc	0.5614035	0.4236214	0.6925846	3
Fg, Ne, An	0.5614035	0.4236214	0.6925846	3
Fg, An, Tc	0.5614035	0.4236214	0.6925846	3
N, Ax, Bu/To	0.5535714	0.4147167	0.6865639	3
N, Ax, Ea	0.5535714	0.4147167	0.6865639	3
N, Ic, Ea	0.5535714	0.4147167	0.6865639	3
N, Ic, Tc	0.5535714	0.4147167	0.6865639	3
Ax, Bu/To, An	0.5535714	0.4147167	0.6865639	3
Ic, Bu/To, An	0.5535714	0.4147167	0.6865639	3
N, Ax, Ne	0.5438596	0.4065597	0.6764458	3
N, Ne, An	0.5438596	0.4065597	0.6764458	3
Fg, Ea, An	0.5438596	0.4065597	0.6764458	3
N, Ax, Tc	0.5357143	0.3974360	0.6700853	3
N, Ea, An	0.5357143	0.3974360	0.6700853	3
Ax, Ic, Ea	0.5357143	0.3974360	0.6700853	3
Ax, Tw, Bu/To	0.5357143	0.3974360	0.6700853	3
Ic, Tw, Bu/To	0.5357143	0.3974360	0.6700853	3
Ic, Tw, Ea	0.5357143	0.3974360	0.6700853	3
Ic, Tw, Ne	0.5263158	0.3896522	0.6601517	3
Fg, Bu/To, Ea	0.5263158	0.3896522	0.6601517	3
Tw, Ne, An	0.5263158	0.3896522	0.6601517	3

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Bu/To, An	0.5178571	0.3803158	0.6534451	3
N, An, Tc	0.5178571	0.3803158	0.6534451	3
Ax, Ic, Tc	0.5178571	0.3803158	0.6534451	3
Ax, Tw, Ea	0.5178571	0.3803158	0.6534451	3
Ic, Tw, Tc	0.5178571	0.3803158	0.6534451	3
Ic, An, Tc	0.5178571	0.3803158	0.6534451	3
Tw, An, Tc	0.5178571	0.3803158	0.6534451	3
Ax, Ic, Ne	0.5087719	0.3728979	0.6437035	3
Ax, Fg, Ea	0.5087719	0.3728979	0.6437035	3
Ic, Ne, An	0.5087719	0.3728979	0.6437035	3
Ax, Tw, Tc	0.5000000	0.3633554	0.6366446	3
Tw, Ea, An	0.5000000	0.3633554	0.6366446	3
Ax, Fg, Ne	0.4912281	0.3562965	0.6271021	3
Ax, Fg, Tc	0.4912281	0.3562965	0.6271021	3
Ax, Tw, Ne	0.4912281	0.3562965	0.6271021	3
Ax, Ne, An	0.4912281	0.3562965	0.6271021	3
Fg, Bu/To, Ne	0.4912281	0.3562965	0.6271021	3
Fg, Bu/To, Tc	0.4912281	0.3562965	0.6271021	3
Ax, Bu/To, Ea	0.4821429	0.3465549	0.6196842	3
Ax, Ea, An	0.4821429	0.3465549	0.6196842	3
Ax, An, Tc	0.4821429	0.3465549	0.6196842	3
Ic, Bu/To, Ea	0.4821429	0.3465549	0.6196842	3
Ic, Ea, An	0.4821429	0.3465549	0.6196842	3
N, Bu/To, Ne	0.4736842	0.3398483	0.6103478	3
Ic, Bu/To, Ne	0.4736842	0.3398483	0.6103478	3
Bu/To, Ne, An	0.4736842	0.3398483	0.6103478	3
N, Bu/To, Ea	0.4642857	0.3299147	0.6025640	3
Tw, Bu/To, Ea	0.4642857	0.3299147	0.6025640	3

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
Bu/To, Ea, An	0.4642857	0.3299147	0.6025640	3
N, Ea, Ne	0.4561404	0.3235542	0.5934403	3
N, Ne, Tc	0.4561404	0.3235542	0.5934403	3
Fg, Ea, Ne	0.4561404	0.3235542	0.5934403	3
Fg, Ea, Tc	0.4561404	0.3235542	0.5934403	3
N, Bu/To, Tc	0.4464286	0.3134361	0.5852833	3
N, Ea, Tc	0.4464286	0.3134361	0.5852833	3
Ic, Bu/To, Tc	0.4464286	0.3134361	0.5852833	3
Bu/To, An, Tc	0.4464286	0.3134361	0.5852833	3
Ax, Bu/To, Ne	0.4385965	0.3074154	0.5763786	3
Tw, Bu/To, Ne	0.4385965	0.3074154	0.5763786	3
Ax, Bu/To, Tc	0.4285714	0.2971209	0.5678410	3
Tw, Bu/To, Tc	0.4285714	0.2971209	0.5678410	3
Ic, Ea, Ne	0.4210526	0.2914341	0.5591614	3
Fg, Ne, Tc	0.4210526	0.2914341	0.5591614	3
Tw, Ea, Ne	0.4210526	0.2914341	0.5591614	3
Ic, Ea, Tc	0.4107143	0.2809715	0.5502351	3
Tw, Ea, Tc	0.4107143	0.2809715	0.5502351	3
Ic, Ne, Tc	0.4035088	0.2756127	0.5417866	3
Ax, Ea, Tc	0.3928571	0.2649913	0.5324631	3
Tw, Ne, Tc	0.3859649	0.2599547	0.5242516	3
Ne, An, Tc	0.3859649	0.2599547	0.5242516	3
Ea, An, Tc	0.3750000	0.2491841	0.5145217	3
Ax, Ea, Ne	0.3684211	0.2444640	0.5065530	3
Ea, Ne, An	0.3684211	0.2444640	0.5065530	3
Ax, Ne, Tc	0.3508772	0.2291456	0.4886866	3
Bu/To, Ea, Ne	0.3157895	0.1990501	0.4524298	3
Bu/To, Ea, Tc	0.3035714	0.1877987	0.4409674	3

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
Bu/To, Ne, Tc	0.2982456	0.1842881	0.4340264	3
Ea, Ne, Tc	0.1929825	0.1004722	0.3191098	3

Four way

- we want to pick the combination of sites (1-way, 2-way, 3-way, etc.) that adds a marginal benefit to the prior set of samples to a master table
- we sort the 4-way comparisons to identify the 4-way site comparisons that maximize our estimate of the proportion of positive patients

```
all4 = combn(unique(newdata1$SiteID)[1:10], 4, simplify = TRUE)

 holder <- vector("list", ncol(all4))
  for(i in 1:ncol(all4)) {
        df  = get_sensitivity_interval_notime(df=newdata1, ref_sites = all4[,i])
      holder[[i]] =  df
}

CI.fourway = do.call("rbind", holder)
colnames(CI.fourway) = c("Prevalence", "95% CI, Lower", "95% CI, Upper", "N.Site
s")
CI.fourway = CI.fourway %>%
  arrange(., desc(`Prevalence`))

kable(CI.fourway, caption = "Sensitivity Analysis") %>%
  kable_styling("basic", full_width = F, bootstrap_options ="condensed")
```

Sensitivity Analysis

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ic, Fg, Tw	0.7543860	0.6223897	0.8587314	4
N, Ic, Fg, An	0.7543860	0.6223897	0.8587314	4
N, Fg, Tw, An	0.7543860	0.6223897	0.8587314	4
N, Ax, Fg, Tw	0.7368421	0.6033725	0.8446152	4
N, Ax, Ic, Fg	0.7192982	0.5845711	0.8302708	4
N, Ax, Fg, An	0.7192982	0.5845711	0.8302708	4
N, Ic, Fg, Bu/To	0.7192982	0.5845711	0.8302708	4
N, Fg, Tw, Bu/To	0.7192982	0.5845711	0.8302708	4
N, Fg, Tw, Ea	0.7192982	0.5845711	0.8302708	4

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Fg, Tw, Ne	0.7192982	0.5845711	0.8302708	4
N, Fg, Tw, Tc	0.7192982	0.5845711	0.8302708	4
N, Ax, Ic, Tw	0.7142857	0.5779013	0.8270482	4
N, Ic, Fg, Ea	0.7017544	0.5659736	0.8157119	4
N, Ic, Fg, Ne	0.7017544	0.5659736	0.8157119	4
N, Ic, Fg, Tc	0.7017544	0.5659736	0.8157119	4
Ic, Fg, Tw, An	0.7017544	0.5659736	0.8157119	4
Fg, Tw, Bu/To, An	0.7017544	0.5659736	0.8157119	4
N, Ax, Ic, An	0.6964286	0.5590326	0.8122013	4
N, Ax, Tw, An	0.6964286	0.5590326	0.8122013	4
N, Fg, Bu/To, An	0.6842105	0.5475702	0.8009499	4
N, Fg, Ea, An	0.6842105	0.5475702	0.8009499	4
N, Fg, Ne, An	0.6842105	0.5475702	0.8009499	4
N, Fg, An, Tc	0.6842105	0.5475702	0.8009499	4
Ax, Fg, Tw, An	0.6842105	0.5475702	0.8009499	4
Ic, Fg, Bu/To, An	0.6842105	0.5475702	0.8009499	4
Fg, Tw, Ne, An	0.6842105	0.5475702	0.8009499	4
Fg, Tw, An, Tc	0.6842105	0.5475702	0.8009499	4
N, Ic, Tw, An	0.6785714	0.5403638	0.7971455	4
N, Ax, Fg, Bu/To	0.6666667	0.5293524	0.7859947	4
Ax, Ic, Fg, An	0.6666667	0.5293524	0.7859947	4
Ax, Fg, Bu/To, An	0.6666667	0.5293524	0.7859947	4
Ic, Fg, Tw, Bu/To	0.6666667	0.5293524	0.7859947	4
Ic, Fg, Tw, Ea	0.6666667	0.5293524	0.7859947	4
Fg, Tw, Ea, An	0.6666667	0.5293524	0.7859947	4
N, Ax, Ic, Bu/To	0.6607143	0.5218863	0.7818907	4
N, Ax, Tw, Bu/To	0.6607143	0.5218863	0.7818907	4
N, Ax, Tw, Ea	0.6607143	0.5218863	0.7818907	4

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Tw, Tc	0.6607143	0.5218863	0.7818907	4
Ax, Ic, Tw, An	0.6607143	0.5218863	0.7818907	4
N, Ax, Fg, Ea	0.6491228	0.5113134	0.7708544	4
N, Ax, Fg, Ne	0.6491228	0.5113134	0.7708544	4
N, Ax, Fg, Tc	0.6491228	0.5113134	0.7708544	4
N, Ax, Tw, Ne	0.6491228	0.5113134	0.7708544	4
Ax, Ic, Fg, Tw	0.6491228	0.5113134	0.7708544	4
Ax, Ic, Fg, Bu/To	0.6491228	0.5113134	0.7708544	4
Ic, Fg, Tw, Ne	0.6491228	0.5113134	0.7708544	4
Ic, Fg, Tw, Tc	0.6491228	0.5113134	0.7708544	4
Ic, Fg, Bu/To, Ea	0.6491228	0.5113134	0.7708544	4
Ic, Fg, Ne, An	0.6491228	0.5113134	0.7708544	4
Ic, Fg, An, Tc	0.6491228	0.5113134	0.7708544	4
N, Ax, Ic, Ea	0.6428571	0.5035931	0.7664452	4
N, Ax, Ic, Tc	0.6428571	0.5035931	0.7664452	4
Ax, Ic, Bu/To, An	0.6428571	0.5035931	0.7664452	4
Ax, Tw, Bu/To, An	0.6428571	0.5035931	0.7664452	4
N, Ax, Ic, Ne	0.6315789	0.4934470	0.7555360	4
N, Ic, Tw, Ne	0.6315789	0.4934470	0.7555360	4
N, Ic, Ne, An	0.6315789	0.4934470	0.7555360	4
N, Fg, Bu/To, Ea	0.6315789	0.4934470	0.7555360	4
N, Fg, Bu/To, Ne	0.6315789	0.4934470	0.7555360	4
N, Fg, Bu/To, Tc	0.6315789	0.4934470	0.7555360	4
N, Tw, Ne, An	0.6315789	0.4934470	0.7555360	4
Ax, Fg, Tw, Bu/To	0.6315789	0.4934470	0.7555360	4
Ax, Fg, Tw, Ea	0.6315789	0.4934470	0.7555360	4
Ic, Fg, Ea, An	0.6315789	0.4934470	0.7555360	4
Fg, Tw, Bu/To, Ea	0.6315789	0.4934470	0.7555360	4

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Ea, An	0.6250000	0.4854783	0.7508159	4
N, Ic, Tw, Bu/To	0.6250000	0.4854783	0.7508159	4
N, Ic, Tw, Ea	0.6250000	0.4854783	0.7508159	4
N, Ic, Tw, Tc	0.6250000	0.4854783	0.7508159	4
N, Ic, Bu/To, An	0.6250000	0.4854783	0.7508159	4
N, Ic, Ea, An	0.6250000	0.4854783	0.7508159	4
N, Ic, An, Tc	0.6250000	0.4854783	0.7508159	4
N, Tw, Bu/To, An	0.6250000	0.4854783	0.7508159	4
N, Tw, Ea, An	0.6250000	0.4854783	0.7508159	4
N, Tw, An, Tc	0.6250000	0.4854783	0.7508159	4
Ax, Ic, Tw, Bu/To	0.6250000	0.4854783	0.7508159	4
Ax, Ic, Tw, Ea	0.6250000	0.4854783	0.7508159	4
Ic, Tw, Bu/To, An	0.6250000	0.4854783	0.7508159	4
N, Ax, Ne, An	0.6140351	0.4757484	0.7400453	4
N, Fg, Ea, Ne	0.6140351	0.4757484	0.7400453	4
N, Fg, Ea, Tc	0.6140351	0.4757484	0.7400453	4
N, Fg, Ne, Tc	0.6140351	0.4757484	0.7400453	4
Ax, Ic, Fg, Ea	0.6140351	0.4757484	0.7400453	4
Ax, Fg, Tw, Ne	0.6140351	0.4757484	0.7400453	4
Ax, Fg, Tw, Tc	0.6140351	0.4757484	0.7400453	4
Ax, Fg, Ne, An	0.6140351	0.4757484	0.7400453	4
Ax, Fg, An, Tc	0.6140351	0.4757484	0.7400453	4
Ic, Fg, Bu/To, Ne	0.6140351	0.4757484	0.7400453	4
Ic, Fg, Bu/To, Tc	0.6140351	0.4757484	0.7400453	4
Ic, Tw, Ne, An	0.6140351	0.4757484	0.7400453	4
Fg, Tw, Ea, Ne	0.6140351	0.4757484	0.7400453	4
Fg, Tw, Ea, Tc	0.6140351	0.4757484	0.7400453	4
Fg, Bu/To, Ea, An	0.6140351	0.4757484	0.7400453	4

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
Fg, Bu/To, Ne, An	0.6140351	0.4757484	0.7400453	4
Fg, Bu/To, An, Tc	0.6140351	0.4757484	0.7400453	4
N, Ax, Bu/To, An	0.6071429	0.4675369	0.7350087	4
N, Ax, An, Tc	0.6071429	0.4675369	0.7350087	4
Ax, Ic, Tw, Tc	0.6071429	0.4675369	0.7350087	4
Ax, Ic, Bu/To, Ea	0.6071429	0.4675369	0.7350087	4
Ax, Ic, An, Tc	0.6071429	0.4675369	0.7350087	4
Ax, Tw, An, Tc	0.6071429	0.4675369	0.7350087	4
Ic, Tw, An, Tc	0.6071429	0.4675369	0.7350087	4
Ax, Ic, Fg, Ne	0.5964912	0.4582134	0.7243873	4
Ax, Ic, Fg, Tc	0.5964912	0.4582134	0.7243873	4
Ax, Ic, Tw, Ne	0.5964912	0.4582134	0.7243873	4
Ax, Ic, Ne, An	0.5964912	0.4582134	0.7243873	4
Ax, Fg, Bu/To, Ea	0.5964912	0.4582134	0.7243873	4
Ax, Fg, Ea, An	0.5964912	0.4582134	0.7243873	4
Ax, Tw, Ne, An	0.5964912	0.4582134	0.7243873	4
Ic, Fg, Ea, Ne	0.5964912	0.4582134	0.7243873	4
Ic, Fg, Ea, Tc	0.5964912	0.4582134	0.7243873	4
Fg, Tw, Bu/To, Ne	0.5964912	0.4582134	0.7243873	4
Fg, Tw, Bu/To, Tc	0.5964912	0.4582134	0.7243873	4
Ax, Ic, Ea, An	0.5892857	0.4497649	0.7190285	4
Ax, Tw, Ea, An	0.5892857	0.4497649	0.7190285	4
Ic, Tw, Ea, An	0.5892857	0.4497649	0.7190285	4
N, Ic, Bu/To, Ne	0.5789474	0.4408386	0.7085659	4
N, Tw, Bu/To, Ne	0.5789474	0.4408386	0.7085659	4
N, Tw, Ea, Ne	0.5789474	0.4408386	0.7085659	4
N, Tw, Ne, Tc	0.5789474	0.4408386	0.7085659	4
Fg, Tw, Ne, Tc	0.5789474	0.4408386	0.7085659	4

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
Tw, Bu/To, Ne, An	0.5789474	0.4408386	0.7085659	4
N, Ax, Bu/To, Ea	0.5714286	0.4321590	0.7028791	4
N, Ic, Bu/To, Ea	0.5714286	0.4321590	0.7028791	4
N, Ic, Bu/To, Tc	0.5714286	0.4321590	0.7028791	4
N, Tw, Bu/To, Ea	0.5714286	0.4321590	0.7028791	4
N, Tw, Bu/To, Tc	0.5714286	0.4321590	0.7028791	4
N, Tw, Ea, Tc	0.5714286	0.4321590	0.7028791	4
Ax, Ic, Bu/To, Tc	0.5714286	0.4321590	0.7028791	4
Ax, Tw, Bu/To, Ea	0.5714286	0.4321590	0.7028791	4
Ax, Bu/To, Ea, An	0.5714286	0.4321590	0.7028791	4
Ic, Tw, Bu/To, Ea	0.5714286	0.4321590	0.7028791	4
Tw, Bu/To, Ea, An	0.5714286	0.4321590	0.7028791	4
Tw, Bu/To, An, Tc	0.5714286	0.4321590	0.7028791	4
N, Ax, Bu/To, Ne	0.5614035	0.4236214	0.6925846	4
N, Ic, Ea, Ne	0.5614035	0.4236214	0.6925846	4
N, Ic, Ne, Tc	0.5614035	0.4236214	0.6925846	4
Ax, Ic, Bu/To, Ne	0.5614035	0.4236214	0.6925846	4
Ax, Fg, Bu/To, Ne	0.5614035	0.4236214	0.6925846	4
Ax, Fg, Bu/To, Tc	0.5614035	0.4236214	0.6925846	4
Ax, Bu/To, Ne, An	0.5614035	0.4236214	0.6925846	4
Ic, Fg, Ne, Tc	0.5614035	0.4236214	0.6925846	4
Ic, Tw, Ea, Ne	0.5614035	0.4236214	0.6925846	4
Ic, Bu/To, Ne, An	0.5614035	0.4236214	0.6925846	4
Fg, Ea, Ne, An	0.5614035	0.4236214	0.6925846	4
Fg, Ea, An, Tc	0.5614035	0.4236214	0.6925846	4
Fg, Ne, An, Tc	0.5614035	0.4236214	0.6925846	4
N, Ax, Bu/To, Tc	0.5535714	0.4147167	0.6865639	4
N, Ax, Ea, Tc	0.5535714	0.4147167	0.6865639	4

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ic, Ea, Tc	0.5535714	0.4147167	0.6865639	4
Ax, Ic, Ea, Tc	0.5535714	0.4147167	0.6865639	4
Ax, Bu/To, An, Tc	0.5535714	0.4147167	0.6865639	4
Ic, Tw, Ea, Tc	0.5535714	0.4147167	0.6865639	4
Ic, Bu/To, Ea, An	0.5535714	0.4147167	0.6865639	4
Ic, Bu/To, An, Tc	0.5535714	0.4147167	0.6865639	4
N, Ax, Ea, Ne	0.5438596	0.4065597	0.6764458	4
N, Ax, Ne, Tc	0.5438596	0.4065597	0.6764458	4
N, Bu/To, Ne, An	0.5438596	0.4065597	0.6764458	4
N, Ea, Ne, An	0.5438596	0.4065597	0.6764458	4
N, Ne, An, Tc	0.5438596	0.4065597	0.6764458	4
Ax, Ic, Ea, Ne	0.5438596	0.4065597	0.6764458	4
Ic, Tw, Bu/To, Ne	0.5438596	0.4065597	0.6764458	4
N, Bu/To, Ea, An	0.5357143	0.3974360	0.6700853	4
N, Ea, An, Tc	0.5357143	0.3974360	0.6700853	4
Ax, Tw, Bu/To, Tc	0.5357143	0.3974360	0.6700853	4
Ax, Tw, Ea, Tc	0.5357143	0.3974360	0.6700853	4
Ic, Tw, Bu/To, Tc	0.5357143	0.3974360	0.6700853	4
Ax, Fg, Ea, Ne	0.5263158	0.3896522	0.6601517	4
Ax, Fg, Ea, Tc	0.5263158	0.3896522	0.6601517	4
Ax, Tw, Bu/To, Ne	0.5263158	0.3896522	0.6601517	4
Ax, Tw, Ea, Ne	0.5263158	0.3896522	0.6601517	4
Ic, Tw, Ne, Tc	0.5263158	0.3896522	0.6601517	4
Ic, Ne, An, Tc	0.5263158	0.3896522	0.6601517	4
Fg, Bu/To, Ea, Ne	0.5263158	0.3896522	0.6601517	4
Fg, Bu/To, Ea, Tc	0.5263158	0.3896522	0.6601517	4
Tw, Ea, Ne, An	0.5263158	0.3896522	0.6601517	4
Tw, Ne, An, Tc	0.5263158	0.3896522	0.6601517	4

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Bu/To, An, Tc	0.5178571	0.3803158	0.6534451	4
Ic, Ea, An, Tc	0.5178571	0.3803158	0.6534451	4
Tw, Ea, An, Tc	0.5178571	0.3803158	0.6534451	4
Ax, Ic, Ne, Tc	0.5087719	0.3728979	0.6437035	4
Ic, Bu/To, Ea, Ne	0.5087719	0.3728979	0.6437035	4
Ic, Ea, Ne, An	0.5087719	0.3728979	0.6437035	4
Ax, Ea, An, Tc	0.5000000	0.3633554	0.6366446	4
Ax, Fg, Ne, Tc	0.4912281	0.3562965	0.6271021	4
Ax, Tw, Ne, Tc	0.4912281	0.3562965	0.6271021	4
Ax, Ea, Ne, An	0.4912281	0.3562965	0.6271021	4
Ax, Ne, An, Tc	0.4912281	0.3562965	0.6271021	4
Fg, Bu/To, Ne, Tc	0.4912281	0.3562965	0.6271021	4
Ax, Bu/To, Ea, Tc	0.4821429	0.3465549	0.6196842	4
Ic, Bu/To, Ea, Tc	0.4821429	0.3465549	0.6196842	4
N, Bu/To, Ea, Ne	0.4736842	0.3398483	0.6103478	4
N, Bu/To, Ne, Tc	0.4736842	0.3398483	0.6103478	4
Ax, Bu/To, Ea, Ne	0.4736842	0.3398483	0.6103478	4
Ic, Bu/To, Ne, Tc	0.4736842	0.3398483	0.6103478	4
Tw, Bu/To, Ea, Ne	0.4736842	0.3398483	0.6103478	4
Bu/To, Ea, Ne, An	0.4736842	0.3398483	0.6103478	4
Bu/To, Ne, An, Tc	0.4736842	0.3398483	0.6103478	4
N, Bu/To, Ea, Tc	0.4642857	0.3299147	0.6025640	4
Tw, Bu/To, Ea, Tc	0.4642857	0.3299147	0.6025640	4
Bu/To, Ea, An, Tc	0.4642857	0.3299147	0.6025640	4
N, Ea, Ne, Tc	0.4561404	0.3235542	0.5934403	4
Fg, Ea, Ne, Tc	0.4561404	0.3235542	0.5934403	4
Ax, Bu/To, Ne, Tc	0.4385965	0.3074154	0.5763786	4
Ic, Ea, Ne, Tc	0.4385965	0.3074154	0.5763786	4

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
Tw, Bu/To, Ne, Tc	0.4385965	0.3074154	0.5763786	4
Tw, Ea, Ne, Tc	0.4210526	0.2914341	0.5591614	4
Ax, Ea, Ne, Tc	0.3859649	0.2599547	0.5242516	4
Ea, Ne, An, Tc	0.3859649	0.2599547	0.5242516	4
Bu/To, Ea, Ne, Tc	0.3333333	0.2140053	0.4706476	4

Five way

- we want to pick the combination of sites (1-way, 2-way, 3-way, etc.) that adds a marginal benefit to the prior set of samples to a master table
- we sort the 5-way comparisons to identify the 5-way site comparisons that maximize our estimate of the proportion of positive patients

```
all5 = combn(unique(newdata1$SiteID)[1:10], 5, simplify = TRUE)

 holder <- vector("list", ncol(all5))
  for(i in 1:ncol(all5)) {
        df  = get_sensitivity_interval_notime(df=newdata1, ref_sites = all5[,i])
      holder[[i]] =  df
}

CI.fiveway = do.call("rbind", holder)
colnames(CI.fiveway) = c("Prevalence", "95% CI, Lower", "95% CI, Upper", "N.Site
s")
CI.fiveway = CI.fiveway %>%
  arrange(., desc(`Prevalence`))

kable(CI.fiveway, caption = "Sensitivity Analysis") %>%
  kable_styling("basic", full_width = F, bootstrap_options ="condensed")
```

Sensitivity Analysis

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ic, Fg, Tw, An	0.7894737	0.6611299	0.8862098	5
N, Ax, Ic, Fg, Tw	0.7719298	0.6416366	0.8726029	5
N, Ax, Ic, Fg, An	0.7719298	0.6416366	0.8726029	5
N, Ax, Fg, Tw, An	0.7719298	0.6416366	0.8726029	5
N, Ic, Fg, Tw, Bu/To	0.7543860	0.6223897	0.8587314	5
N, Ic, Fg, Tw, Ea	0.7543860	0.6223897	0.8587314	5

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ic, Fg, Tw, Ne	0.7543860	0.6223897	0.8587314	5
N, Ic, Fg, Tw, Tc	0.7543860	0.6223897	0.8587314	5
N, Ic, Fg, Bu/To, An	0.7543860	0.6223897	0.8587314	5
N, Ic, Fg, Ea, An	0.7543860	0.6223897	0.8587314	5
N, Ic, Fg, Ne, An	0.7543860	0.6223897	0.8587314	5
N, Ic, Fg, An, Tc	0.7543860	0.6223897	0.8587314	5
N, Fg, Tw, Bu/To, An	0.7543860	0.6223897	0.8587314	5
N, Fg, Tw, Ea, An	0.7543860	0.6223897	0.8587314	5
N, Fg, Tw, Ne, An	0.7543860	0.6223897	0.8587314	5
N, Fg, Tw, An, Tc	0.7543860	0.6223897	0.8587314	5
N, Ax, Ic, Tw, An	0.7500000	0.6162798	0.8560669	5
N, Ax, Ic, Fg, Bu/To	0.7368421	0.6033725	0.8446152	5
N, Ax, Fg, Tw, Bu/To	0.7368421	0.6033725	0.8446152	5
N, Ax, Fg, Tw, Ea	0.7368421	0.6033725	0.8446152	5
N, Ax, Fg, Tw, Ne	0.7368421	0.6033725	0.8446152	5
N, Ax, Fg, Tw, Tc	0.7368421	0.6033725	0.8446152	5
Ic, Fg, Tw, Bu/To, An	0.7368421	0.6033725	0.8446152	5
N, Ax, Ic, Fg, Ea	0.7192982	0.5845711	0.8302708	5
N, Ax, Ic, Fg, Ne	0.7192982	0.5845711	0.8302708	5
N, Ax, Ic, Fg, Tc	0.7192982	0.5845711	0.8302708	5
N, Ax, Fg, Bu/To, An	0.7192982	0.5845711	0.8302708	5
N, Ax, Fg, Ea, An	0.7192982	0.5845711	0.8302708	5
N, Ax, Fg, Ne, An	0.7192982	0.5845711	0.8302708	5
N, Ax, Fg, An, Tc	0.7192982	0.5845711	0.8302708	5
N, Ic, Fg, Bu/To, Ea	0.7192982	0.5845711	0.8302708	5
N, Ic, Fg, Bu/To, Ne	0.7192982	0.5845711	0.8302708	5
N, Ic, Fg, Bu/To, Tc	0.7192982	0.5845711	0.8302708	5
N, Fg, Tw, Bu/To, Ea	0.7192982	0.5845711	0.8302708	5

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Fg, Tw, Bu/To, Ne	0.7192982	0.5845711	0.8302708	5
N, Fg, Tw, Bu/To, Tc	0.7192982	0.5845711	0.8302708	5
N, Fg, Tw, Ea, Ne	0.7192982	0.5845711	0.8302708	5
N, Fg, Tw, Ea, Tc	0.7192982	0.5845711	0.8302708	5
N, Fg, Tw, Ne, Tc	0.7192982	0.5845711	0.8302708	5
Ax, Ic, Fg, Tw, An	0.7192982	0.5845711	0.8302708	5
Ax, Ic, Fg, Bu/To, An	0.7192982	0.5845711	0.8302708	5
Ax, Fg, Tw, Bu/To, An	0.7192982	0.5845711	0.8302708	5
Ic, Fg, Tw, Ne, An	0.7192982	0.5845711	0.8302708	5
Ic, Fg, Tw, An, Tc	0.7192982	0.5845711	0.8302708	5
N, Ax, Ic, Tw, Bu/To	0.7142857	0.5779013	0.8270482	5
N, Ax, Ic, Tw, Ea	0.7142857	0.5779013	0.8270482	5
N, Ax, Ic, Tw, Tc	0.7142857	0.5779013	0.8270482	5
N, Ax, Ic, Tw, Ne	0.7017544	0.5659736	0.8157119	5
N, Ic, Fg, Ea, Ne	0.7017544	0.5659736	0.8157119	5
N, Ic, Fg, Ea, Tc	0.7017544	0.5659736	0.8157119	5
N, Ic, Fg, Ne, Tc	0.7017544	0.5659736	0.8157119	5
Ax, Fg, Tw, Ne, An	0.7017544	0.5659736	0.8157119	5
Ax, Fg, Tw, An, Tc	0.7017544	0.5659736	0.8157119	5
Ic, Fg, Tw, Bu/To, Ea	0.7017544	0.5659736	0.8157119	5
Ic, Fg, Tw, Ea, An	0.7017544	0.5659736	0.8157119	5
Fg, Tw, Bu/To, Ea, An	0.7017544	0.5659736	0.8157119	5
Fg, Tw, Bu/To, Ne, An	0.7017544	0.5659736	0.8157119	5
Fg, Tw, Bu/To, An, Tc	0.7017544	0.5659736	0.8157119	5
N, Ax, Ic, Bu/To, An	0.6964286	0.5590326	0.8122013	5
N, Ax, Ic, Ea, An	0.6964286	0.5590326	0.8122013	5
N, Ax, Ic, An, Tc	0.6964286	0.5590326	0.8122013	5
N, Ax, Tw, Bu/To, An	0.6964286	0.5590326	0.8122013	5

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Tw, Ea, An	0.6964286	0.5590326	0.8122013	5
N, Ax, Tw, An, Tc	0.6964286	0.5590326	0.8122013	5
Ax, Ic, Tw, Bu/To, An	0.6964286	0.5590326	0.8122013	5
N, Ax, Ic, Ne, An	0.6842105	0.5475702	0.8009499	5
N, Ax, Tw, Ne, An	0.6842105	0.5475702	0.8009499	5
N, Ic, Tw, Ne, An	0.6842105	0.5475702	0.8009499	5
N, Fg, Bu/To, Ea, An	0.6842105	0.5475702	0.8009499	5
N, Fg, Bu/To, Ne, An	0.6842105	0.5475702	0.8009499	5
N, Fg, Bu/To, An, Tc	0.6842105	0.5475702	0.8009499	5
N, Fg, Ea, Ne, An	0.6842105	0.5475702	0.8009499	5
N, Fg, Ea, An, Tc	0.6842105	0.5475702	0.8009499	5
N, Fg, Ne, An, Tc	0.6842105	0.5475702	0.8009499	5
Ax, Ic, Fg, Tw, Bu/To	0.6842105	0.5475702	0.8009499	5
Ax, Ic, Fg, Tw, Ea	0.6842105	0.5475702	0.8009499	5
Ax, Ic, Fg, Bu/To, Ea	0.6842105	0.5475702	0.8009499	5
Ax, Ic, Fg, Ne, An	0.6842105	0.5475702	0.8009499	5
Ax, Ic, Fg, An, Tc	0.6842105	0.5475702	0.8009499	5
Ax, Fg, Tw, Ea, An	0.6842105	0.5475702	0.8009499	5
Ic, Fg, Tw, Ea, Ne	0.6842105	0.5475702	0.8009499	5
Ic, Fg, Tw, Ea, Tc	0.6842105	0.5475702	0.8009499	5
Ic, Fg, Bu/To, Ea, An	0.6842105	0.5475702	0.8009499	5
Ic, Fg, Bu/To, Ne, An	0.6842105	0.5475702	0.8009499	5
Ic, Fg, Bu/To, An, Tc	0.6842105	0.5475702	0.8009499	5
Fg, Tw, Ea, Ne, An	0.6842105	0.5475702	0.8009499	5
Fg, Tw, Ea, An, Tc	0.6842105	0.5475702	0.8009499	5
Fg, Tw, Ne, An, Tc	0.6842105	0.5475702	0.8009499	5
N, Ic, Tw, Bu/To, An	0.6785714	0.5403638	0.7971455	5
N, Ic, Tw, Ea, An	0.6785714	0.5403638	0.7971455	5

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ic, Tw, An, Tc	0.6785714	0.5403638	0.7971455	5
Ax, Ic, Tw, An, Tc	0.6785714	0.5403638	0.7971455	5
N, Ax, Fg, Bu/To, Ea	0.6666667	0.5293524	0.7859947	5
N, Ax, Fg, Bu/To, Ne	0.6666667	0.5293524	0.7859947	5
N, Ax, Fg, Bu/To, Tc	0.6666667	0.5293524	0.7859947	5
Ax, Ic, Fg, Tw, Ne	0.6666667	0.5293524	0.7859947	5
Ax, Ic, Fg, Tw, Tc	0.6666667	0.5293524	0.7859947	5
Ax, Ic, Fg, Ea, An	0.6666667	0.5293524	0.7859947	5
Ax, Ic, Tw, Ne, An	0.6666667	0.5293524	0.7859947	5
Ax, Fg, Tw, Bu/To, Ea	0.6666667	0.5293524	0.7859947	5
Ax, Fg, Bu/To, Ea, An	0.6666667	0.5293524	0.7859947	5
Ax, Fg, Bu/To, Ne, An	0.6666667	0.5293524	0.7859947	5
Ax, Fg, Bu/To, An, Tc	0.6666667	0.5293524	0.7859947	5
Ic, Fg, Tw, Bu/To, Ne	0.6666667	0.5293524	0.7859947	5
Ic, Fg, Tw, Bu/To, Tc	0.6666667	0.5293524	0.7859947	5
N, Ax, Ic, Bu/To, Ea	0.6607143	0.5218863	0.7818907	5
N, Ax, Ic, Bu/To, Tc	0.6607143	0.5218863	0.7818907	5
N, Ax, Tw, Bu/To, Ea	0.6607143	0.5218863	0.7818907	5
N, Ax, Tw, Bu/To, Tc	0.6607143	0.5218863	0.7818907	5
N, Ax, Tw, Ea, Tc	0.6607143	0.5218863	0.7818907	5
Ax, Ic, Tw, Bu/To, Ea	0.6607143	0.5218863	0.7818907	5
Ax, Ic, Tw, Ea, An	0.6607143	0.5218863	0.7818907	5
N, Ax, Ic, Bu/To, Ne	0.6491228	0.5113134	0.7708544	5
N, Ax, Fg, Ea, Ne	0.6491228	0.5113134	0.7708544	5
N, Ax, Fg, Ea, Tc	0.6491228	0.5113134	0.7708544	5
N, Ax, Fg, Ne, Tc	0.6491228	0.5113134	0.7708544	5
N, Ax, Tw, Bu/To, Ne	0.6491228	0.5113134	0.7708544	5
N, Ax, Tw, Ea, Ne	0.6491228	0.5113134	0.7708544	5

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Tw, Ne, Tc	0.6491228	0.5113134	0.7708544	5
Ax, Ic, Fg, Bu/To, Ne	0.6491228	0.5113134	0.7708544	5
Ax, Ic, Fg, Bu/To, Tc	0.6491228	0.5113134	0.7708544	5
Ax, Fg, Tw, Ea, Ne	0.6491228	0.5113134	0.7708544	5
Ax, Fg, Tw, Ea, Tc	0.6491228	0.5113134	0.7708544	5
Ic, Fg, Tw, Ne, Tc	0.6491228	0.5113134	0.7708544	5
Ic, Fg, Bu/To, Ea, Ne	0.6491228	0.5113134	0.7708544	5
Ic, Fg, Bu/To, Ea, Tc	0.6491228	0.5113134	0.7708544	5
Ic, Fg, Ea, Ne, An	0.6491228	0.5113134	0.7708544	5
Ic, Fg, Ea, An, Tc	0.6491228	0.5113134	0.7708544	5
Ic, Fg, Ne, An, Tc	0.6491228	0.5113134	0.7708544	5
N, Ax, Ic, Ea, Tc	0.6428571	0.5035931	0.7664452	5
Ax, Ic, Tw, Ea, Tc	0.6428571	0.5035931	0.7664452	5
Ax, Ic, Bu/To, Ea, An	0.6428571	0.5035931	0.7664452	5
Ax, Ic, Bu/To, An, Tc	0.6428571	0.5035931	0.7664452	5
Ax, Tw, Bu/To, Ea, An	0.6428571	0.5035931	0.7664452	5
Ax, Tw, Bu/To, An, Tc	0.6428571	0.5035931	0.7664452	5
N, Ax, Ic, Ea, Ne	0.6315789	0.4934470	0.7555360	5
N, Ax, Ic, Ne, Tc	0.6315789	0.4934470	0.7555360	5
N, Ic, Tw, Bu/To, Ne	0.6315789	0.4934470	0.7555360	5
N, Ic, Tw, Ea, Ne	0.6315789	0.4934470	0.7555360	5
N, Ic, Tw, Ne, Tc	0.6315789	0.4934470	0.7555360	5
N, Ic, Bu/To, Ne, An	0.6315789	0.4934470	0.7555360	5
N, Ic, Ea, Ne, An	0.6315789	0.4934470	0.7555360	5
N, Ic, Ne, An, Tc	0.6315789	0.4934470	0.7555360	5
N, Fg, Bu/To, Ea, Ne	0.6315789	0.4934470	0.7555360	5
N, Fg, Bu/To, Ea, Tc	0.6315789	0.4934470	0.7555360	5
N, Fg, Bu/To, Ne, Tc	0.6315789	0.4934470	0.7555360	5

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Tw, Bu/To, Ne, An	0.6315789	0.4934470	0.7555360	5
N, Tw, Ea, Ne, An	0.6315789	0.4934470	0.7555360	5
N, Tw, Ne, An, Tc	0.6315789	0.4934470	0.7555360	5
Ax, Ic, Fg, Ea, Ne	0.6315789	0.4934470	0.7555360	5
Ax, Ic, Fg, Ea, Tc	0.6315789	0.4934470	0.7555360	5
Ax, Ic, Tw, Ea, Ne	0.6315789	0.4934470	0.7555360	5
Ax, Ic, Bu/To, Ne, An	0.6315789	0.4934470	0.7555360	5
Ax, Fg, Tw, Bu/To, Ne	0.6315789	0.4934470	0.7555360	5
Ax, Fg, Tw, Bu/To, Tc	0.6315789	0.4934470	0.7555360	5
Ax, Tw, Bu/To, Ne, An	0.6315789	0.4934470	0.7555360	5
Ic, Tw, Bu/To, Ne, An	0.6315789	0.4934470	0.7555360	5
Fg, Tw, Bu/To, Ea, Ne	0.6315789	0.4934470	0.7555360	5
Fg, Tw, Bu/To, Ea, Tc	0.6315789	0.4934470	0.7555360	5
N, Ax, Bu/To, Ea, An	0.6250000	0.4854783	0.7508159	5
N, Ax, Ea, An, Tc	0.6250000	0.4854783	0.7508159	5
N, Ic, Tw, Bu/To, Ea	0.6250000	0.4854783	0.7508159	5
N, Ic, Tw, Bu/To, Tc	0.6250000	0.4854783	0.7508159	5
N, Ic, Tw, Ea, Tc	0.6250000	0.4854783	0.7508159	5
N, Ic, Bu/To, Ea, An	0.6250000	0.4854783	0.7508159	5
N, Ic, Bu/To, An, Tc	0.6250000	0.4854783	0.7508159	5
N, Ic, Ea, An, Tc	0.6250000	0.4854783	0.7508159	5
N, Tw, Bu/To, Ea, An	0.6250000	0.4854783	0.7508159	5
N, Tw, Bu/To, An, Tc	0.6250000	0.4854783	0.7508159	5
N, Tw, Ea, An, Tc	0.6250000	0.4854783	0.7508159	5
Ax, Ic, Tw, Bu/To, Tc	0.6250000	0.4854783	0.7508159	5
Ic, Tw, Bu/To, Ea, An	0.6250000	0.4854783	0.7508159	5
Ic, Tw, Bu/To, An, Tc	0.6250000	0.4854783	0.7508159	5
N, Ax, Bu/To, Ne, An	0.6140351	0.4757484	0.7400453	5

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Ea, Ne, An	0.6140351	0.4757484	0.7400453	5
N, Ax, Ne, An, Tc	0.6140351	0.4757484	0.7400453	5
N, Fg, Ea, Ne, Tc	0.6140351	0.4757484	0.7400453	5
Ax, Ic, Tw, Bu/To, Ne	0.6140351	0.4757484	0.7400453	5
Ax, Fg, Tw, Ne, Tc	0.6140351	0.4757484	0.7400453	5
Ax, Fg, Ea, Ne, An	0.6140351	0.4757484	0.7400453	5
Ax, Fg, Ea, An, Tc	0.6140351	0.4757484	0.7400453	5
Ax, Fg, Ne, An, Tc	0.6140351	0.4757484	0.7400453	5
Ic, Fg, Bu/To, Ne, Tc	0.6140351	0.4757484	0.7400453	5
Ic, Tw, Ea, Ne, An	0.6140351	0.4757484	0.7400453	5
Ic, Tw, Ne, An, Tc	0.6140351	0.4757484	0.7400453	5
Fg, Tw, Ea, Ne, Tc	0.6140351	0.4757484	0.7400453	5
Fg, Bu/To, Ea, Ne, An	0.6140351	0.4757484	0.7400453	5
Fg, Bu/To, Ea, An, Tc	0.6140351	0.4757484	0.7400453	5
Fg, Bu/To, Ne, An, Tc	0.6140351	0.4757484	0.7400453	5
N, Ax, Bu/To, An, Tc	0.6071429	0.4675369	0.7350087	5
Ax, Ic, Bu/To, Ea, Tc	0.6071429	0.4675369	0.7350087	5
Ax, Ic, Ea, An, Tc	0.6071429	0.4675369	0.7350087	5
Ax, Tw, Ea, An, Tc	0.6071429	0.4675369	0.7350087	5
Ic, Tw, Ea, An, Tc	0.6071429	0.4675369	0.7350087	5
Ax, Ic, Fg, Ne, Tc	0.5964912	0.4582134	0.7243873	5
Ax, Ic, Tw, Ne, Tc	0.5964912	0.4582134	0.7243873	5
Ax, Ic, Bu/To, Ea, Ne	0.5964912	0.4582134	0.7243873	5
Ax, Ic, Ea, Ne, An	0.5964912	0.4582134	0.7243873	5
Ax, Ic, Ne, An, Tc	0.5964912	0.4582134	0.7243873	5
Ax, Fg, Bu/To, Ea, Ne	0.5964912	0.4582134	0.7243873	5
Ax, Fg, Bu/To, Ea, Tc	0.5964912	0.4582134	0.7243873	5
Ax, Tw, Ea, Ne, An	0.5964912	0.4582134	0.7243873	5

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
Ax, Tw, Ne, An, Tc	0.5964912	0.4582134	0.7243873	5
Ic, Fg, Ea, Ne, Tc	0.5964912	0.4582134	0.7243873	5
Fg, Tw, Bu/To, Ne, Tc	0.5964912	0.4582134	0.7243873	5
N, Ic, Bu/To, Ea, Ne	0.5789474	0.4408386	0.7085659	5
N, Ic, Bu/To, Ne, Tc	0.5789474	0.4408386	0.7085659	5
N, Tw, Bu/To, Ea, Ne	0.5789474	0.4408386	0.7085659	5
N, Tw, Bu/To, Ne, Tc	0.5789474	0.4408386	0.7085659	5
N, Tw, Ea, Ne, Tc	0.5789474	0.4408386	0.7085659	5
Ic, Tw, Bu/To, Ea, Ne	0.5789474	0.4408386	0.7085659	5
Tw, Bu/To, Ea, Ne, An	0.5789474	0.4408386	0.7085659	5
Tw, Bu/To, Ne, An, Tc	0.5789474	0.4408386	0.7085659	5
N, Ax, Bu/To, Ea, Tc	0.5714286	0.4321590	0.7028791	5
N, Ic, Bu/To, Ea, Tc	0.5714286	0.4321590	0.7028791	5
N, Tw, Bu/To, Ea, Tc	0.5714286	0.4321590	0.7028791	5
Ax, Tw, Bu/To, Ea, Tc	0.5714286	0.4321590	0.7028791	5
Ax, Bu/To, Ea, An, Tc	0.5714286	0.4321590	0.7028791	5
Ic, Tw, Bu/To, Ea, Tc	0.5714286	0.4321590	0.7028791	5
Tw, Bu/To, Ea, An, Tc	0.5714286	0.4321590	0.7028791	5
N, Ax, Bu/To, Ea, Ne	0.5614035	0.4236214	0.6925846	5
N, Ax, Bu/To, Ne, Tc	0.5614035	0.4236214	0.6925846	5
N, Ic, Ea, Ne, Tc	0.5614035	0.4236214	0.6925846	5
Ax, Ic, Bu/To, Ne, Tc	0.5614035	0.4236214	0.6925846	5
Ax, Fg, Bu/To, Ne, Tc	0.5614035	0.4236214	0.6925846	5
Ax, Tw, Bu/To, Ea, Ne	0.5614035	0.4236214	0.6925846	5
Ax, Bu/To, Ea, Ne, An	0.5614035	0.4236214	0.6925846	5
Ax, Bu/To, Ne, An, Tc	0.5614035	0.4236214	0.6925846	5
Ic, Tw, Ea, Ne, Tc	0.5614035	0.4236214	0.6925846	5
Ic, Bu/To, Ea, Ne, An	0.5614035	0.4236214	0.6925846	5

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
Ic, Bu/To, Ne, An, Tc	0.5614035	0.4236214	0.6925846	5
Fg, Ea, Ne, An, Tc	0.5614035	0.4236214	0.6925846	5
Ic, Bu/To, Ea, An, Tc	0.5535714	0.4147167	0.6865639	5
N, Ax, Ea, Ne, Tc	0.5438596	0.4065597	0.6764458	5
N, Bu/To, Ea, Ne, An	0.5438596	0.4065597	0.6764458	5
N, Bu/To, Ne, An, Tc	0.5438596	0.4065597	0.6764458	5
N, Ea, Ne, An, Tc	0.5438596	0.4065597	0.6764458	5
Ax, Ic, Ea, Ne, Tc	0.5438596	0.4065597	0.6764458	5
Ic, Tw, Bu/To, Ne, Tc	0.5438596	0.4065597	0.6764458	5
N, Bu/To, Ea, An, Tc	0.5357143	0.3974360	0.6700853	5
Ax, Fg, Ea, Ne, Tc	0.5263158	0.3896522	0.6601517	5
Ax, Tw, Bu/To, Ne, Tc	0.5263158	0.3896522	0.6601517	5
Ax, Tw, Ea, Ne, Tc	0.5263158	0.3896522	0.6601517	5
Ic, Ea, Ne, An, Tc	0.5263158	0.3896522	0.6601517	5
Fg, Bu/To, Ea, Ne, Tc	0.5263158	0.3896522	0.6601517	5
Tw, Ea, Ne, An, Tc	0.5263158	0.3896522	0.6601517	5
Ic, Bu/To, Ea, Ne, Tc	0.5087719	0.3728979	0.6437035	5
Ax, Ea, Ne, An, Tc	0.4912281	0.3562965	0.6271021	5
N, Bu/To, Ea, Ne, Tc	0.4736842	0.3398483	0.6103478	5
Ax, Bu/To, Ea, Ne, Tc	0.4736842	0.3398483	0.6103478	5
Tw, Bu/To, Ea, Ne, Tc	0.4736842	0.3398483	0.6103478	5
Bu/To, Ea, Ne, An, Tc	0.4736842	0.3398483	0.6103478	5

Six way

- we want to pick the combination of sites (1-way, 2-way, 3-way, etc.) that adds a marginal benefit to the prior set of samples to a master table
- we sort the 6-way comparisons to identify the 6-way site comparisons that maximize our estimate of the proportion of positive patients
- 6 sites were required to achieve 100% sensitivity, identifying all patients colonized at any body site

```
all6 = combn(unique(newdata1$SiteID)[1:10], 6, simplify = TRUE)

 holder <- vector("list", ncol(all6))
  for(i in 1:ncol(all6)) {
        df  = get_sensitivity_interval_notime(df=newdata1, ref_sites = all6[,i])
      holder[[i]] =  df
}

CI.sixway = do.call("rbind", holder)
colnames(CI.sixway) = c("Prevalence", "95% CI, Lower", "95% CI, Upper", "N.Site
s")
CI.sixway = CI.sixway %>%
  arrange(., desc(`Prevalence`))

kable(CI.sixway, caption = "Sensitivity Analysis") %>%
  kable_styling("basic", full_width = F, bootstrap_options ="condensed")
```

Sensitivity Analysis

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Ic, Fg, Tw, An	0.8070175	0.6808902	0.8995278	6
N, Ic, Fg, Tw, Bu/To, An	0.7894737	0.6611299	0.8862098	6
N, Ic, Fg, Tw, Ea, An	0.7894737	0.6611299	0.8862098	6
N, Ic, Fg, Tw, Ne, An	0.7894737	0.6611299	0.8862098	6
N, Ic, Fg, Tw, An, Tc	0.7894737	0.6611299	0.8862098	6
N, Ax, Ic, Fg, Tw, Bu/To	0.7719298	0.6416366	0.8726029	6
N, Ax, Ic, Fg, Tw, Ea	0.7719298	0.6416366	0.8726029	6
N, Ax, Ic, Fg, Tw, Ne	0.7719298	0.6416366	0.8726029	6
N, Ax, Ic, Fg, Tw, Tc	0.7719298	0.6416366	0.8726029	6
N, Ax, Ic, Fg, Bu/To, An	0.7719298	0.6416366	0.8726029	6
N, Ax, Ic, Fg, Ea, An	0.7719298	0.6416366	0.8726029	6
N, Ax, Ic, Fg, Ne, An	0.7719298	0.6416366	0.8726029	6
N, Ax, Ic, Fg, An, Tc	0.7719298	0.6416366	0.8726029	6
N, Ax, Fg, Tw, Bu/To, An	0.7719298	0.6416366	0.8726029	6
N, Ax, Fg, Tw, Ea, An	0.7719298	0.6416366	0.8726029	6
N, Ax, Fg, Tw, Ne, An	0.7719298	0.6416366	0.8726029	6

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Fg, Tw, An, Tc	0.7719298	0.6416366	0.8726029	6
N, Ic, Fg, Tw, Bu/To, Ea	0.7543860	0.6223897	0.8587314	6
N, Ic, Fg, Tw, Bu/To, Ne	0.7543860	0.6223897	0.8587314	6
N, Ic, Fg, Tw, Bu/To, Tc	0.7543860	0.6223897	0.8587314	6
N, Ic, Fg, Tw, Ea, Ne	0.7543860	0.6223897	0.8587314	6
N, Ic, Fg, Tw, Ea, Tc	0.7543860	0.6223897	0.8587314	6
N, Ic, Fg, Tw, Ne, Tc	0.7543860	0.6223897	0.8587314	6
N, Ic, Fg, Bu/To, Ea, An	0.7543860	0.6223897	0.8587314	6
N, Ic, Fg, Bu/To, Ne, An	0.7543860	0.6223897	0.8587314	6
N, Ic, Fg, Bu/To, An, Tc	0.7543860	0.6223897	0.8587314	6
N, Ic, Fg, Ea, Ne, An	0.7543860	0.6223897	0.8587314	6
N, Ic, Fg, Ea, An, Tc	0.7543860	0.6223897	0.8587314	6
N, Ic, Fg, Ne, An, Tc	0.7543860	0.6223897	0.8587314	6
N, Fg, Tw, Bu/To, Ea, An	0.7543860	0.6223897	0.8587314	6
N, Fg, Tw, Bu/To, Ne, An	0.7543860	0.6223897	0.8587314	6
N, Fg, Tw, Bu/To, An, Tc	0.7543860	0.6223897	0.8587314	6
N, Fg, Tw, Ea, Ne, An	0.7543860	0.6223897	0.8587314	6
N, Fg, Tw, Ea, An, Tc	0.7543860	0.6223897	0.8587314	6
N, Fg, Tw, Ne, An, Tc	0.7543860	0.6223897	0.8587314	6
Ax, Ic, Fg, Tw, Bu/To, An	0.7543860	0.6223897	0.8587314	6
N, Ax, Ic, Tw, Bu/To, An	0.7500000	0.6162798	0.8560669	6
N, Ax, Ic, Tw, Ea, An	0.7500000	0.6162798	0.8560669	6
N, Ax, Ic, Tw, An, Tc	0.7500000	0.6162798	0.8560669	6
N, Ax, Ic, Fg, Bu/To, Ea	0.7368421	0.6033725	0.8446152	6
N, Ax, Ic, Fg, Bu/To, Ne	0.7368421	0.6033725	0.8446152	6
N, Ax, Ic, Fg, Bu/To, Tc	0.7368421	0.6033725	0.8446152	6
N, Ax, Ic, Tw, Ne, An	0.7368421	0.6033725	0.8446152	6
N, Ax, Fg, Tw, Bu/To, Ea	0.7368421	0.6033725	0.8446152	6

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Fg, Tw, Bu/To, Ne	0.7368421	0.6033725	0.8446152	6
N, Ax, Fg, Tw, Bu/To, Tc	0.7368421	0.6033725	0.8446152	6
N, Ax, Fg, Tw, Ea, Ne	0.7368421	0.6033725	0.8446152	6
N, Ax, Fg, Tw, Ea, Tc	0.7368421	0.6033725	0.8446152	6
N, Ax, Fg, Tw, Ne, Tc	0.7368421	0.6033725	0.8446152	6
Ax, Ic, Fg, Tw, Ne, An	0.7368421	0.6033725	0.8446152	6
Ax, Ic, Fg, Tw, An, Tc	0.7368421	0.6033725	0.8446152	6
Ic, Fg, Tw, Bu/To, Ea, An	0.7368421	0.6033725	0.8446152	6
Ic, Fg, Tw, Bu/To, Ne, An	0.7368421	0.6033725	0.8446152	6
Ic, Fg, Tw, Bu/To, An, Tc	0.7368421	0.6033725	0.8446152	6
N, Ax, Ic, Fg, Ea, Ne	0.7192982	0.5845711	0.8302708	6
N, Ax, Ic, Fg, Ea, Tc	0.7192982	0.5845711	0.8302708	6
N, Ax, Ic, Fg, Ne, Tc	0.7192982	0.5845711	0.8302708	6
N, Ax, Fg, Bu/To, Ea, An	0.7192982	0.5845711	0.8302708	6
N, Ax, Fg, Bu/To, Ne, An	0.7192982	0.5845711	0.8302708	6
N, Ax, Fg, Bu/To, An, Tc	0.7192982	0.5845711	0.8302708	6
N, Ax, Fg, Ea, Ne, An	0.7192982	0.5845711	0.8302708	6
N, Ax, Fg, Ea, An, Tc	0.7192982	0.5845711	0.8302708	6
N, Ax, Fg, Ne, An, Tc	0.7192982	0.5845711	0.8302708	6
N, Ic, Fg, Bu/To, Ea, Ne	0.7192982	0.5845711	0.8302708	6
N, Ic, Fg, Bu/To, Ea, Tc	0.7192982	0.5845711	0.8302708	6
N, Ic, Fg, Bu/To, Ne, Tc	0.7192982	0.5845711	0.8302708	6
N, Fg, Tw, Bu/To, Ea, Ne	0.7192982	0.5845711	0.8302708	6
N, Fg, Tw, Bu/To, Ea, Tc	0.7192982	0.5845711	0.8302708	6
N, Fg, Tw, Bu/To, Ne, Tc	0.7192982	0.5845711	0.8302708	6
N, Fg, Tw, Ea, Ne, Tc	0.7192982	0.5845711	0.8302708	6
Ax, Ic, Fg, Tw, Bu/To, Ea	0.7192982	0.5845711	0.8302708	6
Ax, Ic, Fg, Tw, Ea, An	0.7192982	0.5845711	0.8302708	6

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
Ax, Ic, Fg, Bu/To, Ea, An	0.7192982	0.5845711	0.8302708	6
Ax, Ic, Fg, Bu/To, Ne, An	0.7192982	0.5845711	0.8302708	6
Ax, Ic, Fg, Bu/To, An, Tc	0.7192982	0.5845711	0.8302708	6
Ax, Fg, Tw, Bu/To, Ea, An	0.7192982	0.5845711	0.8302708	6
Ax, Fg, Tw, Bu/To, Ne, An	0.7192982	0.5845711	0.8302708	6
Ax, Fg, Tw, Bu/To, An, Tc	0.7192982	0.5845711	0.8302708	6
Ic, Fg, Tw, Ea, Ne, An	0.7192982	0.5845711	0.8302708	6
Ic, Fg, Tw, Ea, An, Tc	0.7192982	0.5845711	0.8302708	6
Ic, Fg, Tw, Ne, An, Tc	0.7192982	0.5845711	0.8302708	6
N, Ax, Ic, Tw, Bu/To, Ea	0.7142857	0.5779013	0.8270482	6
N, Ax, Ic, Tw, Bu/To, Tc	0.7142857	0.5779013	0.8270482	6
N, Ax, Ic, Tw, Ea, Tc	0.7142857	0.5779013	0.8270482	6
N, Ax, Ic, Tw, Bu/To, Ne	0.7017544	0.5659736	0.8157119	6
N, Ax, Ic, Tw, Ea, Ne	0.7017544	0.5659736	0.8157119	6
N, Ax, Ic, Tw, Ne, Tc	0.7017544	0.5659736	0.8157119	6
N, Ic, Fg, Ea, Ne, Tc	0.7017544	0.5659736	0.8157119	6
Ax, Ic, Fg, Tw, Ea, Ne	0.7017544	0.5659736	0.8157119	6
Ax, Ic, Fg, Tw, Ea, Tc	0.7017544	0.5659736	0.8157119	6
Ax, Fg, Tw, Ea, Ne, An	0.7017544	0.5659736	0.8157119	6
Ax, Fg, Tw, Ea, An, Tc	0.7017544	0.5659736	0.8157119	6
Ax, Fg, Tw, Ne, An, Tc	0.7017544	0.5659736	0.8157119	6
Ic, Fg, Tw, Bu/To, Ea, Ne	0.7017544	0.5659736	0.8157119	6
Ic, Fg, Tw, Bu/To, Ea, Tc	0.7017544	0.5659736	0.8157119	6
Fg, Tw, Bu/To, Ea, Ne, An	0.7017544	0.5659736	0.8157119	6
Fg, Tw, Bu/To, Ea, An, Tc	0.7017544	0.5659736	0.8157119	6
Fg, Tw, Bu/To, Ne, An, Tc	0.7017544	0.5659736	0.8157119	6
N, Ax, Ic, Bu/To, Ea, An	0.6964286	0.5590326	0.8122013	6
N, Ax, Ic, Bu/To, An, Tc	0.6964286	0.5590326	0.8122013	6

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Ic, Ea, An, Tc	0.6964286	0.5590326	0.8122013	6
N, Ax, Tw, Bu/To, Ea, An	0.6964286	0.5590326	0.8122013	6
N, Ax, Tw, Bu/To, An, Tc	0.6964286	0.5590326	0.8122013	6
N, Ax, Tw, Ea, An, Tc	0.6964286	0.5590326	0.8122013	6
Ax, Ic, Tw, Bu/To, Ea, An	0.6964286	0.5590326	0.8122013	6
Ax, Ic, Tw, Bu/To, An, Tc	0.6964286	0.5590326	0.8122013	6
N, Ax, Ic, Bu/To, Ne, An	0.6842105	0.5475702	0.8009499	6
N, Ax, Ic, Ea, Ne, An	0.6842105	0.5475702	0.8009499	6
N, Ax, Ic, Ne, An, Tc	0.6842105	0.5475702	0.8009499	6
N, Ax, Tw, Bu/To, Ne, An	0.6842105	0.5475702	0.8009499	6
N, Ax, Tw, Ea, Ne, An	0.6842105	0.5475702	0.8009499	6
N, Ax, Tw, Ne, An, Tc	0.6842105	0.5475702	0.8009499	6
N, Ic, Tw, Bu/To, Ne, An	0.6842105	0.5475702	0.8009499	6
N, Ic, Tw, Ea, Ne, An	0.6842105	0.5475702	0.8009499	6
N, Ic, Tw, Ne, An, Tc	0.6842105	0.5475702	0.8009499	6
N, Fg, Bu/To, Ea, Ne, An	0.6842105	0.5475702	0.8009499	6
N, Fg, Bu/To, Ea, An, Tc	0.6842105	0.5475702	0.8009499	6
N, Fg, Bu/To, Ne, An, Tc	0.6842105	0.5475702	0.8009499	6
N, Fg, Ea, Ne, An, Tc	0.6842105	0.5475702	0.8009499	6
Ax, Ic, Fg, Tw, Bu/To, Ne	0.6842105	0.5475702	0.8009499	6
Ax, Ic, Fg, Tw, Bu/To, Tc	0.6842105	0.5475702	0.8009499	6
Ax, Ic, Fg, Bu/To, Ea, Ne	0.6842105	0.5475702	0.8009499	6
Ax, Ic, Fg, Bu/To, Ea, Tc	0.6842105	0.5475702	0.8009499	6
Ax, Ic, Fg, Ea, Ne, An	0.6842105	0.5475702	0.8009499	6
Ax, Ic, Fg, Ea, An, Tc	0.6842105	0.5475702	0.8009499	6
Ax, Ic, Fg, Ne, An, Tc	0.6842105	0.5475702	0.8009499	6
Ax, Ic, Tw, Bu/To, Ne, An	0.6842105	0.5475702	0.8009499	6
Ic, Fg, Tw, Ea, Ne, Tc	0.6842105	0.5475702	0.8009499	6

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
Ic, Fg, Bu/To, Ea, Ne, An	0.6842105	0.5475702	0.8009499	6
Ic, Fg, Bu/To, Ea, An, Tc	0.6842105	0.5475702	0.8009499	6
Ic, Fg, Bu/To, Ne, An, Tc	0.6842105	0.5475702	0.8009499	6
Fg, Tw, Ea, Ne, An, Tc	0.6842105	0.5475702	0.8009499	6
N, Ic, Tw, Bu/To, Ea, An	0.6785714	0.5403638	0.7971455	6
N, Ic, Tw, Bu/To, An, Tc	0.6785714	0.5403638	0.7971455	6
N, Ic, Tw, Ea, An, Tc	0.6785714	0.5403638	0.7971455	6
Ax, Ic, Tw, Ea, An, Tc	0.6785714	0.5403638	0.7971455	6
N, Ax, Fg, Bu/To, Ea, Ne	0.6666667	0.5293524	0.7859947	6
N, Ax, Fg, Bu/To, Ea, Tc	0.6666667	0.5293524	0.7859947	6
N, Ax, Fg, Bu/To, Ne, Tc	0.6666667	0.5293524	0.7859947	6
Ax, Ic, Fg, Tw, Ne, Tc	0.6666667	0.5293524	0.7859947	6
Ax, Ic, Tw, Ea, Ne, An	0.6666667	0.5293524	0.7859947	6
Ax, Ic, Tw, Ne, An, Tc	0.6666667	0.5293524	0.7859947	6
Ax, Fg, Tw, Bu/To, Ea, Ne	0.6666667	0.5293524	0.7859947	6
Ax, Fg, Tw, Bu/To, Ea, Tc	0.6666667	0.5293524	0.7859947	6
Ax, Fg, Bu/To, Ea, Ne, An	0.6666667	0.5293524	0.7859947	6
Ax, Fg, Bu/To, Ea, An, Tc	0.6666667	0.5293524	0.7859947	6
Ax, Fg, Bu/To, Ne, An, Tc	0.6666667	0.5293524	0.7859947	6
Ic, Fg, Tw, Bu/To, Ne, Tc	0.6666667	0.5293524	0.7859947	6
N, Ax, Ic, Bu/To, Ea, Tc	0.6607143	0.5218863	0.7818907	6
N, Ax, Tw, Bu/To, Ea, Tc	0.6607143	0.5218863	0.7818907	6
Ax, Ic, Tw, Bu/To, Ea, Tc	0.6607143	0.5218863	0.7818907	6
N, Ax, Ic, Bu/To, Ea, Ne	0.6491228	0.5113134	0.7708544	6
N, Ax, Ic, Bu/To, Ne, Tc	0.6491228	0.5113134	0.7708544	6
N, Ax, Fg, Ea, Ne, Tc	0.6491228	0.5113134	0.7708544	6
N, Ax, Tw, Bu/To, Ea, Ne	0.6491228	0.5113134	0.7708544	6
N, Ax, Tw, Bu/To, Ne, Tc	0.6491228	0.5113134	0.7708544	6

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Tw, Ea, Ne, Tc	0.6491228	0.5113134	0.7708544	6
Ax, Ic, Fg, Bu/To, Ne, Tc	0.6491228	0.5113134	0.7708544	6
Ax, Ic, Tw, Bu/To, Ea, Ne	0.6491228	0.5113134	0.7708544	6
Ax, Fg, Tw, Ea, Ne, Tc	0.6491228	0.5113134	0.7708544	6
Ic, Fg, Bu/To, Ea, Ne, Tc	0.6491228	0.5113134	0.7708544	6
Ic, Fg, Ea, Ne, An, Tc	0.6491228	0.5113134	0.7708544	6
Ax, Ic, Bu/To, Ea, An, Tc	0.6428571	0.5035931	0.7664452	6
Ax, Tw, Bu/To, Ea, An, Tc	0.6428571	0.5035931	0.7664452	6
N, Ax, Ic, Ea, Ne, Tc	0.6315789	0.4934470	0.7555360	6
N, Ic, Tw, Bu/To, Ea, Ne	0.6315789	0.4934470	0.7555360	6
N, Ic, Tw, Bu/To, Ne, Tc	0.6315789	0.4934470	0.7555360	6
N, Ic, Tw, Ea, Ne, Tc	0.6315789	0.4934470	0.7555360	6
N, Ic, Bu/To, Ea, Ne, An	0.6315789	0.4934470	0.7555360	6
N, Ic, Bu/To, Ne, An, Tc	0.6315789	0.4934470	0.7555360	6
N, Ic, Ea, Ne, An, Tc	0.6315789	0.4934470	0.7555360	6
N, Fg, Bu/To, Ea, Ne, Tc	0.6315789	0.4934470	0.7555360	6
N, Tw, Bu/To, Ea, Ne, An	0.6315789	0.4934470	0.7555360	6
N, Tw, Bu/To, Ne, An, Tc	0.6315789	0.4934470	0.7555360	6
N, Tw, Ea, Ne, An, Tc	0.6315789	0.4934470	0.7555360	6
Ax, Ic, Fg, Ea, Ne, Tc	0.6315789	0.4934470	0.7555360	6
Ax, Ic, Tw, Ea, Ne, Tc	0.6315789	0.4934470	0.7555360	6
Ax, Ic, Bu/To, Ea, Ne, An	0.6315789	0.4934470	0.7555360	6
Ax, Ic, Bu/To, Ne, An, Tc	0.6315789	0.4934470	0.7555360	6
Ax, Fg, Tw, Bu/To, Ne, Tc	0.6315789	0.4934470	0.7555360	6
Ax, Tw, Bu/To, Ea, Ne, An	0.6315789	0.4934470	0.7555360	6
Ax, Tw, Bu/To, Ne, An, Tc	0.6315789	0.4934470	0.7555360	6
Ic, Tw, Bu/To, Ea, Ne, An	0.6315789	0.4934470	0.7555360	6
Ic, Tw, Bu/To, Ne, An, Tc	0.6315789	0.4934470	0.7555360	6

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
Fg, Tw, Bu/To, Ea, Ne, Tc	0.6315789	0.4934470	0.7555360	6
N, Ax, Bu/To, Ea, An, Tc	0.6250000	0.4854783	0.7508159	6
N, Ic, Tw, Bu/To, Ea, Tc	0.6250000	0.4854783	0.7508159	6
N, Ic, Bu/To, Ea, An, Tc	0.6250000	0.4854783	0.7508159	6
N, Tw, Bu/To, Ea, An, Tc	0.6250000	0.4854783	0.7508159	6
Ic, Tw, Bu/To, Ea, An, Tc	0.6250000	0.4854783	0.7508159	6
N, Ax, Bu/To, Ea, Ne, An	0.6140351	0.4757484	0.7400453	6
N, Ax, Bu/To, Ne, An, Tc	0.6140351	0.4757484	0.7400453	6
N, Ax, Ea, Ne, An, Tc	0.6140351	0.4757484	0.7400453	6
Ax, Ic, Tw, Bu/To, Ne, Tc	0.6140351	0.4757484	0.7400453	6
Ax, Fg, Ea, Ne, An, Tc	0.6140351	0.4757484	0.7400453	6
Ic, Tw, Ea, Ne, An, Tc	0.6140351	0.4757484	0.7400453	6
Fg, Bu/To, Ea, Ne, An, Tc	0.6140351	0.4757484	0.7400453	6
Ax, Ic, Bu/To, Ea, Ne, Tc	0.5964912	0.4582134	0.7243873	6
Ax, Ic, Ea, Ne, An, Tc	0.5964912	0.4582134	0.7243873	6
Ax, Fg, Bu/To, Ea, Ne, Tc	0.5964912	0.4582134	0.7243873	6
Ax, Tw, Ea, Ne, An, Tc	0.5964912	0.4582134	0.7243873	6
N, Ic, Bu/To, Ea, Ne, Tc	0.5789474	0.4408386	0.7085659	6
N, Tw, Bu/To, Ea, Ne, Tc	0.5789474	0.4408386	0.7085659	6
Ic, Tw, Bu/To, Ea, Ne, Tc	0.5789474	0.4408386	0.7085659	6
Tw, Bu/To, Ea, Ne, An, Tc	0.5789474	0.4408386	0.7085659	6
N, Ax, Bu/To, Ea, Ne, Tc	0.5614035	0.4236214	0.6925846	6
Ax, Tw, Bu/To, Ea, Ne, Tc	0.5614035	0.4236214	0.6925846	6
Ax, Bu/To, Ea, Ne, An, Tc	0.5614035	0.4236214	0.6925846	6
Ic, Bu/To, Ea, Ne, An, Tc	0.5614035	0.4236214	0.6925846	6
N, Bu/To, Ea, Ne, An, Tc	0.5438596	0.4065597	0.6764458	6

Seven way

- we want to pick the combination of sites (1-way, 2-way, 3-way, etc.) that adds a marginal benefit to

the prior set of samples to a master table

- we sort the 6-way comparisons to identify the 6-way site comparisons that maximize our estimate of the proportion of positive patients
- 6 sites were required to achieve 100% sensitivity, identifying all patients colonized at any body site
- 7-way is shown for completeness

```
all7 = combn(unique(newdata1$SiteID)[1:10], 7, simplify = TRUE)

 holder <- vector("list", ncol(all7))
  for(i in 1:ncol(all7)) {
        df  = get_sensitivity_interval_notime(df=newdata1, ref_sites = all7[,i])
      holder[[i]] =  df
}

CI.sevenway = do.call("rbind", holder)
colnames(CI.sevenway) = c("Prevalence", "95% CI, Lower", "95% CI, Upper", "N.Site
s")
CI.sevenway = CI.sevenway %>%
  arrange(., desc(`Prevalence`))

kable(CI.sevenway, caption = "Sensitivity Analysis") %>%
  kable_styling("basic", full_width = F, bootstrap_options ="condensed")
```

Sensitivity Analysis

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Ic, Fg, Tw, Bu/To, An	0.8070175	0.6808902	0.8995278	7
N, Ax, Ic, Fg, Tw, Ea, An	0.8070175	0.6808902	0.8995278	7
N, Ax, Ic, Fg, Tw, Ne, An	0.8070175	0.6808902	0.8995278	7
N, Ax, Ic, Fg, Tw, An, Tc	0.8070175	0.6808902	0.8995278	7
N, Ic, Fg, Tw, Bu/To, Ea, An	0.7894737	0.6611299	0.8862098	7
N, Ic, Fg, Tw, Bu/To, Ne, An	0.7894737	0.6611299	0.8862098	7
N, Ic, Fg, Tw, Bu/To, An, Tc	0.7894737	0.6611299	0.8862098	7
N, Ic, Fg, Tw, Ea, Ne, An	0.7894737	0.6611299	0.8862098	7
N, Ic, Fg, Tw, Ea, An, Tc	0.7894737	0.6611299	0.8862098	7
N, Ic, Fg, Tw, Ne, An, Tc	0.7894737	0.6611299	0.8862098	7
N, Ax, Ic, Fg, Tw, Bu/To, Ea	0.7719298	0.6416366	0.8726029	7
N, Ax, Ic, Fg, Tw, Bu/To, Ne	0.7719298	0.6416366	0.8726029	7
N, Ax, Ic, Fg, Tw, Bu/To, Tc	0.7719298	0.6416366	0.8726029	7

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Ic, Fg, Tw, Ea, Ne	0.7719298	0.6416366	0.8726029	7
N, Ax, Ic, Fg, Tw, Ea, Tc	0.7719298	0.6416366	0.8726029	7
N, Ax, Ic, Fg, Tw, Ne, Tc	0.7719298	0.6416366	0.8726029	7
N, Ax, Ic, Fg, Bu/To, Ea, An	0.7719298	0.6416366	0.8726029	7
N, Ax, Ic, Fg, Bu/To, Ne, An	0.7719298	0.6416366	0.8726029	7
N, Ax, Ic, Fg, Bu/To, An, Tc	0.7719298	0.6416366	0.8726029	7
N, Ax, Ic, Fg, Ea, Ne, An	0.7719298	0.6416366	0.8726029	7
N, Ax, Ic, Fg, Ea, An, Tc	0.7719298	0.6416366	0.8726029	7
N, Ax, Ic, Fg, Ne, An, Tc	0.7719298	0.6416366	0.8726029	7
N, Ax, Fg, Tw, Bu/To, Ea, An	0.7719298	0.6416366	0.8726029	7
N, Ax, Fg, Tw, Bu/To, Ne, An	0.7719298	0.6416366	0.8726029	7
N, Ax, Fg, Tw, Bu/To, An, Tc	0.7719298	0.6416366	0.8726029	7
N, Ax, Fg, Tw, Ea, Ne, An	0.7719298	0.6416366	0.8726029	7
N, Ax, Fg, Tw, Ea, An, Tc	0.7719298	0.6416366	0.8726029	7
N, Ax, Fg, Tw, Ne, An, Tc	0.7719298	0.6416366	0.8726029	7
N, Ic, Fg, Tw, Bu/To, Ea, Ne	0.7543860	0.6223897	0.8587314	7
N, Ic, Fg, Tw, Bu/To, Ea, Tc	0.7543860	0.6223897	0.8587314	7
N, Ic, Fg, Tw, Bu/To, Ne, Tc	0.7543860	0.6223897	0.8587314	7
N, Ic, Fg, Tw, Ea, Ne, Tc	0.7543860	0.6223897	0.8587314	7
N, Ic, Fg, Bu/To, Ea, Ne, An	0.7543860	0.6223897	0.8587314	7
N, Ic, Fg, Bu/To, Ea, An, Tc	0.7543860	0.6223897	0.8587314	7
N, Ic, Fg, Bu/To, Ne, An, Tc	0.7543860	0.6223897	0.8587314	7
N, Ic, Fg, Ea, Ne, An, Tc	0.7543860	0.6223897	0.8587314	7
N, Fg, Tw, Bu/To, Ea, Ne, An	0.7543860	0.6223897	0.8587314	7
N, Fg, Tw, Bu/To, Ea, An, Tc	0.7543860	0.6223897	0.8587314	7
N, Fg, Tw, Bu/To, Ne, An, Tc	0.7543860	0.6223897	0.8587314	7
N, Fg, Tw, Ea, Ne, An, Tc	0.7543860	0.6223897	0.8587314	7
Ax, Ic, Fg, Tw, Bu/To, Ea, An	0.7543860	0.6223897	0.8587314	7

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
Ax, Ic, Fg, Tw, Bu/To, Ne, An	0.7543860	0.6223897	0.8587314	7
Ax, Ic, Fg, Tw, Bu/To, An, Tc	0.7543860	0.6223897	0.8587314	7
N, Ax, Ic, Tw, Bu/To, Ea, An	0.7500000	0.6162798	0.8560669	7
N, Ax, Ic, Tw, Bu/To, An, Tc	0.7500000	0.6162798	0.8560669	7
N, Ax, Ic, Tw, Ea, An, Tc	0.7500000	0.6162798	0.8560669	7
N, Ax, Ic, Fg, Bu/To, Ea, Ne	0.7368421	0.6033725	0.8446152	7
N, Ax, Ic, Fg, Bu/To, Ea, Tc	0.7368421	0.6033725	0.8446152	7
N, Ax, Ic, Fg, Bu/To, Ne, Tc	0.7368421	0.6033725	0.8446152	7
N, Ax, Ic, Tw, Bu/To, Ne, An	0.7368421	0.6033725	0.8446152	7
N, Ax, Ic, Tw, Ea, Ne, An	0.7368421	0.6033725	0.8446152	7
N, Ax, Ic, Tw, Ne, An, Tc	0.7368421	0.6033725	0.8446152	7
N, Ax, Fg, Tw, Bu/To, Ea, Ne	0.7368421	0.6033725	0.8446152	7
N, Ax, Fg, Tw, Bu/To, Ea, Tc	0.7368421	0.6033725	0.8446152	7
N, Ax, Fg, Tw, Bu/To, Ne, Tc	0.7368421	0.6033725	0.8446152	7
N, Ax, Fg, Tw, Ea, Ne, Tc	0.7368421	0.6033725	0.8446152	7
Ax, Ic, Fg, Tw, Ea, Ne, An	0.7368421	0.6033725	0.8446152	7
Ax, Ic, Fg, Tw, Ea, An, Tc	0.7368421	0.6033725	0.8446152	7
Ax, Ic, Fg, Tw, Ne, An, Tc	0.7368421	0.6033725	0.8446152	7
Ic, Fg, Tw, Bu/To, Ea, Ne, An	0.7368421	0.6033725	0.8446152	7
Ic, Fg, Tw, Bu/To, Ea, An, Tc	0.7368421	0.6033725	0.8446152	7
Ic, Fg, Tw, Bu/To, Ne, An, Tc	0.7368421	0.6033725	0.8446152	7
N, Ax, Ic, Fg, Ea, Ne, Tc	0.7192982	0.5845711	0.8302708	7
N, Ax, Fg, Bu/To, Ea, Ne, An	0.7192982	0.5845711	0.8302708	7
N, Ax, Fg, Bu/To, Ea, An, Tc	0.7192982	0.5845711	0.8302708	7
N, Ax, Fg, Bu/To, Ne, An, Tc	0.7192982	0.5845711	0.8302708	7
N, Ax, Fg, Ea, Ne, An, Tc	0.7192982	0.5845711	0.8302708	7
N, Ic, Fg, Bu/To, Ea, Ne, Tc	0.7192982	0.5845711	0.8302708	7
N, Fg, Tw, Bu/To, Ea, Ne, Tc	0.7192982	0.5845711	0.8302708	7

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
Ax, Ic, Fg, Tw, Bu/To, Ea, Ne	0.7192982	0.5845711	0.8302708	7
Ax, Ic, Fg, Tw, Bu/To, Ea, Tc	0.7192982	0.5845711	0.8302708	7
Ax, Ic, Fg, Bu/To, Ea, Ne, An	0.7192982	0.5845711	0.8302708	7
Ax, Ic, Fg, Bu/To, Ea, An, Tc	0.7192982	0.5845711	0.8302708	7
Ax, Ic, Fg, Bu/To, Ne, An, Tc	0.7192982	0.5845711	0.8302708	7
Ax, Fg, Tw, Bu/To, Ea, Ne, An	0.7192982	0.5845711	0.8302708	7
Ax, Fg, Tw, Bu/To, Ea, An, Tc	0.7192982	0.5845711	0.8302708	7
Ax, Fg, Tw, Bu/To, Ne, An, Tc	0.7192982	0.5845711	0.8302708	7
Ic, Fg, Tw, Ea, Ne, An, Tc	0.7192982	0.5845711	0.8302708	7
N, Ax, Ic, Tw, Bu/To, Ea, Tc	0.7142857	0.5779013	0.8270482	7
N, Ax, Ic, Tw, Bu/To, Ea, Ne	0.7017544	0.5659736	0.8157119	7
N, Ax, Ic, Tw, Bu/To, Ne, Tc	0.7017544	0.5659736	0.8157119	7
N, Ax, Ic, Tw, Ea, Ne, Tc	0.7017544	0.5659736	0.8157119	7
Ax, Ic, Fg, Tw, Ea, Ne, Tc	0.7017544	0.5659736	0.8157119	7
Ax, Fg, Tw, Ea, Ne, An, Tc	0.7017544	0.5659736	0.8157119	7
Ic, Fg, Tw, Bu/To, Ea, Ne, Tc	0.7017544	0.5659736	0.8157119	7
Fg, Tw, Bu/To, Ea, Ne, An, Tc	0.7017544	0.5659736	0.8157119	7
N, Ax, Ic, Bu/To, Ea, An, Tc	0.6964286	0.5590326	0.8122013	7
N, Ax, Tw, Bu/To, Ea, An, Tc	0.6964286	0.5590326	0.8122013	7
Ax, Ic, Tw, Bu/To, Ea, An, Tc	0.6964286	0.5590326	0.8122013	7
N, Ax, Ic, Bu/To, Ea, Ne, An	0.6842105	0.5475702	0.8009499	7
N, Ax, Ic, Bu/To, Ne, An, Tc	0.6842105	0.5475702	0.8009499	7
N, Ax, Ic, Ea, Ne, An, Tc	0.6842105	0.5475702	0.8009499	7
N, Ax, Tw, Bu/To, Ea, Ne, An	0.6842105	0.5475702	0.8009499	7
N, Ax, Tw, Bu/To, Ne, An, Tc	0.6842105	0.5475702	0.8009499	7
N, Ax, Tw, Ea, Ne, An, Tc	0.6842105	0.5475702	0.8009499	7
N, Ic, Tw, Bu/To, Ea, Ne, An	0.6842105	0.5475702	0.8009499	7
N, Ic, Tw, Bu/To, Ne, An, Tc	0.6842105	0.5475702	0.8009499	7

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ic, Tw, Ea, Ne, An, Tc	0.6842105	0.5475702	0.8009499	7
N, Fg, Bu/To, Ea, Ne, An, Tc	0.6842105	0.5475702	0.8009499	7
Ax, Ic, Fg, Tw, Bu/To, Ne, Tc	0.6842105	0.5475702	0.8009499	7
Ax, Ic, Fg, Bu/To, Ea, Ne, Tc	0.6842105	0.5475702	0.8009499	7
Ax, Ic, Fg, Ea, Ne, An, Tc	0.6842105	0.5475702	0.8009499	7
Ax, Ic, Tw, Bu/To, Ea, Ne, An	0.6842105	0.5475702	0.8009499	7
Ax, Ic, Tw, Bu/To, Ne, An, Tc	0.6842105	0.5475702	0.8009499	7
Ic, Fg, Bu/To, Ea, Ne, An, Tc	0.6842105	0.5475702	0.8009499	7
N, Ic, Tw, Bu/To, Ea, An, Tc	0.6785714	0.5403638	0.7971455	7
N, Ax, Fg, Bu/To, Ea, Ne, Tc	0.6666667	0.5293524	0.7859947	7
Ax, Ic, Tw, Ea, Ne, An, Tc	0.6666667	0.5293524	0.7859947	7
Ax, Fg, Tw, Bu/To, Ea, Ne, Tc	0.6666667	0.5293524	0.7859947	7
Ax, Fg, Bu/To, Ea, Ne, An, Tc	0.6666667	0.5293524	0.7859947	7
N, Ax, Ic, Bu/To, Ea, Ne, Tc	0.6491228	0.5113134	0.7708544	7
N, Ax, Tw, Bu/To, Ea, Ne, Tc	0.6491228	0.5113134	0.7708544	7
Ax, Ic, Tw, Bu/To, Ea, Ne, Tc	0.6491228	0.5113134	0.7708544	7
N, Ic, Tw, Bu/To, Ea, Ne, Tc	0.6315789	0.4934470	0.7555360	7
N, Ic, Bu/To, Ea, Ne, An, Tc	0.6315789	0.4934470	0.7555360	7
N, Tw, Bu/To, Ea, Ne, An, Tc	0.6315789	0.4934470	0.7555360	7
Ax, Ic, Bu/To, Ea, Ne, An, Tc	0.6315789	0.4934470	0.7555360	7
Ax, Tw, Bu/To, Ea, Ne, An, Tc	0.6315789	0.4934470	0.7555360	7
Ic, Tw, Bu/To, Ea, Ne, An, Tc	0.6315789	0.4934470	0.7555360	7
N, Ax, Bu/To, Ea, Ne, An, Tc	0.6140351	0.4757484	0.7400453	7

Eight way

- we want to pick the combination of sites (1-way, 2-way, 3-way, etc.) that adds a marginal benefit to the prior set of samples to a master table
- we sort the 6-way comparisons to identify the 6-way site comparisons that maximize our estimate of the proportion of positive patients
- 6 sites were required to achieve 100% sensitivity, identifying all patients colonized at any body site
- 8-way is shown for completeness

```
all8 = combn(unique(newdata1$SiteID)[1:10], 8, simplify = TRUE)

 holder <- vector("list", ncol(all8))
  for(i in 1:ncol(all8)) {
        df  = get_sensitivity_interval_notime(df=newdata1, ref_sites = all8[,i])
      holder[[i]] =  df
}

CI.eightway = do.call("rbind", holder)
colnames(CI.eightway) = c("Prevalence", "95% CI, Lower", "95% CI, Upper", "N.Site
s")
CI.eightway = CI.eightway %>%
  arrange(., desc(`Prevalence`))

kable(CI.eightway, caption = "Sensitivity Analysis") %>%
  kable_styling("basic", full_width = F, bootstrap_options ="condensed")
```

Sensitivity Analysis

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Ic, Fg, Tw, Bu/To, Ea, An	0.8070175	0.6808902	0.8995278	8
N, Ax, Ic, Fg, Tw, Bu/To, Ne, An	0.8070175	0.6808902	0.8995278	8
N, Ax, Ic, Fg, Tw, Bu/To, An, Tc	0.8070175	0.6808902	0.8995278	8
N, Ax, Ic, Fg, Tw, Ea, Ne, An	0.8070175	0.6808902	0.8995278	8
N, Ax, Ic, Fg, Tw, Ea, An, Tc	0.8070175	0.6808902	0.8995278	8
N, Ax, Ic, Fg, Tw, Ne, An, Tc	0.8070175	0.6808902	0.8995278	8
N, Ic, Fg, Tw, Bu/To, Ea, Ne, An	0.7894737	0.6611299	0.8862098	8
N, Ic, Fg, Tw, Bu/To, Ea, An, Tc	0.7894737	0.6611299	0.8862098	8
N, Ic, Fg, Tw, Bu/To, Ne, An, Tc	0.7894737	0.6611299	0.8862098	8
N, Ic, Fg, Tw, Ea, Ne, An, Tc	0.7894737	0.6611299	0.8862098	8
N, Ax, Ic, Fg, Tw, Bu/To, Ea, Ne	0.7719298	0.6416366	0.8726029	8
N, Ax, Ic, Fg, Tw, Bu/To, Ea, Tc	0.7719298	0.6416366	0.8726029	8
N, Ax, Ic, Fg, Tw, Bu/To, Ne, Tc	0.7719298	0.6416366	0.8726029	8
N, Ax, Ic, Fg, Tw, Ea, Ne, Tc	0.7719298	0.6416366	0.8726029	8
N, Ax, Ic, Fg, Bu/To, Ea, Ne, An	0.7719298	0.6416366	0.8726029	8
N, Ax, Ic, Fg, Bu/To, Ea, An, Tc	0.7719298	0.6416366	0.8726029	8

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Ic, Fg, Bu/To, Ne, An, Tc	0.7719298	0.6416366	0.8726029	8
N, Ax, Ic, Fg, Ea, Ne, An, Tc	0.7719298	0.6416366	0.8726029	8
N, Ax, Fg, Tw, Bu/To, Ea, Ne, An	0.7719298	0.6416366	0.8726029	8
N, Ax, Fg, Tw, Bu/To, Ea, An, Tc	0.7719298	0.6416366	0.8726029	8
N, Ax, Fg, Tw, Bu/To, Ne, An, Tc	0.7719298	0.6416366	0.8726029	8
N, Ax, Fg, Tw, Ea, Ne, An, Tc	0.7719298	0.6416366	0.8726029	8
N, Ic, Fg, Tw, Bu/To, Ea, Ne, Tc	0.7543860	0.6223897	0.8587314	8
N, Ic, Fg, Bu/To, Ea, Ne, An, Tc	0.7543860	0.6223897	0.8587314	8
N, Fg, Tw, Bu/To, Ea, Ne, An, Tc	0.7543860	0.6223897	0.8587314	8
Ax, Ic, Fg, Tw, Bu/To, Ea, Ne, An	0.7543860	0.6223897	0.8587314	8
Ax, Ic, Fg, Tw, Bu/To, Ea, An, Tc	0.7543860	0.6223897	0.8587314	8
Ax, Ic, Fg, Tw, Bu/To, Ne, An, Tc	0.7543860	0.6223897	0.8587314	8
N, Ax, Ic, Tw, Bu/To, Ea, An, Tc	0.7500000	0.6162798	0.8560669	8
N, Ax, Ic, Fg, Bu/To, Ea, Ne, Tc	0.7368421	0.6033725	0.8446152	8
N, Ax, Ic, Tw, Bu/To, Ea, Ne, An	0.7368421	0.6033725	0.8446152	8
N, Ax, Ic, Tw, Bu/To, Ne, An, Tc	0.7368421	0.6033725	0.8446152	8
N, Ax, Ic, Tw, Ea, Ne, An, Tc	0.7368421	0.6033725	0.8446152	8
N, Ax, Fg, Tw, Bu/To, Ea, Ne, Tc	0.7368421	0.6033725	0.8446152	8
Ax, Ic, Fg, Tw, Ea, Ne, An, Tc	0.7368421	0.6033725	0.8446152	8
Ic, Fg, Tw, Bu/To, Ea, Ne, An, Tc	0.7368421	0.6033725	0.8446152	8
N, Ax, Fg, Bu/To, Ea, Ne, An, Tc	0.7192982	0.5845711	0.8302708	8
Ax, Ic, Fg, Tw, Bu/To, Ea, Ne, Tc	0.7192982	0.5845711	0.8302708	8
Ax, Ic, Fg, Bu/To, Ea, Ne, An, Tc	0.7192982	0.5845711	0.8302708	8
Ax, Fg, Tw, Bu/To, Ea, Ne, An, Tc	0.7192982	0.5845711	0.8302708	8
N, Ax, Ic, Tw, Bu/To, Ea, Ne, Tc	0.7017544	0.5659736	0.8157119	8
N, Ax, Ic, Bu/To, Ea, Ne, An, Tc	0.6842105	0.5475702	0.8009499	8
N, Ax, Tw, Bu/To, Ea, Ne, An, Tc	0.6842105	0.5475702	0.8009499	8
N, Ic, Tw, Bu/To, Ea, Ne, An, Tc	0.6842105	0.5475702	0.8009499	8

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
Ax, Ic, Tw, Bu/To, Ea, Ne, An, Tc	0.6842105	0.5475702	0.8009499	8

Nine way

- we want to pick the combination of sites (1-way, 2-way, 3-way, etc.) that adds a marginal benefit to the prior set of samples to a master table
- we sort the 6-way comparisons to identify the 6-way site comparisons that maximize our estimate of the proportion of positive patients
- 6 sites were required to achieve 100% sensitivity, identifying all patients colonized at any body site
- 9-way is shown for completeness

```
all9 = combn(unique(newdata1$SiteID)[1:10], 9, simplify = TRUE)

 holder <- vector("list", ncol(all9))
  for(i in 1:ncol(all9)) {
        df  = get_sensitivity_interval_notime(df=newdata1, ref_sites = all9[,i])
      holder[[i]] =  df
}

CI.nineway = do.call("rbind", holder)
colnames(CI.nineway) = c("Prevalence", "95% CI, Lower", "95% CI, Upper", "N.Site
s")
CI.nineway = CI.nineway %>%
  arrange(., desc(`Prevalence`))

kable(CI.nineway, caption = "Sensitivity Analysis") %>%
  kable_styling("basic", full_width = F, bootstrap_options ="condensed")
```

Sensitivity Analysis

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Ic, Fg, Tw, Bu/To, Ea, Ne, An	0.8070175	0.6808902	0.8995278	9
N, Ax, Ic, Fg, Tw, Bu/To, Ea, An, Tc	0.8070175	0.6808902	0.8995278	9
N, Ax, Ic, Fg, Tw, Bu/To, Ne, An, Tc	0.8070175	0.6808902	0.8995278	9
N, Ax, Ic, Fg, Tw, Ea, Ne, An, Tc	0.8070175	0.6808902	0.8995278	9
N, Ic, Fg, Tw, Bu/To, Ea, Ne, An, Tc	0.7894737	0.6611299	0.8862098	9
N, Ax, Ic, Fg, Tw, Bu/To, Ea, Ne, Tc	0.7719298	0.6416366	0.8726029	9
N, Ax, Ic, Fg, Bu/To, Ea, Ne, An, Tc	0.7719298	0.6416366	0.8726029	9
N, Ax, Fg, Tw, Bu/To, Ea, Ne, An, Tc	0.7719298	0.6416366	0.8726029	9
Ax, Ic, Fg, Tw, Bu/To, Ea, Ne, An, Tc	0.7543860	0.6223897	0.8587314	9

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Ic, Tw, Bu/To, Ea, Ne, An, Tc	0.7368421	0.6033725	0.8446152	9

all 10

```
all10 = combn(unique(newdata1$SiteID)[1:10], 10, simplify = TRUE)

 holder <- vector("list", ncol(all10))
  for(i in 1:ncol(all10)) {
        df  = get_sensitivity_interval_notime(df=newdata1, ref_sites = all10[,i])
      holder[[i]] =  df
}

CI.tenway = do.call("rbind", holder)

colnames(CI.tenway) = c("Prevalence", "95% CI, Lower", "95% CI, Upper", "N.Site
s")
CI.tenway = CI.tenway %>%
  arrange(., desc(`Prevalence`))

kable(CI.tenway, caption = "Sensitivity Analysis") %>%
  kable_styling("basic", full_width = F, bootstrap_options ="condensed")
```

Sensitivity Analysis

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
N, Ax, Ic, Fg, Tw, Bu/To, Ea, Ne, An, Tc	0.8070175	0.6808902	0.8995278	10

Let's get ax, ic, n since this is an often used surveillance combination

```
#get the single site sensitivity estimates
survey.sites  = get_sensitivity_interval_notime(df=newdata1, ref_sites = c("Ax",
"Ic", "N"))
colnames(survey.sites) = c("Prevalence", "95% CI, Lower", "95% CI, Upper", "N.Sit
es")
```

Make a table with the best hit for all sitewise permutations

```
CI.combo = rbind(
  survey.sites,
  CI.single[1,],
  CI.twoway[1,],
  CI.twoway[9,],
  CI.threeway[1,],
  CI.fourway[1,],
  CI.fiveway[1,],
  CI.sixway[1,],
  CI.sevenway[1,],
  CI.eightway[1,],
  CI.nineway[1,],
  CI.tenway[1,])


kable(CI.combo, caption = "Sensitivity Analysis") %>%
  kable_styling("basic", full_width = F, bootstrap_options ="condensed") %>%
  footnote(general = "CI = Confidence Interval",
           alphabet = c("Footnote A; Prevalence is defined as the percentage of c
olonized subjects identified by the body site"))
```

Sensitivity Analysis

	Prevalence	95% CI, Lower	95% CI, Upper	N.Sites
Ax, Ic, N	0.6428571	0.5035931	0.7664452	3
N	0.4285714	0.2971209	0.5678410	1
N, Fg	0.6140351	0.4757484	0.7400453	2
Ax, Ic	0.5000000	0.3633554	0.6366446	2
N, Fg, Tw	0.7192982	0.5845711	0.8302708	3
N, Ic, Fg, Tw	0.7543860	0.6223897	0.8587314	4
N, Ic, Fg, Tw, An	0.7894737	0.6611299	0.8862098	5
N, Ax, Ic, Fg, Tw, An	0.8070175	0.6808902	0.8995278	6
N, Ax, Ic, Fg, Tw, Bu/To, An	0.8070175	0.6808902	0.8995278	7
N, Ax, Ic, Fg, Tw, Bu/To, Ea, An	0.8070175	0.6808902	0.8995278	8
N, Ax, Ic, Fg, Tw, Bu/To, Ea, Ne, An	0.8070175	0.6808902	0.8995278	9
N, Ax, Ic, Fg, Tw, Bu/To, Ea, Ne, An, Tc	0.8070175	0.6808902	0.8995278	10

Note:

CI = Confidence Interval

[a] Footnote A; Prevalence is defined as the percentage of colonized subjects identified by the body site

Figure 1A : Estimated prevalence is plotted per site.

Figure 1A: Surveillance of multiple skin, nares, perianal, oral body sites for prevalence and bioburden of C. auris colonization. a, Frequency of colonization by C.auris at 10 body sites, color-coded by habitat, for 57 subjects at the time of first surveillance. The point estimate reflects the proportion of subjects colonized at each body site and error bars encompass the 95% confidence interval.

```
CI.single$site = as.factor(rownames(CI.single))
colnames(CI.single) = c("Prevalence", "95% CI, Lower", "95% CI, Upper", "N.Site
s", "SiteID")
CI.single = plyr::join(CI.single, site_codes)
CI.single$Site.Extended = as.factor(as.character(CI.single$Site.Extended))
ordering = forcats::fct_inorder(levels(CI.single$Site.Extended))
CI.single$Habitat = str_replace_all(CI.single$Habitat, "Stool", "GI Tract")
CI.single$Habitat = str_replace_all(CI.single$Habitat, "Oral", "GI Tract")
Figure1A = ggplot(CI.single, aes(y =  `Prevalence`,
                                      x = forcats::fct_inorder(Site.Extended)))
+
        geom_point(size=3) +
        theme_classic() +
        geom_errorbar(aes(ymin=`95% CI, Lower`,
                          ymax=`95% CI, Upper`), width=.2,
                 position=position_dodge(.9))  + xlab("") + coord_flip() +
        ylim(0, 1) + ylab("Proportion Colonized")+
  theme(text = element_text(size = 14 ))

Figure1A
```

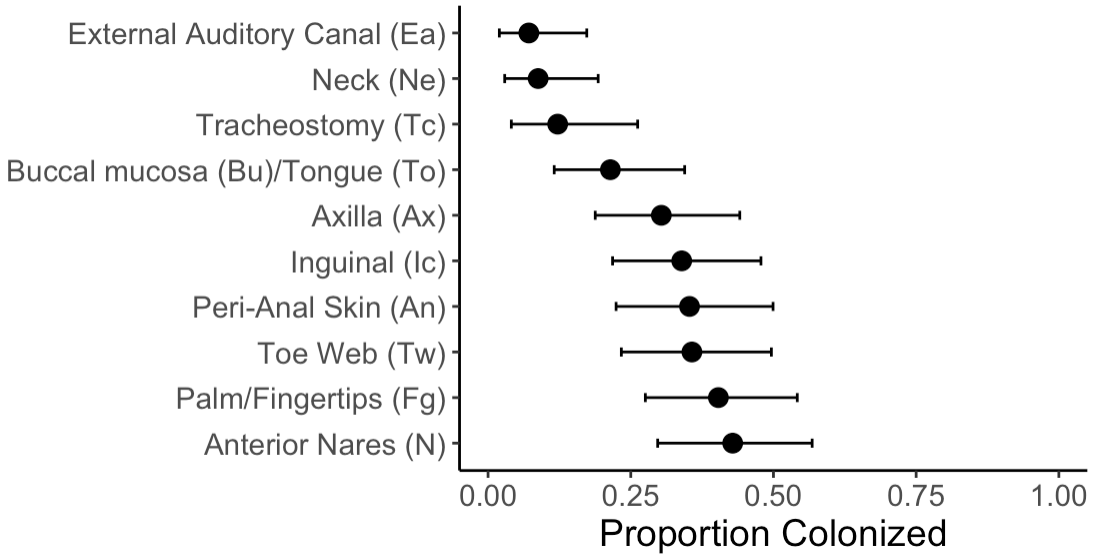


Figure 1B: Sensitivity analysis identifying sites that capture observed prevalence of C. auris colonization based on different pairings of sites

- note that we do not have data for false positives and false negatives; therefore, we cannot formally compute sensitivity, which is defined as TP / (TP+FN). Due to the lack of FP/FN data our

prevalence and sensitivity estimates are equal. Instead, sensitivity is defined by dividing the "estimated site-wise prevalence" by the number of colonized subjects using all sites (0.8070175)

- A minimum of 6 sites was required to achieve 100% sensitivity, capturing all colonized individuals.

Figure 1: Surveillance of multiple skin, nares, perianal, oral body sites for prevalence and bioburden of C. auris colonization. Sensitivity analysis to calculate the proportion of colonized individuals captured by performing surveillance on different pairings of sites. Sensitivity is defined as the Proportion Colonized at each site divided by the total number of individuals identified as colonized at any body site. From left to right, the two vertical lines correspond to the sensitivity of Axilla/Inguinal Crease and Axilla/Inguinal Crease/ Nares. A minimum of 6 sites was required to achieve 100% sensitivity, capturing all colonized individuals.

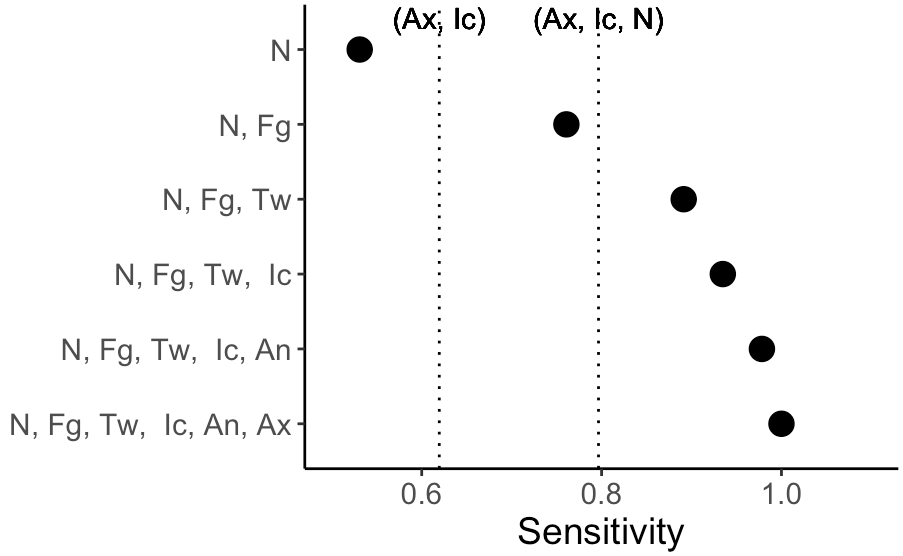
```
true.prevalence.t1 =  0.8070175

sensitivity = data.frame(CI.combo[,1:3]/true.prevalence.t1, CI.combo[,4])
colnames(sensitivity) = c("Sensitivity", "95% CI, Lower", "95% CI, Upper", "N.Sit
es")
sensitivity$Site = rownames(sensitivity)
sensitivity = subset(sensitivity, N.Sites <7)

sensitivity = sensitivity %>%
  arrange(., desc(Sensitivity))
sensitivity$name <- factor(sensitivity$Site, levels = sensitivity$Site)


#make plot
#let's annotate by adding lines at "Ax, Ic" and with "Ax, Ic, N"  rather than inc
luding them as data points
myannotations = subset(sensitivity, name %in% c("Ax, Ic","Ax, Ic, N"))
sensitivity1 = subset(sensitivity, !(name %in% c("Ax, Ic","Ax, Ic, N")))
ax.ic = data.frame(subset(sensitivity, name == "Ax, Ic"))$Sensitivity
ax.ic.n = data.frame(subset(sensitivity, name == "Ax, Ic, N"))$Sensitivity
ax.ic
```

```
## [1] 0.6195653
```

```
ax.ic.n
```

```
## [1] 0.7965839
```

```
#relabel the site codes
sensitivity1$name = plyr::revalue(sensitivity1$name,
                        c("N, Ic, Fg, Tw" = "N, Fg, Tw,  Ic",
                          "N, Ic, Fg, Tw, An" = "N, Fg, Tw,  Ic, An",
                          "N, Ax, Ic, Fg, Tw, An" = "N, Fg, Tw,  Ic, An, Ax"))



Figure1C = ggplot(sensitivity1, aes(y =  `Sensitivity`,
                                    x = `name`)) +
        geom_point(size=4) +
        theme_classic()  + xlab("") + coord_flip() +
        labs(col="Number of sites")  +ylab("Sensitivity") +
        geom_hline(yintercept = ax.ic , linetype = "dotted") +
        geom_text(aes(6, ax.ic,label = "(Ax, Ic)", vjust = -1)) +
        geom_hline(yintercept = ax.ic.n, linetype = "dotted" ) +
        geom_text(aes(6, ax.ic.n,label = "(Ax, Ic, N)", vjust = -1))+
        ylim(0.5, 1.1)+
  theme(text = element_text(size = 14 ))
Figure1C
```



Supplementary Figure 3: Ridgeline plot for samples across sites during the first survey.

Supplementary Figure 3: Ridgeline plot for sample colony counts across sites during the first survey. The cumulative distribution for each ridgeline sums to 1, with peaks corresponding to peak bioburden (log colony forming units), for each site. Sites with low level colonization include Sebaceous sites (External Auditory Canal, Neck) while sites having the highest bioburden include Nares and Inguinal crease.

```
mpn$CFU.Missing = is.na(mpn$CFUTransform_Cauris)
mpn$Log.cfu = log(mpn$CFUTransform_Cauris)
mpn$sqrt.cfu = sqrt(mpn$CFUTransform_Cauris)
cfu = subset(mpn, CFU.Missing==FALSE) %>%
  subset(., CFUTransform_Cauris > 1)  %>%
  subset(., Survey_Period == 1) %>%
  dplyr::select(., c("Log.cfu", "SiteID")) %>%
  plyr::join(., site_codes)

cfu$Habitat = str_replace_all(cfu$Habitat, "Oral", "GI Tract")
cfu$Habitat = str_replace_all(cfu$Habitat, "Stool", "GI Tract")

foo = arrange(Figure1A$data, -Prevalence)
ordering = foo$Site.Extended

cfu$Site.Extended2 <- factor(cfu$Site.Extended, levels = ordering)
SupplementaryFigure3= ggplot(cfu, aes(x = Log.cfu, y = Site.Extended2)) +
  geom_density_ridges(scale = 5, alpha=0.8) + theme_classic() + xlim(0, 12) +
  ylab("") + xlab("Colony Forming Units (CFUs, Log Scale)")

SupplementaryFigure3
```

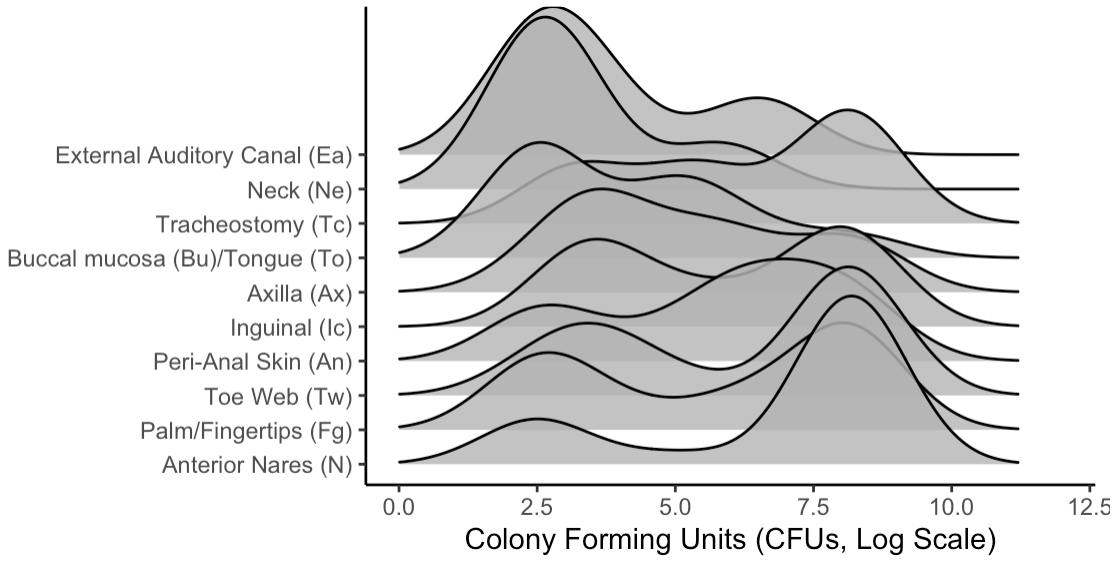


Figure 1C: MPN counts differ by body site and are highest at the nares

Figure 1: Surveillance of multiple skin, nares, perianal, oral body sites for prevalence and bioburden of C. auris colonization.

Number of cultured C. auris colonies, determined by Most Probable Number (MPN), plotted for Inguinal Crease (Ic), Nares (N), and Axilla (Ax). Group wise medians are demarcated with blue lines. Statistical significance of differences was assessed with the Kruskal-Wallis test.

```
mpn =  dplyr::select(mpn, c("Unique_ptid", "Survey_Period", "Cauris_Result", "sit
e",
                      "CDC_MPN")) %>%
  plyr::join(., site_codes)  %>%
  subset(., SiteID %in% c("Ax", "N", "Ic")) %>%
  subset(., Survey_Period==1)

mpn = subset(mpn, CDC_MPN > 0)
mpn$CDC_MPN_Mod = mpn$CDC_MPN
mpn$CDC_MPN_Mod <- ifelse(is.na(mpn$CDC_MPN_Mod), mpn$Cauris_Result, mpn$CDC_MPN_
Mod)

dfa = subset(mpn, SiteID %in% c("N", "Ic"))
dfa$subset = "a"
dfa$logMPN = log(1+dfa$CDC_MPN_Mod)
dfb = subset(mpn, SiteID %in% c("N", "Ax"))
dfb$subset = "b"
dfb$logMPN = log(1+dfb$CDC_MPN_Mod)
newdf = data.frame(rbind(dfa, dfb))
ordering = c("Anterior Nares (N)", "Axilla (Ax)", "Inguinal (Ic)")

newdf$Site.Extended <- factor(newdf$Site.Extended, levels = ordering)
mpn$Site.Extended <- factor(mpn$Site.Extended, levels = ordering)

table(mpn$Site.Extended)
```

```
##
## Anterior Nares (N)        Axilla (Ax)      Inguinal (Ic)
##                 19                 14                 16
```

```
#make the figure
Figure1B =
  ggplot(mpn, aes(x = factor(Site.Extended,
                    level = c("Anterior Nares (N)", "Axilla (Ax)", "Inguinal (I
c)")),
                  y = CDC_MPN))  +
    stat_summary(fun = median, fun.min = median, fun.max = median,
                 geom = "crossbar", width = 0.5) +
  geom_jitter(width=0.05) +
  theme_classic()  + scale_y_log10(label=comma) +
  xlab("") + ylab("Most Probable Number (MPN)")   +
    theme(text = element_text(size=14),
        axis.text.x = element_text(angle=0, hjust=1))
Figure1B
```

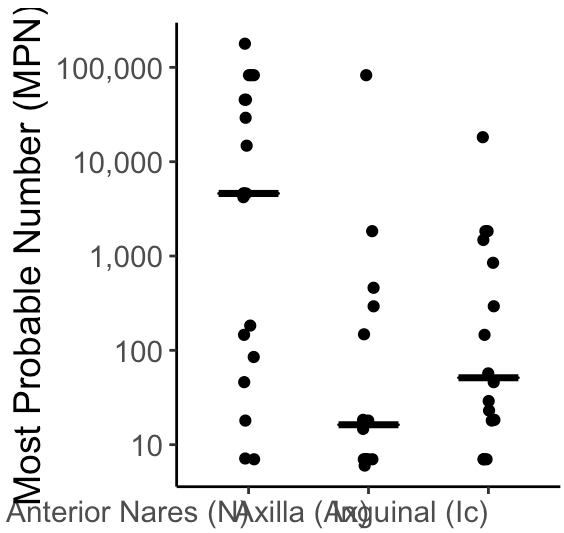
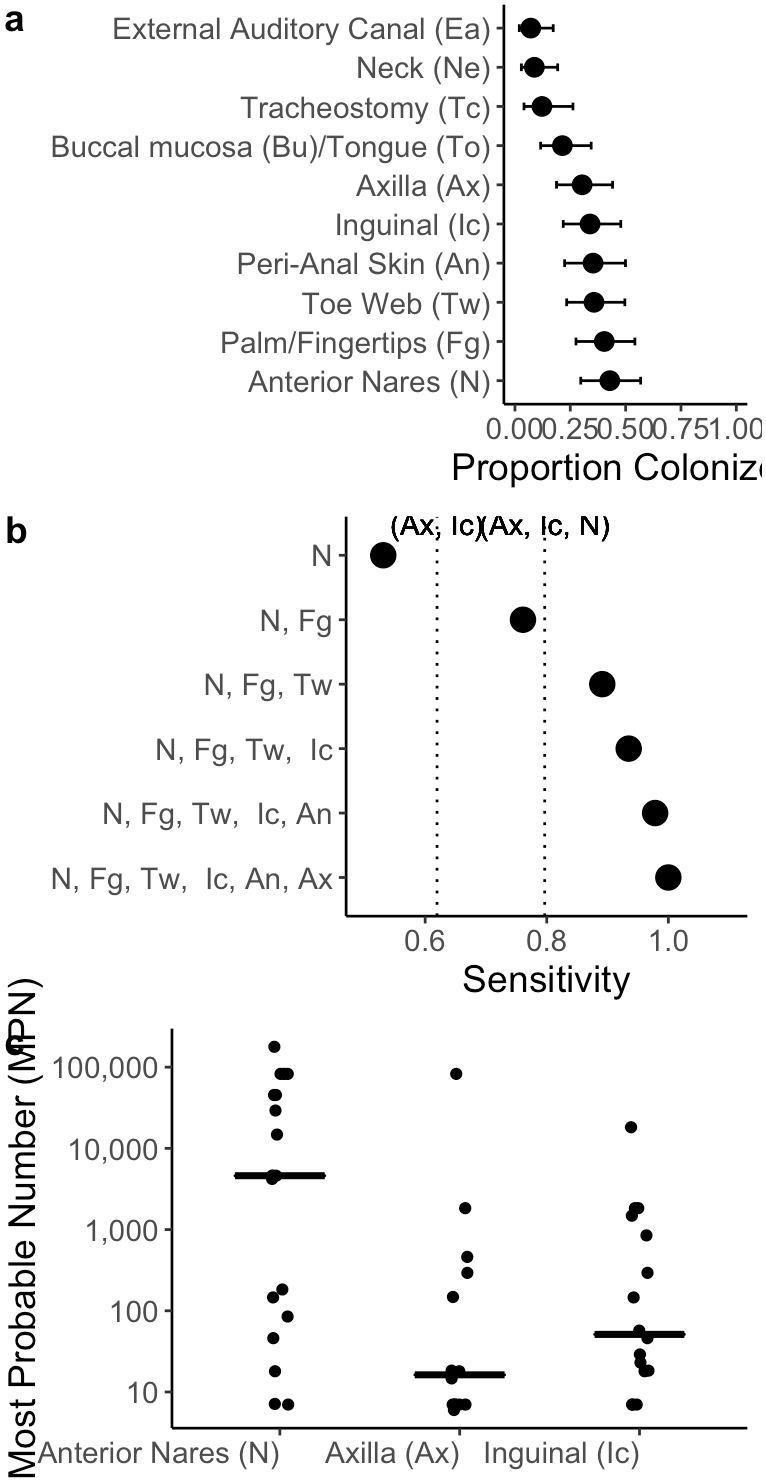


Figure 1: Surveillance of multiple skin, nares, anal, oral body sites for prevalence and patient-level bioburden of C. auris colonization.

Figure 1: Surveillance of multiple skin, nares, perianal, oral body sites for prevalence and bioburden of C. auris colonization. a) Frequency of colonization by C.auris at 10 body sites, color-coded by habitat. Point estimate reflects the proportion of subjects colonized at each body site. Error bars encompass the 95% confidence interval. b) Number of cultured C. auris colonies, determined by Most Probable Number (MPN), plotted for Inguinal Crease (Ic), Nares (N), and Axilla (Ax). Group wise medians are demarcated with blue lines. Statistical significance of differences was assessed with the Kruskal-Wallis test. c) Sensitivity analysis to calculate the proportion of colonized individuals captured by performing surveillance on different pairings of sites. Sensitivity is defined as the Proportion Colonized at each site divided by the total number of individuals identified as colonized at any body site. From left to right, the two vertical lines correspond to the sensitivity of Axilla/Inguinal Crease and Axilla/Inguinal Crease/ Nares. A minimum of 6 sites was required to achieve 100% sensitivity, capturing all colonized individuals.

```
cowplot::plot_grid(
  Figure1A,
  Figure1C,
  Figure1B,
  labels=c("a", "b", "c"), ncol=1,
  rel_heights = c(1,1, 1),
  rel_widths = c(1,1, 1)
)
```

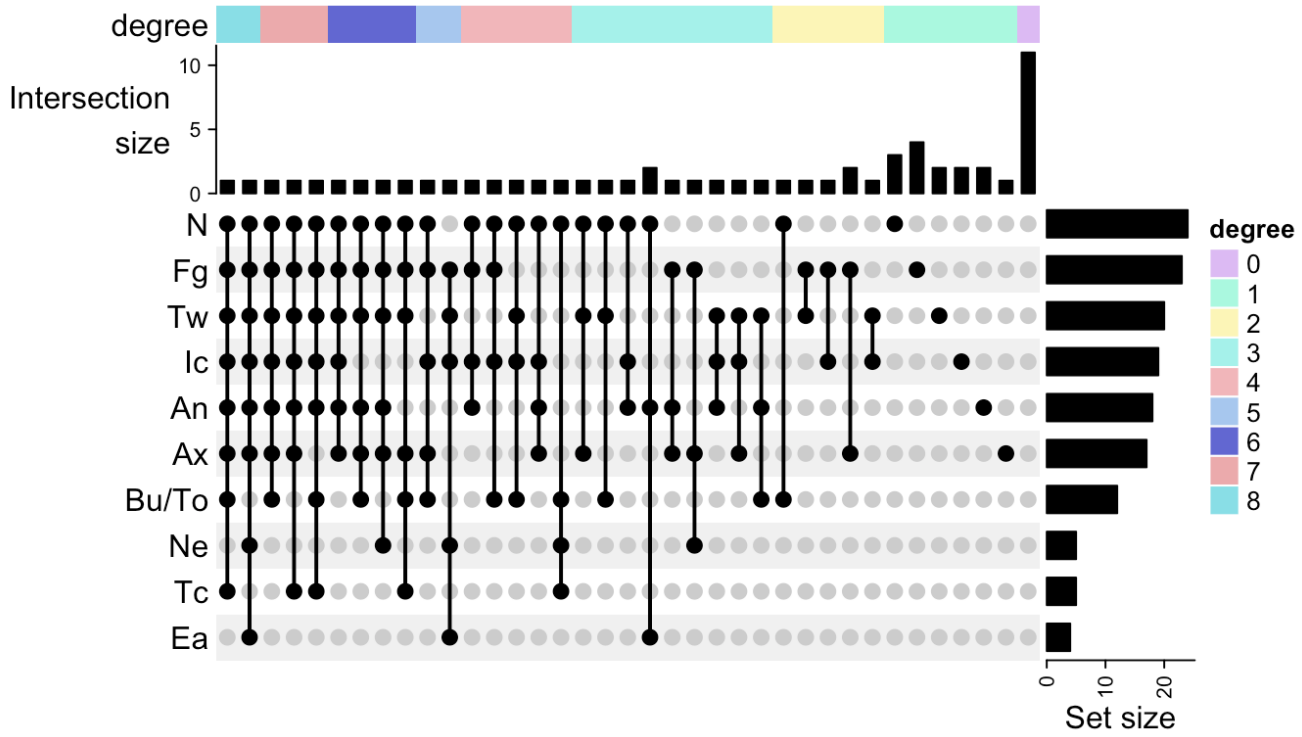

a
External Auditory Canal (Ea)
Neck (Ne)
Tracheostomy (Tc)
Buccal mucosa (Bu)/Tongue (To)
Axilla (Ax)
Inguinal (Ic)
Peri-Anal Skin (An)
Toe Web (Tw)
Palm/Fingertips (Fg)
Anterior Nares (N)
0.000.250.500.751.00
Proportion Coloniz
b
(Ax, Ic)
(Ax, Ic, N)
N
N, Fg
N, Fg, Tw
N, Fg, Tw, Ic
N, Fg, Tw, Ic, An
N, Fg, Tw, Ic, An, Ax
0.6
0.8
1.0
Sensitivity
Most Probable Number (MPN)
100,000
10,000
1,000
100
10
Anterior Nares (N)
Axilla (Ax)
Inguinal (Ic)

```
ggsave(Figure1A, file="~/Desktop/proctor_manuscript/Figure1/Figure1a.eps", device
="eps")
ggsave(Figure1B, file="~/Desktop/proctor_manuscript/Figure1/Figure1b.eps", device
="eps")
ggsave(Figure1C, file="~/Desktop/proctor_manuscript/Figure1/Figure1c.eps", device
="eps")

#save to pdf
ggsave(cowplot::plot_grid(
  Figure1A,
  Figure1B,
  NULL,
  Figure1C,
  labels=c("a", "b", "", "c"), ncol=2,
  rel_heights = c(1,1, 1,1),
  rel_widths = c(1.1,1, 1.1, 1)
), file="~/Desktop/proctor_manuscript/Figure1/Figure1.pdf", width = 12, height =
9, device="pdf")

#save to eps
ggsave(cowplot::plot_grid(
  Figure1A,
  Figure1B,
  NULL,
  Figure1C,
  labels=c("a", "b", "", "c"), ncol=2,
  rel_heights = c(1,1, 1,1),
  rel_widths = c(1.1,1, 1.1, 1)
), file="~/Desktop/proctor_manuscript/Figure1/Figure1.eps", width = 12, height =
9, device="eps")
```

Supplementary Figure 2: Patterns of body site occupancy visualized with UpSetR.

Supplementary Figure 2: Patterns of body site colonization visualized with UpSetR. Colors map to degree, a measure of the number of co-colonized sizes. A total of 36 distinct co-colonization patterns were observed, each arranged from the left to the right as a function of decreasing degree. The intersection size is the number of subjects whose body-site colonization matches the points connecting sites for each of the 36 unique co-colonization patterns. For example, the nares and fingertips are more frequently mono-colonized than any of the other sites while the Bu/To, Neck, Tc, and Ea are never mono-colonized. Most patients have a distinct pattern of co-colonization with the most frequent pattern being the Fg, Ax and the N, An, and Ea. The set size corresponds to the frequency of colonization for each site for the first time point. Colors map to degree, a measure of the number of co-colonized sizes per set.

```
############ Let's subset on the first time point so that we don't have 2-3 entr
ies per patient
newdata1 = subset(data, Survey_Period==1)

#note that subject 29 was not sampled on day 1, just on day 2, so let's grab that
individual, subsetting on time point 2 only
subject29 = subset(data, Unique_ptid=="29" & Survey_Period=="2")
newdata1 = data.frame(rbind(newdata1, subject29))
newdata2 = newdata1  %>%
    subset(., SiteID != "Ax, Ic")
newdata3 = dplyr::select(newdata2, c("Unique_ptid", "SiteID", "Cauris_Result"))


#verify we have 57 subjects
#length(unique(newdata3$Unique_ptid))
rownames(newdata3) = NULL


cast.df<-dcast(newdata3, Unique_ptid~SiteID, value.var="Cauris_Result")
cast.df[is.na(cast.df)] = 0

#make the heatmap
m = make_comb_mat(cast.df)

UpSet(m, top_annotation = HeatmapAnnotation(
    degree = as.character(comb_degree(m)),
    "Intersection\nsize" = anno_barplot(comb_size(m),
        border = FALSE,
        gp = gpar(fill = "black"),
        height = unit(2, "cm")
    ),
    annotation_name_side = "left",
    annotation_name_rot = 0))
```



How many people are person colonized at >=3 sites?

- wee can see that

```
foo = cast.df
rownames(foo) = foo$Unique_ptid
foo$Unique_ptid = NULL
foo$N.positive = rowSums(foo)

tab = data.frame(table(foo$N.positive))

#number of people colonized at >2 sites
sum(tab$Freq[2:8])/57
```

```
## [1] 0.7719298
```

```
#number of people colonized at >2 sites
sum(tab$Freq[2:8])/57
```

```
## [1] 0.7719298
```

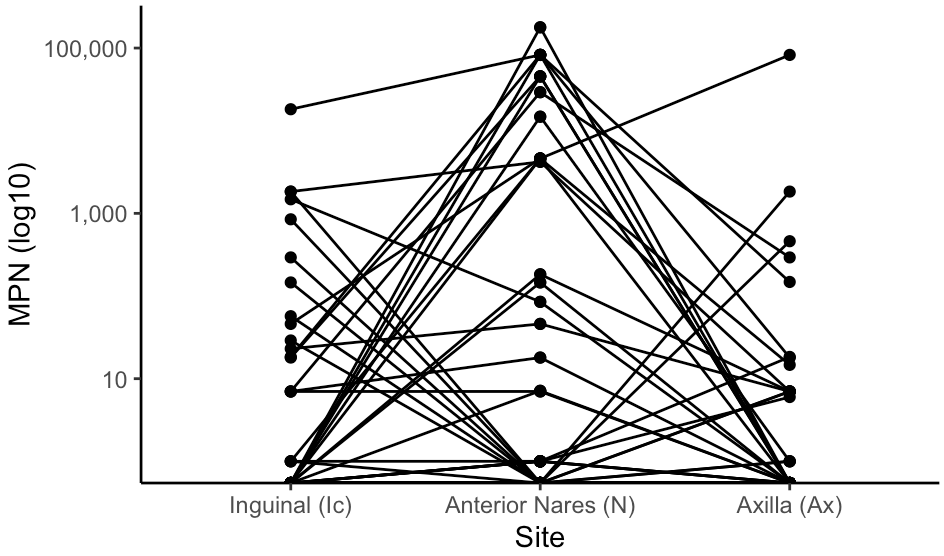
```
#number of people colonized at >3 sites
sum(tab$Freq[3:8])/57
```

```
## [1] 0.5263158
```

Supplementary Figure 4 - paired MPN analysis

Supplementary Figure 4: Paired Most Probable Number (MPN). MPN estimates are shown for the inguinal crease, anterior nares, and axilla. Data represented are from the first point prevalence survey. Each line represents an individual. Individual trajectories reveal a large number of individuals with high counts at the nares and either absent or low level colonization at the axilla or inguinal crease

```
mpn = read.csv("~/Desktop/candida_auris_rush/manuscript/data/Cauris_Analytic_2020
-5-20.csv") %>%
  dplyr::select(., c("Unique_ptid", "Survey_Period", "Cauris_Result", "site",
                     "CDC_MPN")) %>%
  plyr::join(., site_codes)  %>%
  subset(., SiteID %in% c("Ax", "N", "Ic")) %>%
  subset(., Survey_Period==1)


mpn$CDC_MPN_Mod = mpn$CDC_MPN
mpn$CDC_MPN_Mod <- ifelse(is.na(mpn$CDC_MPN_Mod), mpn$Cauris_Result, mpn$CDC_MPN_
Mod)

ordering = c("N", "Ax", "Ic")
mpn$SiteID  <- factor(mpn$SiteID, levels = ordering)

dfa = subset(mpn, SiteID %in% c("N", "Ic"))
dfa$subset = "a"
dfa$logMPN = log(1+dfa$CDC_MPN_Mod)
dfb = subset(mpn, SiteID %in% c("N", "Ax"))
dfb$subset = "b"
dfb$logMPN = log(1+dfb$CDC_MPN_Mod)
newdf = data.frame(rbind(dfa, dfb))
ordering = c("Ax", "N", "Ic")
newdf$SiteID <- factor(newdf$SiteID, levels = ordering)

SupplementaryFigure4 =
  ggplot(newdf, aes(x = factor(Site.Extended,
                    level = c("Inguinal (Ic)" ,  "Anterior Nares (N)",   "Axilla
(Ax)"  )),
                  y = CDC_MPN_Mod, group=Unique_ptid)) +
  geom_point() +
  geom_line()  +
  theme_classic()  + scale_y_log10(label=comma) +
  theme(
  strip.background = element_blank(),
  strip.text.x = element_blank()
)+ xlab("Site") + ylab("MPN (log10)")
SupplementaryFigure4
```



What are the version numbers of all packages and ultilies used in this script?

```
sessionInfo()
```

```
## R version 4.0.2 (2020-06-22)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Mojave 10.14.6
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dy
lib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/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] Hmisc_4.4-2          Formula_1.2-4        lattice_0.20-41
##  [4] cowplot_1.1.1        ComplexHeatmap_2.4.3 dabestr_0.3.0
##  [7] magrittr_2.0.1       ggpubr_0.4.0         ggridges_0.5.3
## [10] harrypotter_2.1.1    wesanderson_0.3.6    phyloseq_1.32.0
## [13] gridExtra_2.3        reshape2_1.4.4       UpSetR_1.4.0
## [16] kableExtra_1.3.1     scales_1.1.1         viridis_0.5.1
## [19] viridisLite_0.3.0    epiR_2.0.19          survival_3.2-7
## [22] forcats_0.5.0        stringr_1.4.0        dplyr_1.0.3
## [25] purrr_0.3.4          readr_1.4.0          tidyr_1.1.2
## [28] tibble_3.0.5         ggplot2_3.3.3        tidyverse_1.3.0
## [31] knitr_1.30
##
## loaded via a namespace (and not attached):
##   [1] readxl_1.3.1        backports_1.2.1     circlize_0.4.12
##   [4] plyr_1.8.6          igraph_1.2.6        splines_4.0.2
##   [7] digest_0.6.27       foreach_1.5.1       htmltools_0.5.1
##  [10] fansi_0.4.2         checkmate_2.0.0     cluster_2.1.0
##  [13] openxlsx_4.2.3      Biostrings_2.56.0   modelr_0.1.8
##  [16] prettyunits_1.1.1   jpeg_0.1-8.1        colorspace_2.0-0
##  [19] rvest_0.3.6         BiasedUrn_1.07      haven_2.3.1
##  [22] xfun_0.20           crayon_1.3.4        jsonlite_1.7.2
##  [25] iterators_1.0.13    ape_5.4-1           glue_1.4.2
##  [28] gtable_0.3.0        zlibbioc_1.34.0     XVector_0.28.0
##  [31] webshot_0.5.2       GetoptLong_1.0.5    car_3.0-10
##  [34] Rhdf5lib_1.10.1     shape_1.4.5         BiocGenerics_0.34.0
##  [37] abind_1.4-5         DBI_1.1.1           rstatix_0.6.0
##  [40] Rcpp_1.0.6          progress_1.2.2      htmlTable_2.1.0
##  [43] clue_0.3-58         foreign_0.8-81      stats4_4.0.2
##  [46] htmlwidgets_1.5.3   httr_1.4.2          RColorBrewer_1.1-2
##  [49] ellipsis_0.3.1      farver_2.0.3        pkgconfig_2.0.3
```

```
##  [52] nnet_7.3-14          dbplyr_2.0.0         labeling_0.4.2
##  [55] tidyselect_1.1.0     rlang_0.4.10         munsell_0.5.0
##  [58] cellranger_1.1.0     tools_4.0.2          cli_2.2.0
##  [61] generics_0.1.0       ade4_1.7-16          broom_0.7.3
##  [64] evaluate_0.14        biomformat_1.16.0    yaml_2.2.1
##  [67] fs_1.5.0             zip_2.1.1            pander_0.6.3
##  [70] nlme_3.1-151         xml2_1.3.2           doBy_4.6.8
##  [73] compiler_4.0.2       rstudioapi_0.13      curl_4.3
##  [76] png_0.1-7            ggsignif_0.6.0       reprex_0.3.0
##  [79] stringi_1.5.3        highr_0.8            Matrix_1.3-2
##  [82] vegan_2.5-7          permute_0.9-5        multtest_2.44.0
##  [85] vctrs_0.3.6          pillar_1.4.7         lifecycle_0.2.0
##  [88] BiocManager_1.30.10 GlobalOptions_0.1.2 data.table_1.13.6
##  [91] R6_2.5.0             latticeExtra_0.6-29 rio_0.5.16
##  [94] IRanges_2.22.2       codetools_0.2-18     boot_1.3-25
##  [97] MASS_7.3-53          assertthat_0.2.1     rhdf5_2.32.4
## [100] rjson_0.2.20         withr_2.4.0          Deriv_4.1.2
## [103] S4Vectors_0.26.1     mgcv_1.8-33          parallel_4.0.2
## [106] hms_1.0.0            rpart_4.1-15         rmarkdown_2.6
## [109] carData_3.0-4        Biobase_2.48.0       lubridate_1.7.9.2
## [112] base64enc_0.1-3
```

Supplementary Data 3

Figure 2: High concentrations of CHG reduce the likelihood of colonization, but are rarely achieved

Diana Proctor

03/26/220

last updated: April 22 2021

---

draft: submission to Nature Medicine October 2020, Revised submission Feb 2021

Manuscript Title: Integrated genomic, epidemiologic investigation of Candida auris skin colonization in a skilled nursing facility

Authors: Diana M Proctor1, Thelma Dangana2, D. Joseph Sexton3, Christine Fukuda2, Rachel D Yelin2, Mary Stanley2, Pamela B Bell2, Sangeetha Baskaran2, Clay Deming1, Qiong Chen1, Sean Conlan1, Morgan Park4, NISC Comparative Sequencing Program4, Rory M Welsh3, Snigdha Vallabhaneni3,5, Tom Chiller3, Kaitlin Forsberg3, Stephanie R. Black6, Massimo Pacilli6, Heidi H Kong7, Michael Y Lin2, Michael E Schoeny8, Anastasia P Litvintseva3, Julia A Segre1+, Mary K Hayden2+

Affiliations:

- 1Microbial Genomics Section, Translational and Functional Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA
- 2 Department of Internal Medicine, Division of Infectious Diseases, Rush University Medical Center, Chicago, IL 60612, USA.
- 3Mycotic Diseases Branch, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA.
- 4NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA.
- 5Division of Healthcare Quality Promotion, NCEZID, CDC, USA
- 6Communicable Disease Program, Chicago Department of Public Health, Chicago, IL, 60612, USA.
- 7Dermatology Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD 20892, USA
- 8College of Nursing, Rush University, Chicago, IL 60612, USA.
- +Contributed equally

Description of the dataset

Here, we seek to examine the impact of CHG on the microbiome, as well as levels of variation in CHG concentration by body site. We have CHG concentrations for the first time point at each body site (Ic, An, Fg, Ax, Ne, Tw). We also have the bacterial and fungal data tables, which includes 365 taxa across 1115 samples.

To accomplish this, we read in the following data:

1. bac_match.rds
2. its_match.rds
3. sitecode_to_factored_sites.csv
4. Cauris_Analytic_2020-5-20.csv
5. CHG_oddsRatios_Schoeny.csv (not provided)

Here, we render the following figures:

- Figure 3
- Supplementary Figure 4

Footnote: The logistic mixed effects model presented in Panel B, and embedded within data object 5 in the list above, was performed by M. Schoeny at Rush University. The code to render the figure is presented here, but the analysis itself was conducted in SAS.

```
#set global knitting options
knitr::opts_chunk$set(echo = TRUE, warning = FALSE, message = FALSE, error = FALSE, fig.width = 6, fig.height = 5
)
```

```
# load package method from from Dan Sprokett

# set seed
set.seed(78979)
#define packages
packages <- c("knitr",
              "ggplot2",
              "scales",
              "ggeffects",
              "lme4",
              "stringr",
              "RColorBrewer",
              "gridExtra",
              "phyloseq",
              "dabestr",
              "tidyr",
              "dplyr",
              "purrr",
              "ggpubr",
              "ggrepel",
              "broom.mixed",
              "lmerTest",
              "cowplot",
              "reshape2")

# install packages from bioconductor
BiocManager::install(setdiff(packages,installed.packages()), update=FALSE)
n <- length(packages) - sum(sapply(packages,require,character.only=TRUE))

# print if packages loaded properly
if(n == 0){
  print("All necessary R packages loaded properly")
} else {
  print(paste0(n, " R packages did not load properly"))
}
```

```
## [1] "All necessary R packages loaded properly"
```

Define functions and then read in the data

```
pal_freiburg_info <- c("#2a6ebb", "#a7c1e3", "#7b2927", "#de3831", "#739600", "#92d400",
                       "#4d4f53", "#747678", "#b2b4b3", "#d5d6d2", "#e98300", "#efbd47")

ISU_secondary_palette <- c("#3E4827", "#76881D", "#A2A569",
                           "#003D4C", "#006BA6", "#7A99AC",
                           "#7C2529", "#9A3324", "#BE531C",
                           "#8B5B29", "#B9975B", "#EED484",
                           "#6E6259", "#707372", "#ACA39A", "#C8102E")

ISU_primary_palette   <- c("#C8102E", "#F1BE48", "#524727",
                           "#9B945F", "#CAC7A7")

#define color palettes
set.seed(1009)
n <- 50
qual_col_pals = brewer.pal.info[brewer.pal.info$category == 'qual',]
col_vector = unlist(mapply(brewer.pal, qual_col_pals$maxcolors, rownames(qual_col_pals)))
```

Read in the data

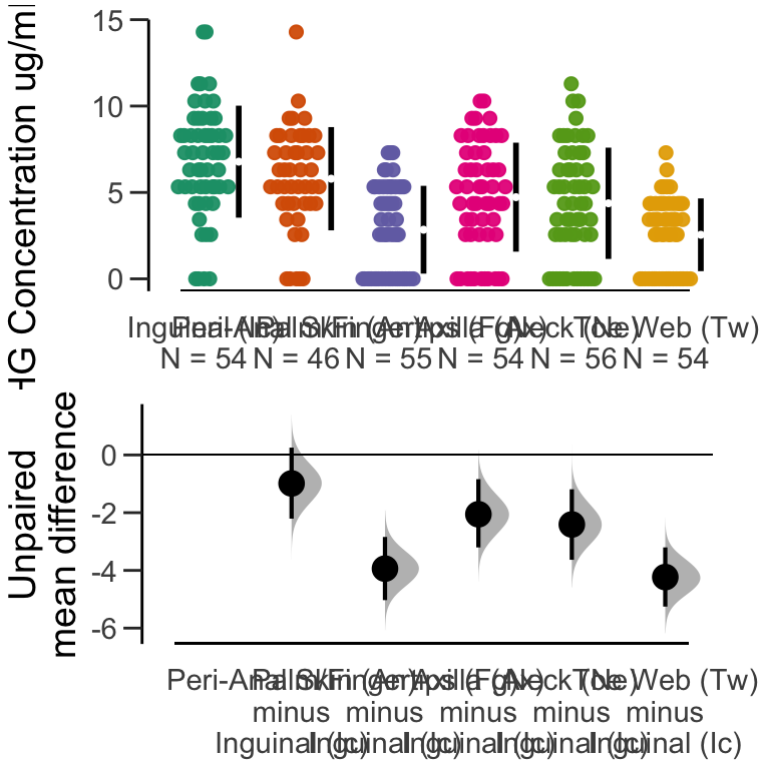
```
bac_match = readRDS(file="~/Desktop/candida_auris_rush/merged_16s_bac_match_cauris_clinical_map_withsqrt_withtree
_withcoo_2020-03-19.rds")
merged_16s = readRDS(file="~/Desktop/candida_auris_rush/merged_16s_nonglom_cauris_clinical_map_withsqrt_withtree_
withcoo_2020-03-19.rds")
its_match = readRDS(file="~/Desktop/candida_auris_rush/merged_its_its_match3_2020-04-02.rds") %>%
      subset_taxa(., Highest.Rank != "Less_than_10_per_ASV Less_than_10_per_ASV") %>%
      subset_taxa(., Highest.Rank != "Fungal sp.")

sites =  read.csv("~/Desktop/candida_auris_rush/sitecode_to_factored_sites.csv")
full.dat = read.csv("~/Desktop/candida_auris_rush/manuscript/data/Cauris_Analytic_2020-5-20.csv") %>%
  dplyr::select(., c("Unique_ptid", "Survey_Period", "Cauris_Result",
                     "site", "chg_conc", "CFUTotal_Cauris", "chgswab", "CFUTransform_Cauris"))  %>%
  plyr::join(., sites)
```

Figure 2A: Let's look at chg concentration as a function of body site

Figure 2: High concentrations of CHG reduce the likelihood of colonization, but are rarely achieved. a) Gardner Altman estimation plot comparing the mean difference in CHG concentrations (ug/mL) across all body sites. Upper panel: scatter plot of CHG concentration plotted as a function of body site (Ic: Inguinal Crease; An: Perianus; Fg: Fingertips; Ax: Axilla; Ne: Neck; Tw: Toeweb) for survey 1. Lower Panel: point estimates for the

mean difference between CHG concentration (ug/mL) at each body site and the Ic, the site reaching the highest average CHG dosage. Error bars encompass the 95% confidence interval surrounding each estimate while the histogram reflects the sampling distribution from a non-parametric bootstrap.

```
full.dat$complete = complete.cases(full.dat$chg_conc)
full.dat = subset(full.dat, complete==TRUE)
full.dat$CFUTransform_Cauris[is.na(full.dat$CFUTransform_Cauris)] <- 0
full.dat$group = paste0(full.dat$Unique_ptid, ";", full.dat$SiteID)
full.dat$`CHG Concentration ug/mL (log2)` = log2(1+full.dat$chg_conc)

t1 = subset(full.dat, Survey_Period==1)
Figure3A <- dabestr::dabest(t1, Site.Extended, `CHG Concentration ug/mL (log2)` ,
                            idx=c("Inguinal (Ic)", "Peri-Anal Skin (An)",  "Palm/Fingertips (Fg)",
                                  "Axilla (Ax)", "Neck (Ne)" , "Toe Web (Tw)" ),
                            paired = FALSE) %>%
                      mean_diff()

plot(Figure3A, palette = "Dark2")
```



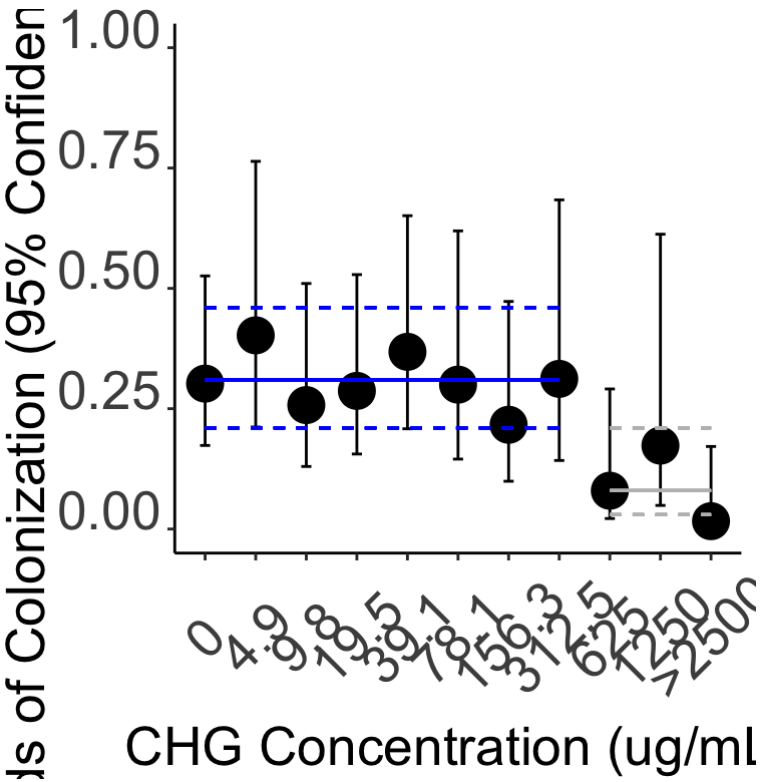
```
ggsave(plot(Figure3A, palette = "Dark2"), file="~/Desktop/proctor_manuscript/Figure3/Figure3a.pdf", device="pdf",
       width = 12, height = 8)
```

Figure 2B: Plot the logistic mixed effects model from Michael Schoeny

Each point represents the odds of C. auris colonization from a logistic mixed effects model plotted against CHG concentration (ug/mL). Error bars encompass 95% confidence intervals. The solid horizontal lines demarcate the odds of colonization per respective group while the dashed lines encompass the 95% CI surrounding each estimate.

```
dat = read.csv("~/Desktop/candida_auris_rush/manuscript/CHG_oddsRatios_Schoeny.csv")
dat$foo = dat$CHG
dat$foo = ifelse(dat$foo==">2500", 2500, dat$CHG)
ordering = unique(dat$CHG)
dat$CHG  <- factor(dat$CHG, levels = ordering)



y1 = mean(subset(dat, CHG %in% c("625", "1250", ">2500"))$Odds)
y2 = mean(subset(dat, CHG %in% c("625", "1250", ">2500"))$Odds)

high = subset(dat, CHG %in% c("625", "1250", ">2500"))
high$start=615
high$end=">2500"
low = subset(dat, !(CHG %in% c("625", "1250", ">2500")))
low$start=0
low$end="313"

#generate the plot
Figure3B = ggplot(dat, aes(x =CHG, y = Odds)) + geom_point(size=6) +
        theme_classic() +
        geom_errorbar(aes(ymin=X95U.u, ymax=X95L.l), width=.2,
                 position=position_dodge(.9))  + xlab("") +
        ylim(0, 1) + xlab("CHG Concentration (ug/mL)") +
        ylab("Odds of Colonization (95% Confidence Intervals)")  +
  geom_segment(aes(x = "0", y = 0.31, xend = "312.5", yend = 0.31),
               colour="blue") +
  geom_segment(aes(x = "0", y = 0.21, xend = "312.5", yend = 0.21), linetype="dashed",
               colour="blue") +
  geom_segment(aes(x = "0", y = 0.46, xend = "312.5", yend = 0.46), linetype="dashed",
               colour="blue") +
  geom_segment(aes(x = "625", y = 0.08, xend = ">2500", yend = 0.08),
               colour="gray")+
  geom_segment(aes(x = "625", y = 0.03, xend = ">2500", yend = 0.03),
               linetype="dashed", colour="gray")+
  geom_segment(aes(x = "625", y = 0.21, xend = ">2500", yend = 0.21),
               linetype="dashed", colour="gray") +
  theme(axis.text.x = element_text( size = 20, angle = 45, hjust = .5, vjust = .5, face = "plain"),
        axis.text.y = element_text(size = 20, angle = 0, hjust = 1, vjust = 0, face = "plain"),
        axis.title.x = element_text( size = 20, angle = 0, hjust = .5, vjust = 0, face = "plain"),
```

```
        axis.title.y = element_text(size = 20, angle = 90, hjust = .5, vjust = .5, face = "plain"))

Figure3B
```




```
ggsave(Figure3B, file="~/Desktop/proctor_manuscript/Figure3/Figure3b.pdf", device="pdf",
       width = 6, height = 6)
```

Supplementay Figure 4: CHG and microbial abundance

- The model: abundance ~ cauris chg_conc + SiteID + (1 | Unique_ptid/Survey_Period) +ϵ

Volcano plot of statistical significance (-Log adjusted P-value) against the regression coefficients from the linear mixed effects models. Each point represents a regression coefficient for a bacterial or fungal species. The vertical lines demarcate regression coefficients of -0.2 and 0.2. Species having Holm adjusted p-values < 0.05 are highlighted in green while non-significant taxa are in blue. Species exhibiting a positive association with CHG concentration (estimate > 0.2, Holm adjusted p < 0.05) include Providencia stuartii, Proteus mirabilis, Candida tropicalis, Saccharomyces cerevisiae and Morganella morganii. Species exhibiting a negative correlation with CHG (estimate < -0.2, Holm adjusted p < 0.05) include Staphylococcus pettenkoferi, Anaerococcus octavius, Malassezia slooffiae, and Campylobacter ureolyticus.

```
#add a variable testing for nas in chg conc' we'll use this to subset the data
sample_data(bac_match)$chg_conc_yn = is.na(sample_data(bac_match)$chg_conc)
sample_data(its_match)$chg_conc_yn = is.na(sample_data(its_match)$chg_conc)

#get rid of the tree - bacteria
otus = otu_table(bac_match)
map = sample_data(bac_match)
tax = bac_match@tax_table@.Data
bac_match = merge_phyloseq(otus, map, tax_table(tax))

#get rid of the tree - fungal
otus = otu_table(its_match)
map = sample_data(its_match)
tax = its_match@tax_table@.Data
its_match = merge_phyloseq(otus, map, tax_table(tax))


#subset the its table to eliminate noisy taxa; otherwise model fails
library(DESeq2);library(genefilter)
filtergroup = genefilter::filterfun(genefilter::kOverA(k=20, A=10)) #k = number of samples; A = abundance
#filter  taxa
        filtPhy = filter_taxa(its_match, filtergroup, prune=TRUE)
        filtPhy = prune_taxa(taxa_sums(filtPhy) > 0, filtPhy)
        filt_its = subset_samples(filtPhy, Unique_ptid != 32)

#generate a combined fungal/bacterial table so we can adjust pvalues in the regression appropriately
#get the number of taxa
phy = merge_phyloseq(bac_match, filt_its)
ntaxa(phy)
```

```
## [1] 365
```

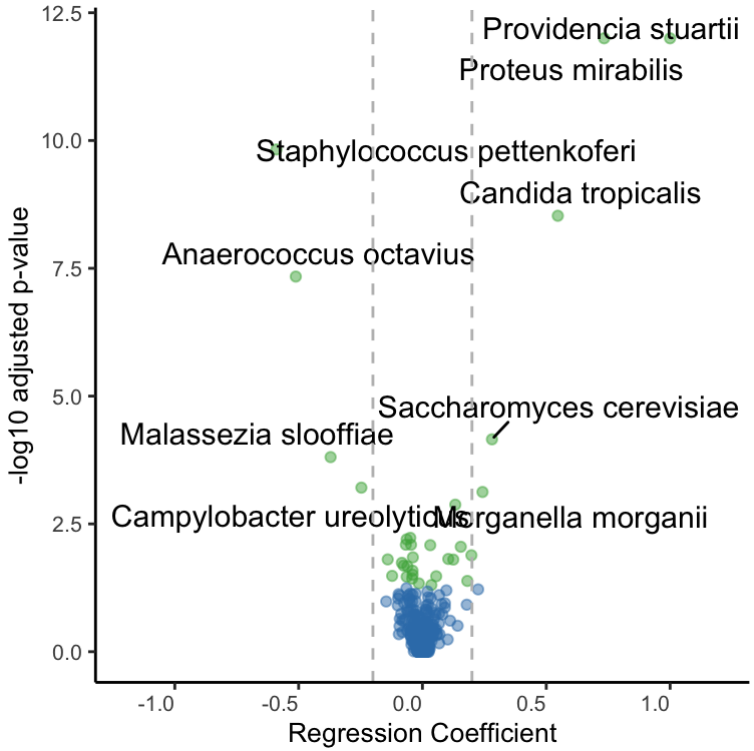
```
nsamples(phy)
```

```
## [1] 1115
```

```
#subset the phyloseq object on samples with chg conc, specific sites shown in panel a, and clr transform
chgPhy= subset_samples(phy, chg_conc_yn =="FALSE") %>%
  transform_sample_counts(., function(x) compositions::clr(x)) %>%
  subset_samples(., SiteID %in% c("Tw", "Fg", "Ic", "An", "Ax"))


set.seed(78927)
#make a map for the regression
map = data.frame(sample_data(chgPhy)) %>%
  dplyr::select(., c("chg_conc", "Unique_ptid", "SiteID", "Survey_Period"))

#log transform chg conc
map$chg_conc = log2(1+map$chg_conc)

#convert site and survey period to numeric
map$SiteID = as.numeric(factor(map$SiteID))
map$Survey_Period = as.numeric(as.character(map$Survey_Period))
map = data.frame(scale(map))

#get the otu table of the centered log ratio table
otus = data.frame(otu_table(chgPhy))

attach(map)
all=data.frame(cbind(otus, map))

#set up empty lists
mod <- list()
out <- list()
adjp <- list()

#https://stackoverflow.com/questions/57590176/adjust-p-values-obtained-with-lmertestlmer-for-multiple-comparisons
adjMC <- function( model_name ) {
  model_glht <- glht(model_name)
  model_MCadj <- summary(model_glht, test = adjusted('holm')) # Bonferroni-Holm
  return(model_MCadj)
}

library(multcomp)
for(i in names(otus)[-1]){
  mod[[i]] <- lmerTest::lmer(get(i) ~ chg_conc  + SiteID +
```

```
                                (1 | Unique_ptid / Survey_Period  ),
      data = all)
  adjp[[i]]  = adjMC(mod[[i]])
  out[[i]] = broom.mixed::tidy(adjp[[i]], conf.int = TRUE, .name_repair = "unique")
}
keep = colnames(bac_match@tax_table@.Data)[1:8]
tax = data.frame(filt_its@tax_table@.Data) %>%
  dplyr::select(., all_of(keep))
out = out %>% map_dfr(~ .x %>% as_tibble(), .id = "Highest.Rank")
out$Highest.Rank = str_replace_all(out$Highest.Rank, "([.])", " ")

df = data.frame(out) %>%
  plyr::join(tax) %>%
  subset(., contrast=="chg_conc")

tax.df = df

#make a volcano plot
pal <- brewer.pal(n = 4, name = 'Set1')
df$adj.p.value = as.numeric(as.character(df$adj.p.value))
my.annotation = subset(df, adj.p.value  < 0.05)
keepers = c("Staphylococcus pettenkoferi",
    "Anaerococcus octavius",
    "Malassezia slooffiae",
    "Campylobacter ureolyticus",
    "Morganella morganii",
    "Saccharomyces cerevisiae",
    "Candida tropicalis",
    "Proteus mirabilis",
    "Providencia stuartii")

df$Highest.Rank = as.character(as.factor(df$Highest.Rank))
df$neglog = -log10(df$adj.p.value)
df$neglog  = ifelse(df$neglog=="Inf", 12, df$neglog)

## Create a column to indicate which genes to label
df$species.label = ifelse(df$Highest.Rank %in% keepers, "TRUE", "FALSE")
df$significance = ifelse(df$adj.p.value < 0.05, "significant", "not")
myannotations = subset(df, species.label==TRUE & estimate > 0.2 | estimate < -0.2)

#plot figure 3c
```

```
Figure3C = ggplot(df) +
  geom_point(aes(x = estimate, y = neglog,
                                          color=significance), alpha=0.5) +
  geom_text_repel(data=myannotations, aes(x = estimate, y = neglog, label = Highest.Rank),size=4) +
  xlab("Regression Coefficient") +
  ylab("-log10 adjusted p-value")  +
  xlim(-1.2, 1.2)  +
  geom_vline(xintercept = -0.2, linetype='dashed', color="gray") +
  geom_vline(xintercept = 0.2, linetype='dashed', color="gray") +
  theme_classic() +
  scale_color_manual(values=c("#377EB8" ,"#4DAF4A"))+
    theme(text = element_text(size=10),
        axis.text.x = element_text(angle=0, hjust=1)) + theme(legend.position = "none") +
  scale_y_continuous(label=comma)

Figure3C
```



```
ggsave(Figure3C, file="~/Desktop/candida_auris_rush/manuscript/NatureMedicine_revision/Figure3c.pdf", device="pd
f",
       width = 10, height = 10)
```

Supplementary Figure 4

Let's look at the taxa with significant coefficients. We'll plot abundance vs. CHG concentration.

```
keep = filter(df, adj.p.value < 0.05) %>%
  filter(., estimate > 0.25 | estimate < -0.25 )
keep.names = str_replace_all(keep$Highest.Rank, "([.])", " ")

sigPhy = prune_taxa(keep.names, chgPhy) %>%
  subset_samples(., abx_rx %in% c(0, 1))  %>%
  prune_samples(sample_sums(.) > 0, .)


df = data.frame(otu_table(sigPhy), sample_data(sigPhy)) %>%
  melt(., id.vars=colnames(sample_data(sigPhy)))
foo = colsplit(df$variable, "(.)", c("Genus", "Species"))
df = data.frame(cbind(df, foo))
df = subset(df, value !=0)

Figure3D = ggplot(df, aes(as.factor(chg_conc), value, color=variable, group=variable)) + geom_point() +
  facet_wrap(~variable, scales="free") + geom_smooth(method="lm") +
  theme_classic() + theme(legend.position = "none")+
  theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank()) + scale_y_log10()


ggsave(Figure3D, file="~/Desktop/candida_auris_rush/manuscript/NatureMedicine_revision/Figure3d.pdf", device="pd
f",
       width = 10, height = 10)
```

Repeat this analysis but this time use an even more stringent filter

```
#add a variable testing for nas in chg conc' we'll use this to subset the data
sample_data(bac_match)$chg_conc_yn = is.na(sample_data(bac_match)$chg_conc)
sample_data(its_match)$chg_conc_yn = is.na(sample_data(its_match)$chg_conc)

#get rid of the tree - bacteria
otus = otu_table(bac_match)
map = sample_data(bac_match)
tax = bac_match@tax_table@.Data
bac_match = merge_phyloseq(otus, map, tax_table(tax))

#get rid of the tree - fungal
otus = otu_table(its_match)
map = sample_data(its_match)
tax = its_match@tax_table@.Data
its_match = merge_phyloseq(otus, map, tax_table(tax))


#subset the its table to eliminate noisy taxa; otherwise model fails
library(DESeq2);library(genefilter)
filtergroup = genefilter::filterfun(genefilter::kOverA(k=20, A=10)) #k = number of samples; A = abundance
#filter  taxa
        filtPhy = filter_taxa(its_match, filtergroup, prune=TRUE)
        filtPhy = prune_taxa(taxa_sums(filtPhy) > 0, filtPhy)
        filt_its = subset_samples(filtPhy, Unique_ptid != 32)

#generate a combined fungal/bacterial table so we can adjust pvalues in the regression appropriately
#get the number of taxa
phy = merge_phyloseq(bac_match, filt_its)
ntaxa(phy)
```

```
## [1] 365
```

```
nsamples(phy)
```

```
## [1] 1115
```

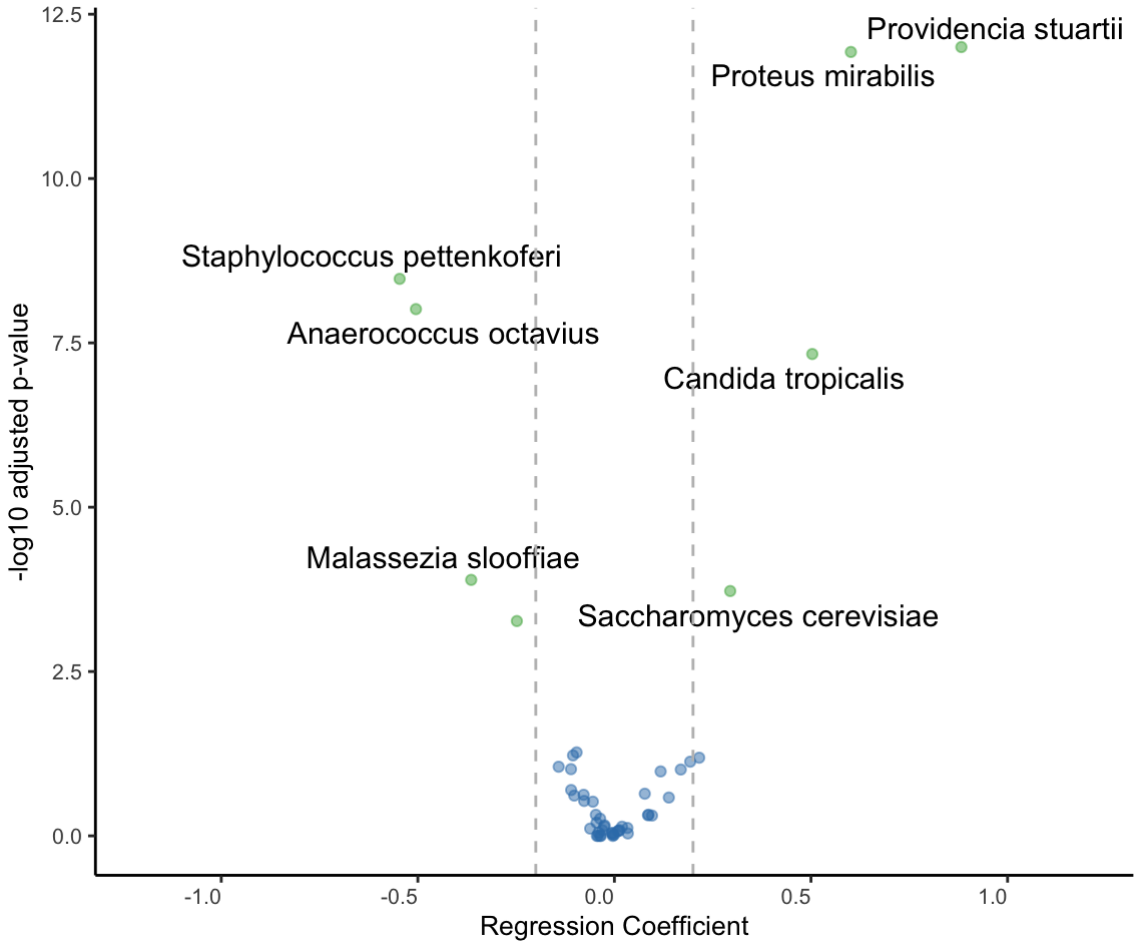
```
top50 = names(sort(taxa_sums(phy), TRUE)[1:50])
phy50 = prune_taxa(top50, phy) %>%
  prune_samples(sample_sums(.) > 0, .)
```

analysis on top 50 taxa

```
#subset the phyloseq object on samples with chg conc, specific sites shown in panel a, and clr transform
chgPhy= subset_samples(phy50, chg_conc_yn =="FALSE") %>%
  transform_sample_counts(., function(x) compositions::clr(x)) %>%
  subset_samples(., SiteID %in% c("Tw", "Fg", "Ic", "An", "Ax"))


set.seed(78927)
#make a map for the regression
map = data.frame(sample_data(chgPhy)) %>%
  dplyr::select(., c("chg_conc", "Unique_ptid", "SiteID", "Survey_Period"))

#log transform chg conc
map$chg_conc = log2(1+map$chg_conc)

#convert site and survey period to numeric
map$SiteID = as.numeric(factor(map$SiteID))
map$Survey_Period = as.numeric(as.character(map$Survey_Period))
map = data.frame(scale(map))

#get the otu table of the centered log ratio table
otus = data.frame(otu_table(chgPhy))

attach(map)
all=data.frame(cbind(otus, map))

#set up empty lists
mod <- list()
out <- list()
adjp <- list()

#https://stackoverflow.com/questions/57590176/adjust-p-values-obtained-with-lmertestlmer-for-multiple-comparisons
adjMC <- function( model_name ) {
  model_glht <- glht(model_name)
  model_MCadj <- summary(model_glht, test = adjusted('holm')) # Bonferroni-Holm
  return(model_MCadj)
}

library(multcomp)
for(i in names(otus)[-1]){
  mod[[i]] <- lmerTest::lmer(get(i) ~ chg_conc  + SiteID +
```

```
                                (1 | Unique_ptid / Survey_Period  ),
      data = all)
  adjp[[i]]  = adjMC(mod[[i]])
  out[[i]] = broom.mixed::tidy(adjp[[i]], conf.int = TRUE, .name_repair = "unique")
}
keep = colnames(bac_match@tax_table@.Data)[1:8]
tax = data.frame(filt_its@tax_table@.Data) %>%
  dplyr::select(., all_of(keep))
out = out %>% map_dfr(~ .x %>% as_tibble(), .id = "Highest.Rank")
out$Highest.Rank = str_replace_all(out$Highest.Rank, "([.])", " ")

df = data.frame(out) %>%
  plyr::join(tax) %>%
  subset(., contrast=="chg_conc")

tax.df = df

#make a volcano plot
pal <- brewer.pal(n = 4, name = 'Set1')
df$adj.p.value = as.numeric(as.character(df$adj.p.value))
my.annotation = subset(df, adj.p.value  < 0.05)

df$Highest.Rank = as.character(as.factor(df$Highest.Rank))
df$neglog = -log10(df$adj.p.value)
df$neglog  = ifelse(df$neglog=="Inf", 12, df$neglog)

## Create a column to indicate which genes to label
df$significance = ifelse(df$adj.p.value < 0.05, "significant", "not")

#plot figure 3c
SupplementaryFigureTop50 = ggplot(df) +
  geom_point(aes(x = estimate, y = neglog,
                                        color=significance), alpha=0.5) +
  geom_text_repel(data=subset(df, adj.p.value < 0.05 & estimate > 0.25 | adj.p.value < 0.05 & estimate < -0.25)
                  , aes(x = estimate, y = neglog, label = Highest.Rank),size=4) +
  xlab("Regression Coefficient") +
  ylab("-log10 adjusted p-value")  +
  xlim(-1.2, 1.2)  +
  geom_vline(xintercept = -0.2, linetype='dashed', color="gray") +
  geom_vline(xintercept = 0.2, linetype='dashed', color="gray") +
  theme_classic() +
```

```
  scale_color_manual(values=c("#377EB8" ,"#4DAF4A"))+
    theme(text = element_text(size=10),
        axis.text.x = element_text(angle=0, hjust=1)) + theme(legend.position = "none") +
  scale_y_continuous(label=comma)

SupplementaryFigureTop50
```

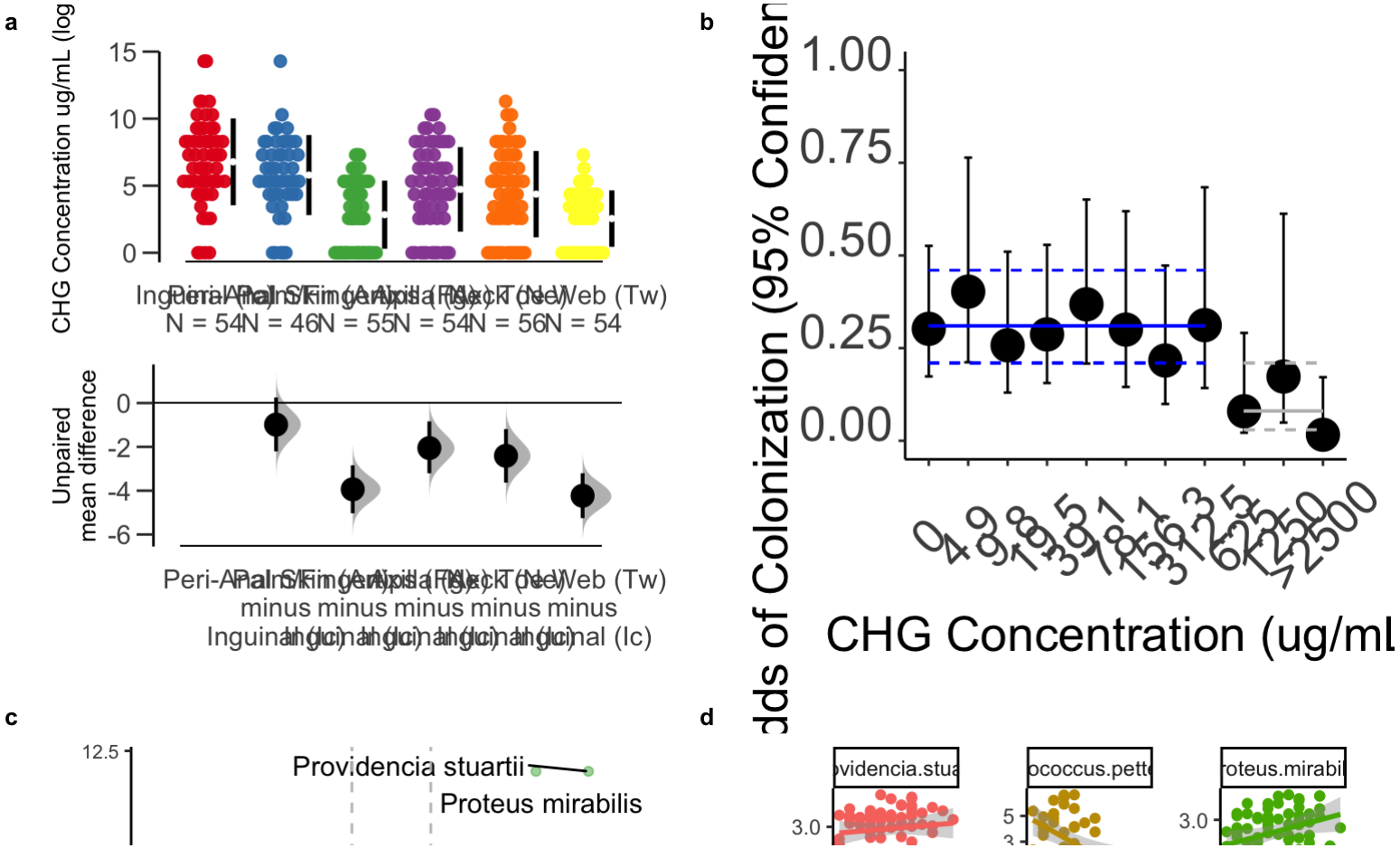


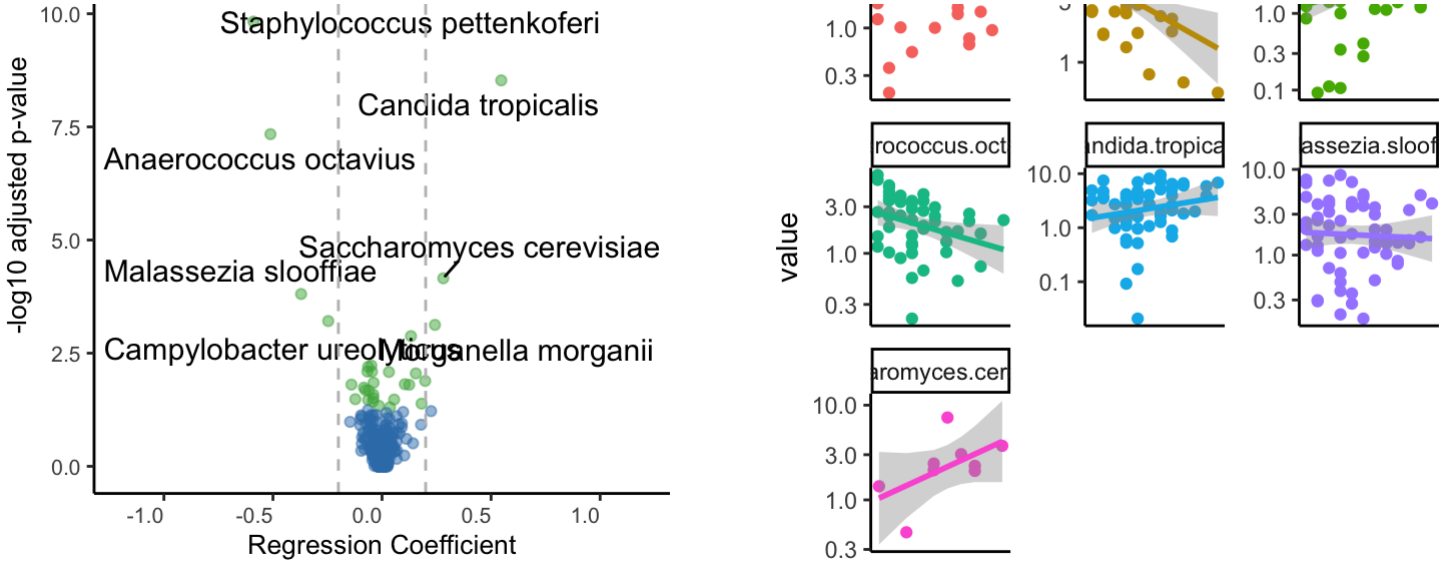
Render Figure 2

Figure 2: High concentrations of CHG reduce the likelihood of colonization, but are rarely achieved. a) Gardner Altman estimation plot comparing the mean difference in CHG concentrations (ug/mL) across all body sites. Upper panel: scatterplot of CHG concentration plotted as a function of body site (Ic: Inguinal Crease; An: Perianus; Fg: Fingertips; Ax: Axilla; Ne: Neck; Tw: Toeweb) for survey 1. Lower Panel: point estimates for the mean difference between CHG concentration (ug/mL) at each body site and the Ic, the site reaching the highest average CHG dosage. Error bars

encompass the 95% confidence interval surrounding each estimate while the histogram reflects the sampling distribution from a nonparametric bootstrap. b) Each point represents the odds of C. auris colonization from a logistic mixed effects model plotted against CHG concentration (ug/mL). Error bars encompass 95% confidence intervals. The solid horizontal lines demarcate the odds of colonization per respective group while the dashed lines encompass the 95% CI surrounding each estimate. c) Volcano plot of statistical significance (-Log adjusted P-value) against the regression coefficients from the linear mixed effects models. Each point represents a regression coefficient for a bacterial or fungal species. The vertical lines demarcate regression coefficients of -0.2 and 0.2. Species having Holm adjusted p-values < 0.05 are highlighted in green while non-significant taxa are in blue. Species exhibiting a positive association with CHG concentration (estimate > 0.2, Holm adjusted p < 0.05) include Providencia stuartii, Proteus mirabilis, Candida tropicalis, Saccharomyces cerevisiae and Morganella morganii. Species exhibiting a negative correlation with CHG (estimate < -0.2, Holm adjusted p < 0.05) include Staphylococcus pettenkoferi, Anaerococcus octavius, Malassezia slooffiae, and Campylobacter ureolyticus.

```
cowplot::plot_grid(plot(Figure3A, axes.title.fontsize = 10), Figure3B, Figure3C,  Figure3D,
                   labels = "auto", label_size = 10, label_x = 0, hjust = 0, scale=0.9)
```






```
ggsave(cowplot::plot_grid(plot(Figure3A, axes.title.fontsize = 10), Figure3B, Figure3C, Figure3D,
                    labels = "auto", label_size = 10, label_x = 0, hjust = 0, scale=0.9),
file="~/Desktop/candida_auris_rush/manuscript/NatureMedicine_revision/Figure3.pdf", device="pdf", width = 12, hei
ght = 10)
```

What are the version numbers of all packages and ultilies used in this script?

```
sessionInfo()
```

```
## R version 4.0.2 (2020-06-22)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Mojave 10.14.6
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/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] parallel  stats4    stats     graphics  grDevices utils     datasets
## [8] methods   base
##
## other attached packages:
##  [1] multcomp_1.4-15             TH.data_1.0-10
##  [3] MASS_7.3-53                 survival_3.2-7
##  [5] mvtnorm_1.1-1               genefilter_1.70.0
##  [7] DESeq2_1.28.1               SummarizedExperiment_1.18.2
##  [9] DelayedArray_0.14.1         matrixStats_0.57.0
## [11] Biobase_2.48.0              GenomicRanges_1.40.0
## [13] GenomeInfoDb_1.24.2         IRanges_2.22.2
## [15] S4Vectors_0.26.1            BiocGenerics_0.34.0
## [17] reshape2_1.4.4              cowplot_1.1.1
## [19] lmerTest_3.1-3              broom.mixed_0.2.6
## [21] ggrepel_0.9.1               ggpubr_0.4.0
## [23] purrr_0.3.4                 dplyr_1.0.5
## [25] tidyr_1.1.2                 dabestr_0.3.0
## [27] magrittr_2.0.1              phyloseq_1.32.0
## [29] gridExtra_2.3               RColorBrewer_1.1-2
## [31] stringr_1.4.0               lme4_1.1-26
## [33] Matrix_1.3-2                ggeffects_1.0.1
## [35] scales_1.1.1                ggplot2_3.3.3
## [37] knitr_1.30
##
## loaded via a namespace (and not attached):
##   [1] readxl_1.3.1           backports_1.2.1        plyr_1.8.6
##   [4] igraph_1.2.6           TMB_1.7.18             splines_4.0.2
##   [7] BiocParallel_1.22.0    digest_0.6.27          foreach_1.5.1
```

```
##  [10] htmltools_0.5.1        fansi_0.4.2            memoise_1.1.0
##  [13] cluster_2.1.0          openxlsx_4.2.3         Biostrings_2.56.0
##  [16] annotate_1.66.0        bayesm_3.1-4           sandwich_3.0-0
##  [19] prettyunits_1.1.1      colorspace_2.0-0       blob_1.2.1
##  [22] haven_2.3.1            xfun_0.20              crayon_1.4.1
##  [25] RCurl_1.98-1.2         jsonlite_1.7.2         zoo_1.8-8
##  [28] iterators_1.0.13       ape_5.4-1              glue_1.4.2
##  [31] gtable_0.3.0           zlibbioc_1.34.0        XVector_0.28.0
##  [34] compositions_2.0-1     car_3.0-10             Rhdf5lib_1.10.1
##  [37] DEoptimR_1.0-8         abind_1.4-5            DBI_1.1.1
##  [40] rstatix_0.6.0          Rcpp_1.0.6             xtable_1.8-4
##  [43] progress_1.2.2         foreign_0.8-81         bit_4.0.4
##  [46] ellipsis_0.3.1         pkgconfig_2.0.3        XML_3.99-0.5
##  [49] farver_2.1.0           locfit_1.5-9.4         utf8_1.1.4
##  [52] tidyselect_1.1.0       labeling_0.4.2         rlang_0.4.10
##  [55] AnnotationDbi_1.50.3   munsell_0.5.0          cellranger_1.1.0
##  [58] tools_4.0.2            generics_0.1.0         RSQLite_2.2.2
##  [61] ade4_1.7-16            sjlabelled_1.1.7       broom_0.7.3
##  [64] evaluate_0.14          biomformat_1.16.0      yaml_2.2.1
##  [67] bit64_4.0.5            robustbase_0.93-7      zip_2.1.1
##  [70] nlme_3.1-151           compiler_4.0.2         beeswarm_0.2.3
##  [73] curl_4.3               ggsignif_0.6.0         geneplotter_1.66.0
##  [76] tibble_3.1.0           statmod_1.4.35         stringi_1.5.3
##  [79] forcats_0.5.0          lattice_0.20-41        nloptr_1.2.2.2
##  [82] tensorA_0.36.2         vegan_2.5-7            permute_0.9-5
##  [85] multtest_2.44.0        vctrs_0.3.6            pillar_1.5.1
##  [88] lifecycle_1.0.0        BiocManager_1.30.10    data.table_1.13.6
##  [91] bitops_1.0-6           insight_0.12.0         R6_2.5.0
##  [94] rio_0.5.16             vipor_0.4.5            codetools_0.2-18
##  [97] boot_1.3-25            assertthat_0.2.1       rhdf5_2.32.4
## [100] withr_2.4.1            GenomeInfoDbData_1.2.3 mgcv_1.8-33
## [103] hms_1.0.0              grid_4.0.2             coda_0.19-4
## [106] minqa_1.2.4            rmarkdown_2.6          carData_3.0-4
## [109] numDeriv_2016.8-1.1    ggbeeswarm_0.6.0
```

Supplementary Data 4

Figure 3: Underlying skin microbiome (fungal and bacterial communities) integrated with C. auris colonization status

Diana Proctor

06/30/2020

draft: submission to Nature Medicine October 2020, Revised submission Feb 2021

Manuscript Title: Integrated genomic, epidemiologic investigation of Candida auris skin colonization in a skilled nursing facility

Authors: Diana M Proctor1, Thelma Dangana2, D. Joseph Sexton3, Christine Fukuda2, Rachel D Yelin2, Mary Stanley2, Pamela B Bell2, Sangeetha Baskaran2, Clay Deming1, Qiong Chen1, Sean Conlan1, Morgan Park4, NISC Comparative Sequencing Program4, Rory M Welsh3, Snigdha Vallabhaneni3,5, Tom Chiller3, Kaitlin Forsberg3, Stephanie R. Black6, Massimo Pacilli6, Heidi H Kong7, Michael Y Lin2, Michael E Schoeny8, Anastasia P Litvintseva3, Julia A Segre1+, Mary K Hayden2+

Affiliations:

- 1Microbial Genomics Section, Translational and Functional Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA
- 2 Department of Internal Medicine, Division of Infectious Diseases, Rush University Medical Center, Chicago, IL 60612, USA.
- 3Mycotic Diseases Branch, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA.
- 4NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA.
- 5Division of Healthcare Quality Promotion, NCEZID, CDC, USA
- 6Communicable Disease Program, Chicago Department of Public Health, Chicago, IL, 60612, USA.
- 7Dermatology Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD 20892, USA
- 8College of Nursing, Rush University, Chicago, IL 60612, USA.
- +Contributed equally

Description of the dataset

For this data set, we have samples for 51 subjects who were sampled for 16S rRNA and ITS1 sequencing. In addition, we have a clinical data set including the variables that were used to generate Table 1. The purpose of this script is to identify associations between bacteria and fungi with respect to Candida auris colonization outcomes (0, 1).

In order to accomplish this, we read in the following data:

1. its_match.rds
2. bac_match.rds

Here, we render the following figures:

1. Figure 3

```
#set global knitting options
knitr::opts_chunk$set(echo = TRUE, warning = FALSE, message = FALSE, error = FALSE)
```

```
packages <- c("knitr",
              "tidyverse",
              "colorspace",
              "viridis",
              "scales",
              "kableExtra",
              "ggplot2",
              "reshape2",
              "gridExtra",
              "phyloseq",
              "ggordiplots",
              "DESeq2",
              "mixOmics",
              "ggpubr",
              "dabestr",
              "compositions",
              "ggsci",
              "RColorBrewer",
              "ggrepel",
              "stringr",
              "yarrr",
              "mixOmics")

# install packages from bioconductor
BiocManager::install(setdiff(packages,installed.packages()), update=FALSE)

#From Dan Sprockett
# load packages
n <- length(packages) - sum(sapply(packages,require,character.only=TRUE))

# print if packages loaded properly
if(n == 0){
  print("All necessary R packages loaded properly")
} else {
  print(paste0(n, " R packages did not load properly"))
}
```

```
## [1] "All necessary R packages loaded properly"
```

set plotting options and define functions

```
safe_colorblind_palette <- c("#88CCEE", "#CC6677", "#DDCC77", "#117733", "#332288", "#AA4499",
                             "#44AA99", "#999933", "#882255", "#661100", "#6699CC", "#888888")
mypal = brewer.pal(9, "Set1")

ISU_secondary_palette <- c("#3E4827", "#76881D", "#A2A569",
                           "#003D4C", "#006BA6", "#7A99AC",
                           "#7C2529", "#9A3324", "#BE531C",
                           "#8B5B29", "#B9975B", "#EED484",
                           "#6E6259", "#707372", "#ACA39A", "#C8102E")



phyToDf <- function(phy, level) {
      ra = transform_sample_counts(phy, function(x) x/sum(x))
      #domain.phy <- tax_glom(ra, taxrank=level)
      tax.count <- data.frame(data.frame(ra@tax_table@.Data, t(otu_table(ra))))
      dfm = melt(tax.count, colnames(tax_table(ra)))
      colnames(dfm)[colnames(dfm) == 'variable'] <- 'Fungal.Matcher'
      sample_data(ra)$Fungal.Matcher =
        str_replace_all(sample_data(ra)$Fungal.Matcher, "([;])", ".")
      df = plyr::join(dfm,data.frame(sample_data(ra)) )
      return(df)
}


#define some functions
add_NSubjects <- function(phy) {
    map = as.data.frame(as.matrix(sample_data(phy))) %>%
      dplyr::count(., PID, sort = FALSE, name = "NSamplesPerSubject")  %>%
      plyr::join(., data.frame(sample_data(phy)), by = "PID") %>%
      tibble::column_to_rownames(., "Fungal.Matcher")%>%
      sample_data(.)
}

add_NSites <- function(phy) {
      map = as.data.frame(as.matrix(sample_data(phy))) %>%
      dplyr::group_by(., Unique_ptid) %>%
      dplyr::count(., SiteID,  name = "NSamplesPerSubjectSite")
      df = as.data.frame(map)
      new_map = merge(df, sample_data(phy), by=c("Unique_ptid", "SiteID"))
```

```
      rownames(new_map) = new_map$Fungal.Matcher
      new_map = sample_data(new_map)
      return(new_map)
}
```

Read in in the data and clean it up

rename the candida genera to candida rename the highest rank to the candida variant of the species name

g__Candida s__jadinii = Candida utils Issatchenkia orientalis = Candida krusei

one asv assigned to the genus candida which couldn't be assigned to the species level and was preesent in only 59 reads, dropped

```
its_match = readRDS(file="~/Desktop/candida_auris_rush/its_match_CDI-out.rds") %>%
                  subset_samples(., Survey_Period==1) %>%
  subset_samples(., Unique_ptid !=32)
sample_names(its_match) = str_replace_all(sample_names(its_match), "Bu/To", "BuTo")
sample_data(its_match)$Fungal.Matcher = sample_names(its_match)

tax = data.frame(its_match@tax_table@.Data)
taxa_names(its_match) = tax$Seq
write.csv(tax, "~/Desktop/candida_auris_rush/its_match_CDI-out-tax_uncorrected.csv")

tax = read.csv("~/Desktop/candida_auris_rush/its_match_CDI-out-tax_corrected.csv")
rownames(tax) = tax$X
tax = as.matrix(tax) %>%
  tax_table(.)
tax_table(its_match) = tax

#update taxa names to highest rank
tax = data.frame(its_match@tax_table@.Data)
new.names = tax$Highest.Rank
taxa_names(its_match) = new.names


#drop taxa that aren't assigned to the genus level

its_match2 = subset_taxa(its_match, !(ASV_Number %in% c(
            "ASV_458",
            "ASV_647",
            "ASV_1314",
            "ASV_1400",
            "ASV_1724",
            "ASV_1792",
            "ASV_1860",
            "ASV_1861",
            "ASV_1897",
            "ASV_2340",
            "ASV_3824")))

its_match2 =  subset_taxa(its_match2, Highest.Rank != "Less_than_10_per_ASV Less_than_10_per_ASV")
#read in the bacterial data
bac_match = readRDS(file="~/Desktop/candida_auris_rush/bac_match_CDI-out.rds") %>%
```

```
  subset_samples(., Unique_ptid !=32) %>%
  subset_samples(., Survey_Period=="1") %>%
  prune_taxa(taxa_sums(.) > 0, .)

sample_names(bac_match) = str_replace_all(sample_names(bac_match), "Bu/To", "BuTo")
sample_data(bac_match)$Fungal.Matcher = sample_names(bac_match)

filtergroup = genefilter::filterfun(genefilter::kOverA(k=76, A=10)) #k = number of samples; A = abundance
bac_filt =filter_taxa(bac_match, filtergroup, prune=TRUE)  %>%
        prune_taxa(taxa_sums(.) > 0, .)


its_filt = its_match2


### get the clinical data
clindf = data.frame(sample_data(its_filt),
                    Unique_ptid=sample_data(its_filt)$Unique_ptid,
                    Site=sample_data(its_filt)$site,
                    Cauris_Result = sample_data(its_filt)$Cauris_Result,
                    Survey_Period=sample_data(its_filt)$Survey_Period)
 clindf = dplyr::select(clindf, c("Unique_ptid", "Survey_Period", "Site",  "Cauris_Result",
                     "trach", "gtube", "urinary_cath", "mech_vent", "age", "braden_score",
                     "pasthospitaladmit", "antifungal_rx", "abx_rx", "contact_iso",
                     "sex", "braden_score"))

clindf$Cauris_Result = as.factor(as.character(clindf$Cauris_Result))
clindf$Cauris_Result = plyr::revalue(clindf$Cauris_Result, c(
      "0"="Colonization Negative",
      "1"="Colonization Positive"))
clindf = subset(clindf, Survey_Period==1)
#fix the clinical data
Y1 = clindf$Cauris_Result
clindf = subset(clindf,  Unique_ptid != 32)
design = dplyr::select(clindf, c("Unique_ptid", "Site", "Survey_Period"))
clinSave = clindf
#clindf = subset(clindf, Unique_ptid != "32")
clindf$Unique_ptid = NULL
clindf$Site = NULL
clindf$Survey_Period = NULL
clindf$Cauris_Result = NULL
```

```
#scale the variables in the clinical table
clin_scales = scale(clindf)


#vst the taxa tables for each dataset independently
clrF = compositions::clr(data.frame(otu_table(its_filt)))
clrB = compositions::clr(data.frame(otu_table(bac_filt)))

#order the data frames so they are the same
clrB = clrB[rownames(clin_scales),]
clrF = clrF[rownames(clin_scales),]

### specify a design matrix
# in design we only need to mention the repeated measurements to split the one level variation
foo = subset_samples(its_filt, Survey_Period==1)
design1 = data.frame(sample_data(foo)) %>%
  dplyr::select("Unique_ptid")

design2 = data.frame(sample_data(foo)) %>%
  dplyr::select("SiteID")

### Within analysis
clinW <- mixOmics::withinVariation(X=clin_scales, design=design1)
clrBw <- mixOmics::withinVariation(X = clrB, design = design2)
clrFw <- mixOmics::withinVariation(X = clrF, design = design2)

### make table lists
Xa <- list(fungi=clrFw, bacteria=clrBw)
```

Figure 3A: Fungal data

Panel A: Each panel encompasses samples for the specified body site with bars representing the relative abundance of taxa for each patient. The inner black curve represents the relative abundance of C. auris for each sample at that site. a) Relative abundance of top fungal genera at each body site surveyed for each individual. Genera included in the `Other` category include: Saccharomyces, Trichosporon, Trichophyton, Aspergillus.

```
ordering =    c("Malassezia",
                "Malasseziales",
                "Other",
                "Candida",
                "Candida utils",
                "Candida glabrata",
                "Candida duobushaemulonii",
                "Candida orthopsilosis",
                "Candida albicans",
                "Candida parapsilosis",
                "Candida tropicalis",
                "Candida auris")

colours= c("Malassezia"="#77AADD",
            "Malasseziales" = "#77AADD",
             "Other"="#EE8866",
             "Candida"= "#1B7837",
              "Candida utils"="#AA4499", #
             "Candida glabrata"="#AAAA00", #
             "Candida duobushaemulonii"="#BBCC33", #
             "Candida orthopsilosis"="#44BB99", #
             "Candida albicans"="#99DDFF", #
            "Candida parapsilosis"="#993404", #
            "Candida tropicalis"="#EEDD88", #
            "Candida auris"="#E7D4E8")

label.vector = c( "Candida utils",
             "Candida glabrata",
             "Candida duobushaemulonii",
             "Candida orthopsilosis",
             "Candida albicans",
            "Candida parapsilosis",
            "Candida tropicalis",
            "Candida auris")
```

Figure 3

```
show_candida_plot_by_rank <- function(phy, Site){
    map = data.frame(sample_data(phy)) %>%
      dplyr::select(., c("Fungal.Matcher", "percentAuris", "Cauris_Result"))
    rownames(map) = NULL
    myranks = map %>%
      arrange(., percentAuris) %>%
      mutate(., Rank=order(percentAuris))%>%
      dplyr::select(., c("Fungal.Matcher", "Rank"))
    myranks$Fungal.Matcher = str_replace_all(myranks$Fungal.Matcher, ";", ".")
    myranks$Fungal.Matcher = as.factor(as.character(myranks$Fungal.Matcher))
    mydf = phyToDf(phy, "Genus")
    df = plyr::join(mydf, myranks)
      df$Genus = str_remove_all(df$Genus, "g__")
      df$Genus = ifelse(!(df$Genus %in% c("Malassezia", "Candida", "Malasseziales")), "Other", df$Genus)
      df$Genus = ifelse(df$Genus=="Malasseziales", "Malassezia", df$Genus)
      df$Label = ifelse(df$Highest.Rank %in% label.vector, df$Highest.Rank, df$Genus)
      df$Label=factor(df$Label, levels = ordering)
      df = subset(df, Cauris_Result %in% c(0, 1))
    p = ggplot(df) +
            geom_col(aes(x=Rank, y=value, fill=Label), position="stack", width=1) +
            geom_line(aes(x=Rank, y = percentAuris, group = 1)) + theme_classic() +
            ylab("") +
            ylim(0, 1) +
            ggtitle(paste0(Site)) +
        theme(plot.title = element_text(size = 12),
        axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank()) +
             scale_fill_manual(values = colours)
    return(p)
    }

#set up the percent auris variable
sample_data(its_match2)$percentAuris = sample_data(its_match2)$caurisReads/sample_sums(its_match2)
map = data.frame(sample_data(its_match2))

sample_data(bac_match) = sample_data(map)

myphy =  subset_samples(its_match2, SiteID=="An") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
```

```
      prune_taxa(taxa_sums(.) > 0, .)
p1 = show_candida_plot_by_rank(phy=myphy, Site="Peri-Anus")  +
             theme(legend.position = "none",
                  text = element_text(size=14),
        axis.text.y = element_text(angle=0, hjust=1))  +
  ylab("Relative Abundance")


myphy =  subset_samples(its_match2, SiteID=="Ax") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)
p2 = show_candida_plot_by_rank(phy=myphy, Site="Axilla") +
             theme(legend.position = "none",
              axis.text.y = element_blank())

myphy =  subset_samples(its_match2, SiteID=="Bu/To") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)

p3 = show_candida_plot_by_rank(phy=myphy, Site="Buccal Mucosa/Tongue") +
             theme(legend.position = "none",
              axis.text.y = element_blank())



myphy =  subset_samples(its_match2, SiteID=="N") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)
p4 = show_candida_plot_by_rank(phy=myphy, Site="Nares")  +
             theme(legend.position = "none",
              axis.text.y = element_blank())

myphy =  subset_samples(its_match2, SiteID=="Ea") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)
p5 = show_candida_plot_by_rank(phy=myphy, Site="Ear canal")  +
             theme(legend.position = "none",
              axis.text.y = element_blank())
```

```
myphy =  subset_samples(its_match2, SiteID=="Tc") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)
p6 = show_candida_plot_by_rank(phy=myphy, Site="Tracheostomy site")  +
             theme(legend.position = "none",
              axis.text.y = element_blank())

myphy =  subset_samples(its_match2, SiteID=="Ic") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)
p7 = show_candida_plot_by_rank(phy=myphy, Site="Inguinal crease")  +
             theme(legend.position = "none",
              axis.text.y = element_blank())


myphy =  subset_samples(its_match2, SiteID=="Fg") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)
p8 = show_candida_plot_by_rank(phy=myphy, Site="Fingertips/Palms") +
             theme(legend.position = "none",
              axis.text.y = element_blank())


myphy =  subset_samples(its_match2, SiteID=="Tw") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)
p9 = show_candida_plot_by_rank(phy=myphy, Site="Toeweb") +
             theme(legend.position = "none",
              axis.text.y = element_blank())


myphy =  subset_samples(its_match2, SiteID=="Ne") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)
p10 = show_candida_plot_by_rank(phy=myphy, Site="Neck") +
             theme(legend.position = "none",
              axis.text.y = element_blank())

#### FIGURE LEGEND
mylegend = show_candida_plot_by_rank(phy=its_match2, Site="All")  +
             theme(legend.position = "bottom") +guides(fill=guide_legend(ncol=10))
```

```
fungi_p = get_legend(mylegend)
```

```
### Make Figure 2a
Figure2_Legend <- cowplot::get_legend(mylegend)
Figure2 <- cowplot::plot_grid((p1 + theme(legend.position = "none")),
                              (p2 + theme(legend.position = "none")),
                              (p3 + theme(legend.position = "none")),
                              (p4 + theme(legend.position = "none")),
                              (p5 + theme(legend.position = "none")),
                              (p6 + theme(legend.position = "none")),
                              (p7 + theme(legend.position = "none")),
                              (p8 + theme(legend.position = "none")),
                              (p9 + theme(legend.position = "none")),
                              (p10 + theme(legend.position = "none")),
                             ncol = 10, hjust = -2.75, vjust = 1.5)
Figure2a <- cowplot::plot_grid(Figure2, Figure2_Legend, nrow = 2)
ggsave(Figure2a, file="~/Desktop/proctor_manuscript/Figure2/Figure2a.pdf",  width = 30, height = 6, units ="in",d
pi = 300, device = "pdf")
ggsave(Figure2a, file="~/Desktop/proctor_manuscript/Figure2/Figure2a.eps",  width = 30, height = 6, units ="in",d
pi = 300, device = "eps")

Figure2a
```

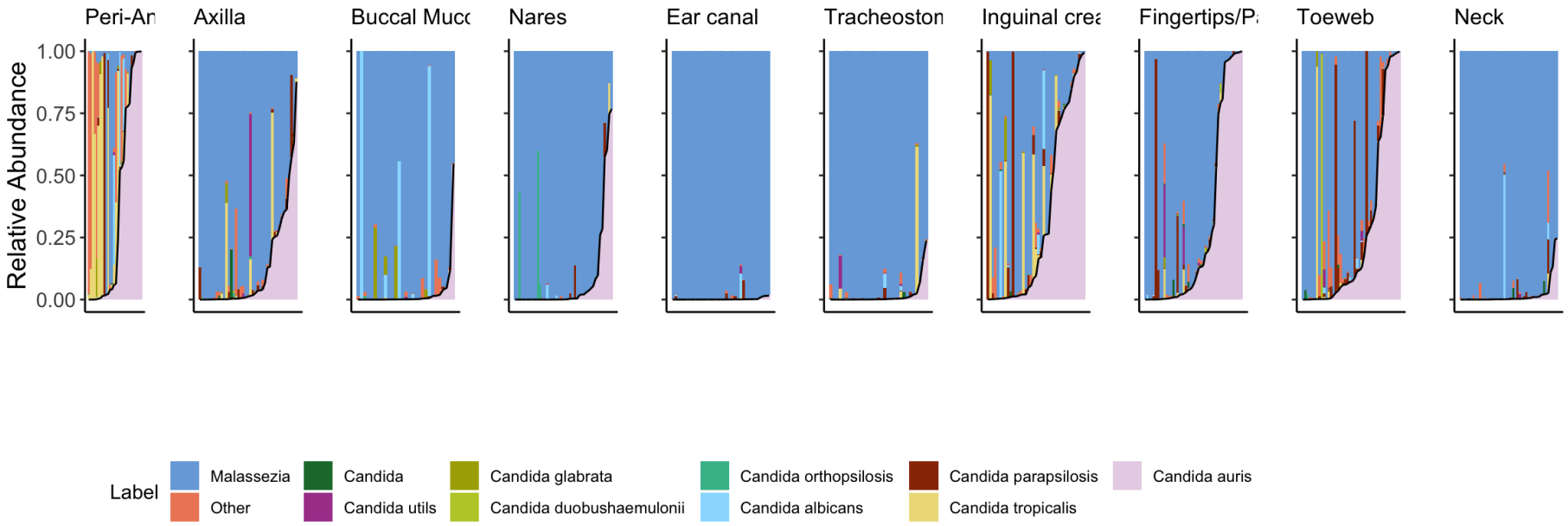


Figure 3B: Plot the bacterial data

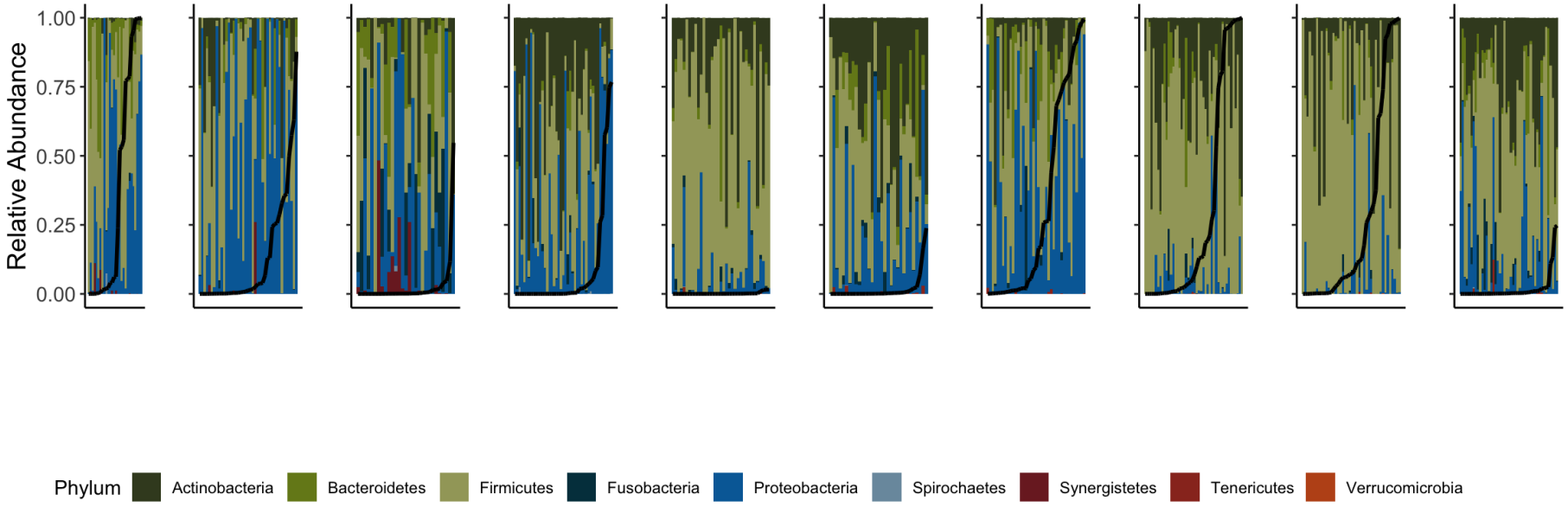
Panel B: Each panel encompasses samples for the specified body site with bars representing the relative abundance of taxa for each patient. The inner black curve represents the relative abundance of C. auris for each sample at that site. b) Relative abundance of bacteria colored by Phylum reveals site-specific associations of C. auris with Proteobacteria.

```
rank_samples_bygenus_and_plot <- function(phy, Site){
    map = data.frame(sample_data(phy)) %>%
      dplyr::select(., c("Fungal.Matcher", "percentAuris", "Cauris_Result"))
    rownames(map) = NULL
    myranks = map %>%
      arrange(., percentAuris) %>%
      mutate(., Rank=order(percentAuris))%>%
      dplyr::select(., c("Fungal.Matcher", "Rank"))
    myranks$Fungal.Matcher = as.factor(as.character(myranks$Fungal.Matcher))
    myranks$Fungal.Matcher = str_replace_all(myranks$Fungal.Matcher, ";", ".")

    mydf = phyToDf(myphy, "Phylum")
    df = plyr::join(mydf, myranks)
    df = subset(df, Cauris_Result %in% c(0, 1))
      p = ggplot(df) +
            geom_col(aes(x=Rank, y=value, fill=Phylum), position="stack", width=1)  +
            ylab("") +   scale_color_manual(values=ISU_secondary_palette) +
            scale_fill_manual(values=ISU_secondary_palette) + ylim(0, 1) +
            geom_line(aes(x=Rank, y = percentAuris, group = 1), size=1) + theme_classic()+
        theme(plot.title = element_text(size = 6),
              axis.title.x=element_blank(),
              axis.text.x=element_blank(),
              axis.ticks.x=element_blank())
    return(p)
    }


myphy =  subset_samples(bac_match, SiteID=="An") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)
p11 = rank_samples_bygenus_and_plot(phy=myphy, Site="Peri-anus")+
              theme(legend.position = "none",
                    text = element_text(size=14),
        axis.text.y = element_text(angle=0, hjust=1))  +
  ylab("Relative Abundance")



myphy =  subset_samples(bac_match, SiteID=="Ax") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
```

```
      prune_taxa(taxa_sums(.) > 0, .)
p12 = rank_samples_bygenus_and_plot(phy=myphy, Site="Axilla") +
             theme(legend.position = "none",
              axis.text.y = element_blank())


myphy =  subset_samples(bac_match, SiteID=="Bu/To") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)
p13 = rank_samples_bygenus_and_plot(phy=myphy, Site="Buccal Mucosa / Tongue")+
             theme(legend.position = "none",
              axis.text.y = element_blank())


myphy =  subset_samples(bac_match, SiteID=="N") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)
p14 = rank_samples_bygenus_and_plot(phy=myphy, Site="Nares")+
             theme(legend.position = "none",
              axis.text.y = element_blank())

myphy =  subset_samples(bac_match, SiteID=="Ea") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)
p15 = rank_samples_bygenus_and_plot(phy=myphy, Site="Ear canal")+
             theme(legend.position = "none",
              axis.text.y = element_blank())



myphy =  subset_samples(bac_match, SiteID=="Tc") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)
p16 = rank_samples_bygenus_and_plot(phy=myphy, Site="Tracheostomy site")+
             theme(legend.position = "none",
              axis.text.y = element_blank())

myphy =  subset_samples(bac_match, SiteID=="Ic") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)
p17 = rank_samples_bygenus_and_plot(phy=myphy, Site="Inguinal crease")+
```

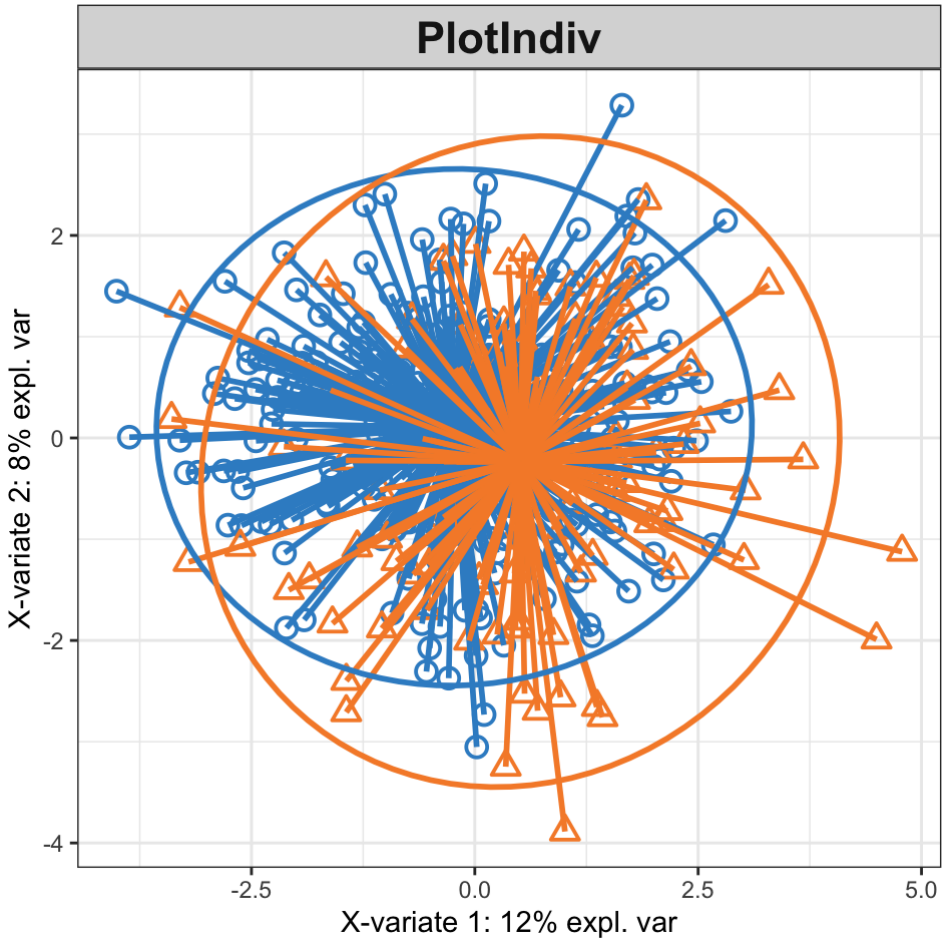
```
             theme(legend.position = "none",
              axis.text.y = element_blank())


myphy =  subset_samples(bac_match, SiteID=="Fg") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)
p18 = rank_samples_bygenus_and_plot(phy=myphy, Site="Fingertips/Palms")+
             theme(legend.position = "none",
              axis.text.y = element_blank())


myphy =  subset_samples(bac_match, SiteID=="Tw") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)
p19 = rank_samples_bygenus_and_plot(phy=myphy, Site="Toeweb")+
             theme(legend.position = "none",
              axis.text.y = element_blank())


myphy =  subset_samples(bac_match, SiteID=="Ne") %>%
      prune_samples(sample_sums(.) > 0, .) %>%
      prune_taxa(taxa_sums(.) > 0, .)
p20 = rank_samples_bygenus_and_plot(phy=myphy, Site="Neck")+
             theme(legend.position = "none",
              axis.text.y = element_blank())

mylegend = rank_samples_bygenus_and_plot(phy=bac_match, Site="Peri-Anus") +
             theme(legend.position = "bottom") +guides(fill=guide_legend(ncol=10))
```

```
### Make Figure 2b
Figure2_Legend <- cowplot::get_legend(mylegend)
Figure2 <- cowplot::plot_grid((p11 + theme(legend.position = "none")),
                              (p12 + theme(legend.position = "none")),
                              (p13 + theme(legend.position = "none")),
                              (p14 + theme(legend.position = "none")),
                              (p15 + theme(legend.position = "none")),
                              (p16 + theme(legend.position = "none")),
                              (p17 + theme(legend.position = "none")),
                              (p18 + theme(legend.position = "none")),
                              (p19 + theme(legend.position = "none")),
                              (p20 + theme(legend.position = "none")),
                             ncol = 10, hjust = -2.75, vjust = 1.5)
Figure2b <- cowplot::plot_grid(Figure2, Figure2_Legend, nrow = 2)
ggsave(Figure2b, file="~/Desktop/proctor_manuscript/Figure2/Figure2b.eps",  width = 30, height = 6, units ="in",d
pi = 300, device = "eps")
ggsave(Figure2b, file="~/Desktop/proctor_manuscript/Figure2/Figure2b.pdf",  width = 30, height = 6, units ="in",d
pi = 300, device = "pdf")
Figure2b
```



Now let's look at variables int he bacterial table that segregate with colonization status

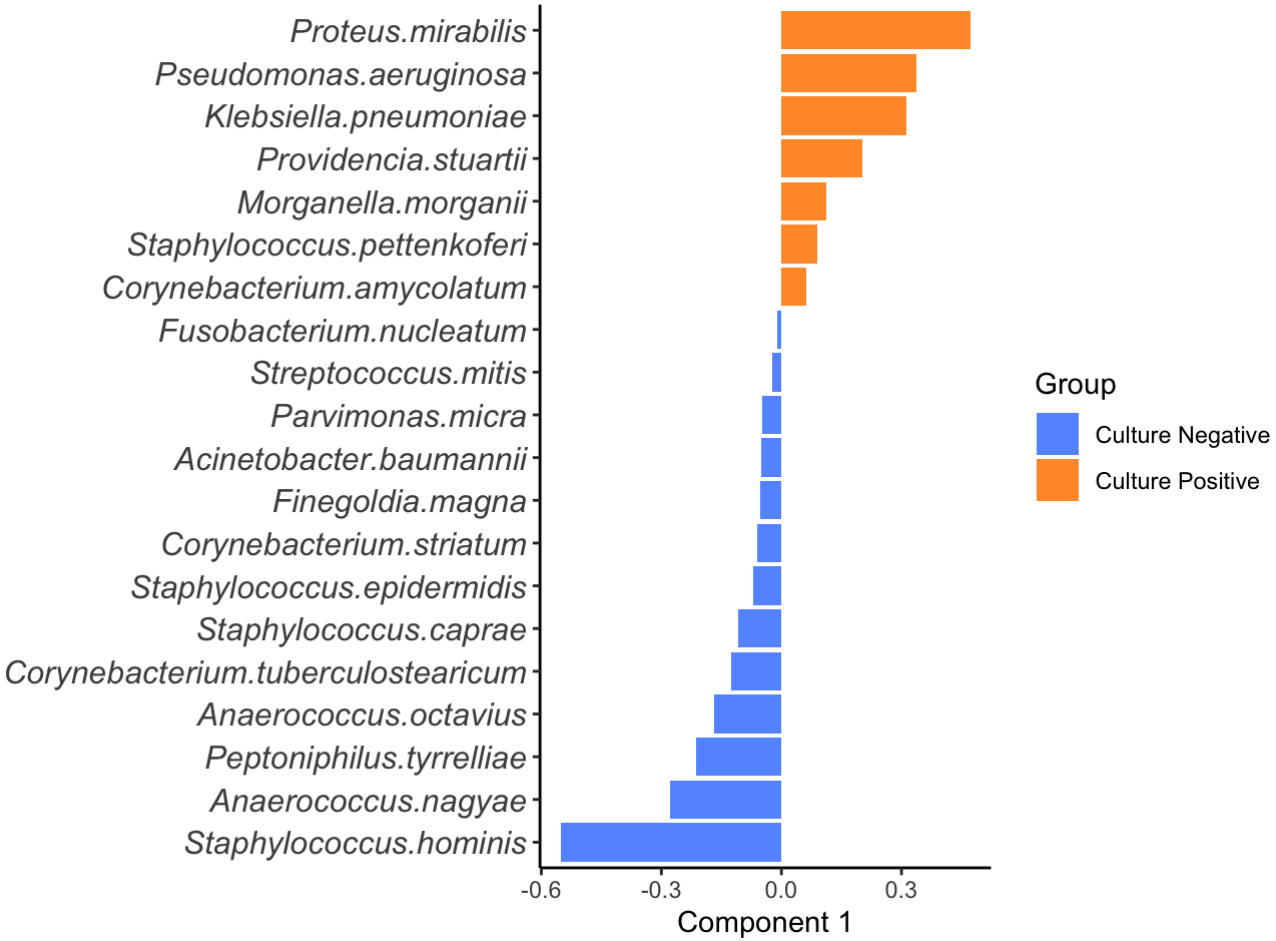
```
bacterial.splsda <- mixOmics::splsda(clrB, Y1, ncomp=2, near.zero.var=TRUE) # 1 Run the method
mixOmics::plotIndiv(bacterial.splsda, ellipse = TRUE, star=TRUE, ind.names = FALSE)                                    # 2
 Plot the samples
```




```
setEPS()
pdf("~/Desktop/proctor_manuscript/Figure2/Figure2c.pdf")

par(mfrow=c(1,1))
mixOmics::plotIndiv(bacterial.splsda, ellipse = TRUE, star=TRUE, ind.names = FALSE)                    # 2
 Plot the samples
dev.off()
```

```
## quartz_off_screen
##                 2
```

```
tax1 = data.frame(bac_filt@tax_table@.Data) %>%
  dplyr::select(., c("Kingdom", "Class", "Order", "Family", "Genus", "Highest.Rank"))
tax1$Highest.Rank = str_replace_all(tax1$Highest.Rank, " ", ".")
loadings.16s = data.frame(bacterial.splsda$loadings$X)
loadings.16s$Highest.Rank= rownames(loadings.16s)
loadings.16s$Highest.Rank = str_replace_all(loadings.16s$Highest.Rank, " ", ".")
loadings.16s = plyr::join(loadings.16s, tax1)

loadings.16s = loadings.16s[with(loadings.16s, order(comp1)),]
ordering16s = as.vector(loadings.16s$Highest.Rank)
loadings.16s$Highest.Rank  <- factor(loadings.16s$Highest.Rank, levels = ordering16s)
loadings.16s$group = ifelse(loadings.16s$comp1 < 0, "Culture Negative", "Culture Positive")


myfont <- element_text(face = "italic",size = 12)

Figure2d = ggplot() +
  geom_col(data=loadings.16s, aes(Highest.Rank, comp1, fill=group)) +
  coord_flip()+ scale_fill_manual(values=c("#6699FF", "#FF9933"), name="Group")+
  theme_classic() +
  ylab("Component 1")+ xlab("")+ theme(axis.text.y = myfont)


Figure2d
```




```
ggsave(Figure2d, file="~/Desktop/proctor_manuscript/Figure2/Figure2d.eps",  width = 6, height = 6, units ="in",dpi = 300, device = "eps")
```

Let's look at thee performance of the model

This suggests that the taxa that segregate culture positive and negative samples discriminate between theem 68% of the time, outperforming a random model, using 2-fold cross validation.

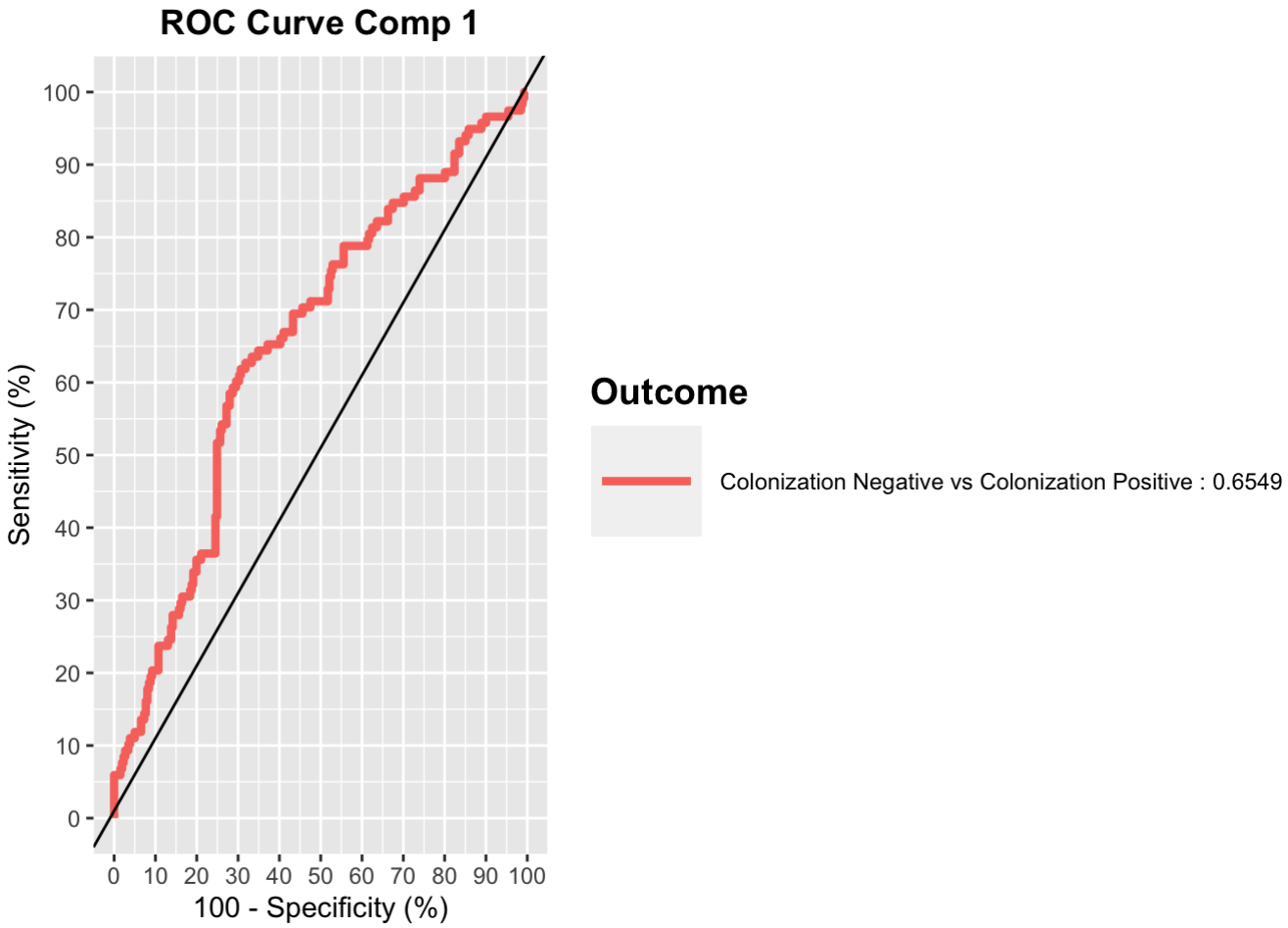
```
#Performance
set.seed(123) # for reproducibility
MyPerf.selected.diablo <- mixOmics::perf(bacterial.splsda, validation = 'Mfold', folds = 2,
                  nrepeat = 100,
                  dist = c('centroids.dist', "max.dist", "mahalanobis.dist")) #,

bacterial.splsda$choice.ncomp
```

```
## NULL
```

```
bacterial.splsda$MajorityVote.error.rate
```

```
## NULL
```

```
auroc(bacterial.splsda)
```



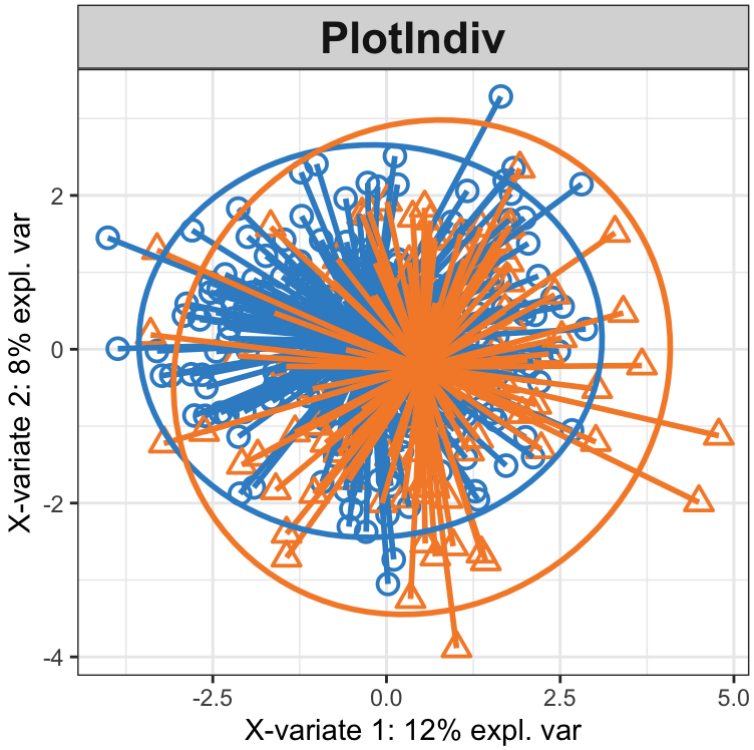

```
## $Comp1
##                                                    AUC   p-value
## Colonization Negative vs Colonization Positive 0.6549 1.365e-06
##
## $Comp2
##                                                    AUC   p-value
## Colonization Negative vs Colonization Positive 0.6654 2.507e-07
```

Let's build a multiple table model

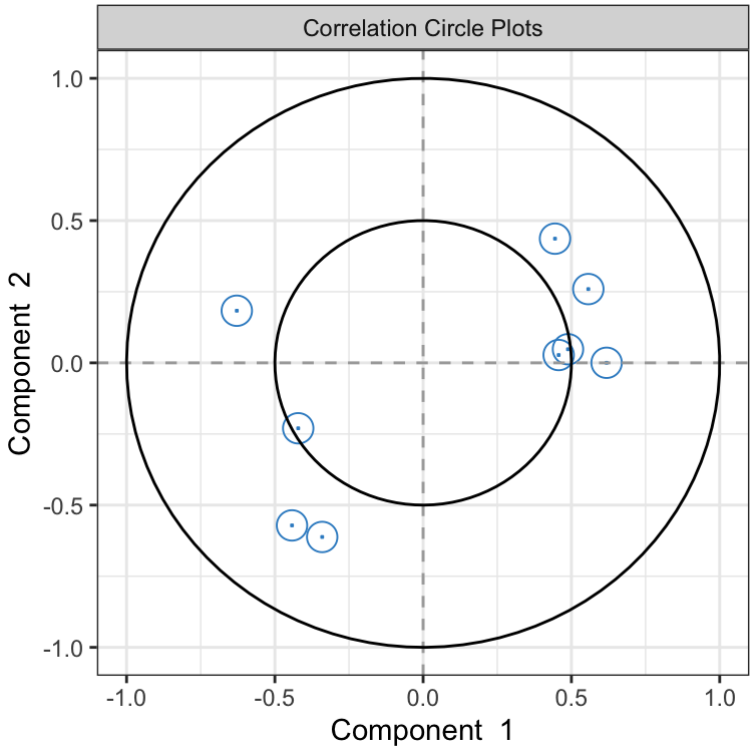
We want to look at the relationships betweeen the taxa in the bacterial and fungal tables within a single analysis. First, we will repeat variable selection on the bacterial table.

Select variables in the bacterial table

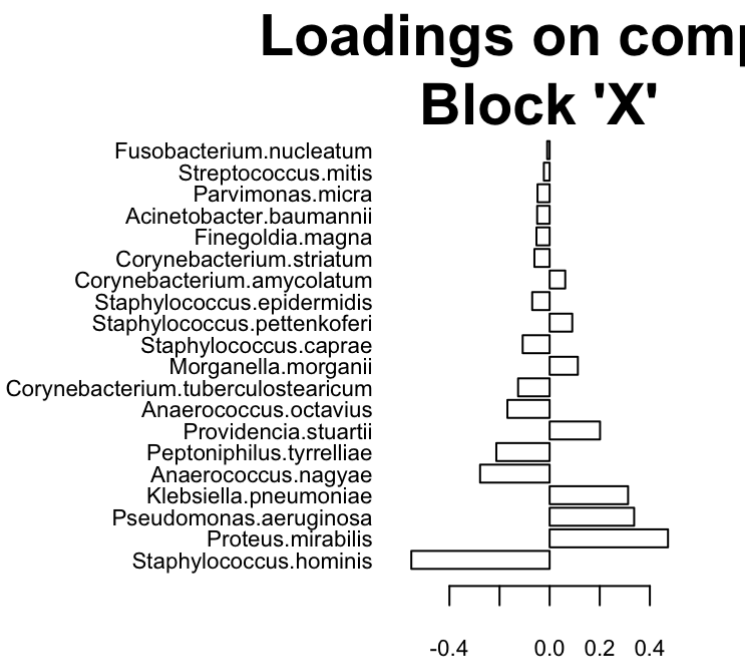
```
bacterial.splsda <- splsda(clrB, Y1, ncomp=8, near.zero.var=TRUE) # 1 Run the method
plotIndiv(bacterial.splsda, ellipse = TRUE, star=TRUE, ind.names = FALSE)                    # 2 Plot the s
amples
```




```
plotVar(bacterial.splsda, cutoff=0.4, var.names = FALSE)
```



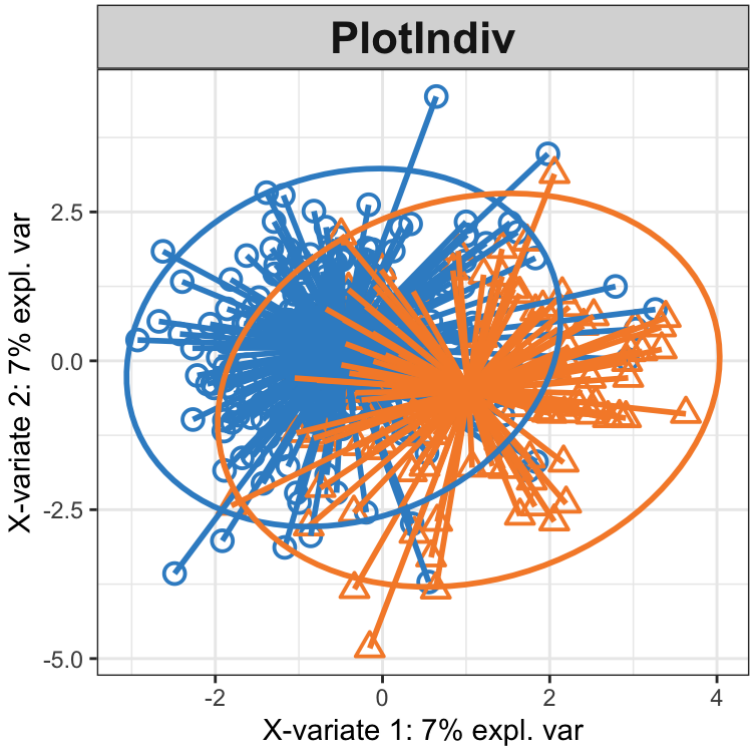

```
plotLoadings(bacterial.splsda)
```



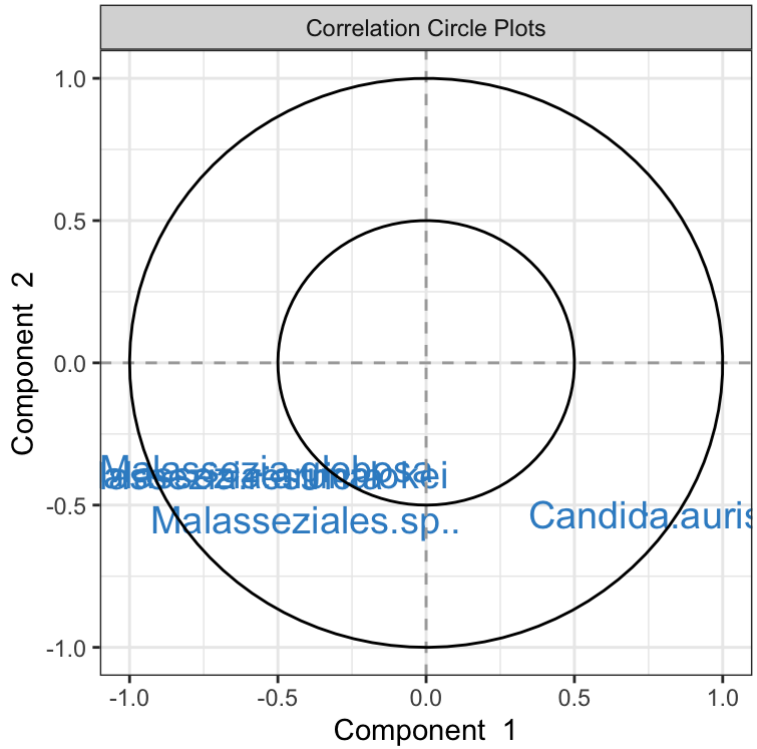

```
keep.bacterial.taxa = selectVar(bacterial.splsda, comp=1)$name
```

Now, we select variables in the fungal table

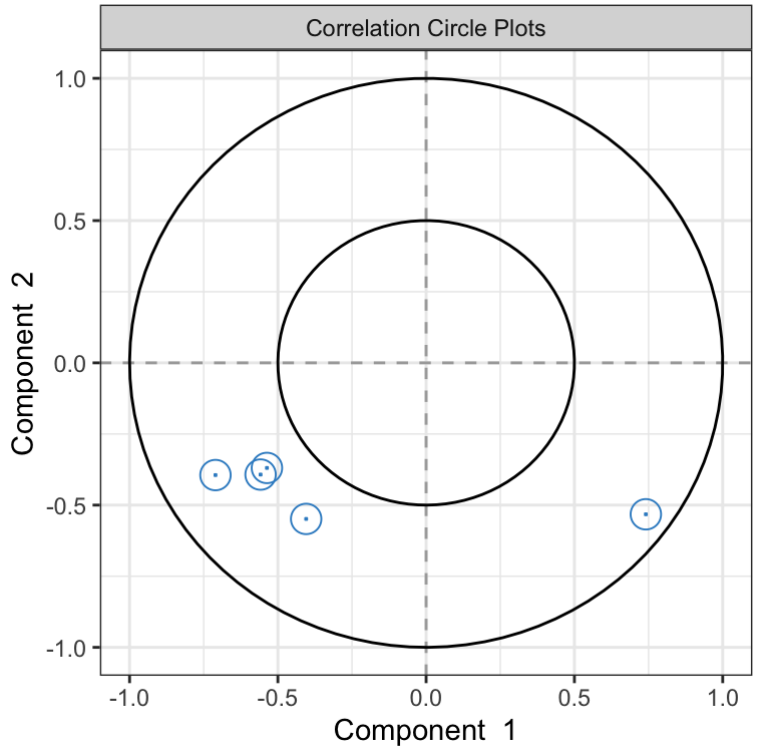
```
fungal.splsda <- splsda(clrF, Y1, ncomp=8, near.zero.var=TRUE) # 1 Run the method
plotIndiv(fungal.splsda, ellipse = TRUE, star=TRUE, ind.names = FALSE)                    # 2 Plot the samp
les
```




```
plotVar(fungal.splsda, cutoff=0.4, var.names = TRUE)
```




```
plotVar(fungal.splsda, cutoff=0.4, var.names = FALSE)
```



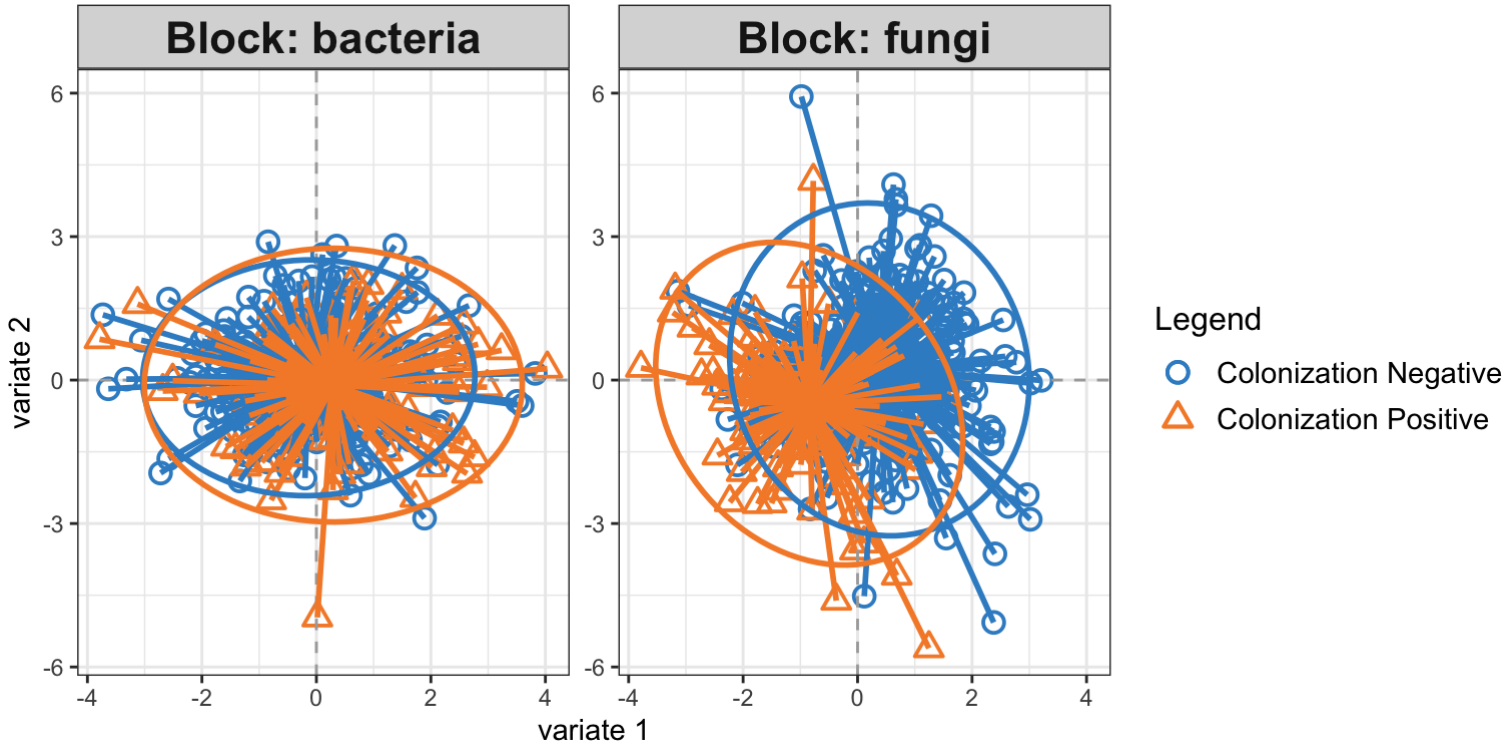

```
keep.fungal.taxa = selectVar(fungal.splsda, comp=1)$name
```

We subset our taxa tables on the selected variables and account for variability between sites.

```
#select the taxa
clrB2 <- clrB[, keep.bacterial.taxa]
clrF2 <- clrF[, keep.fungal.taxa]

#order the data frames so they are the same
clrB2 = clrB2[rownames(clin_scales),]
clrF2 = clrF2[rownames(clin_scales),]


### Within analysis
clrBw <- withinVariation(X = clrB2, design = design2)
clrFw <- withinVariation(X = clrF2, design = design2)

### make table lists
Xa <- list( fungi=clrFw, bacteria=clrBw)
```

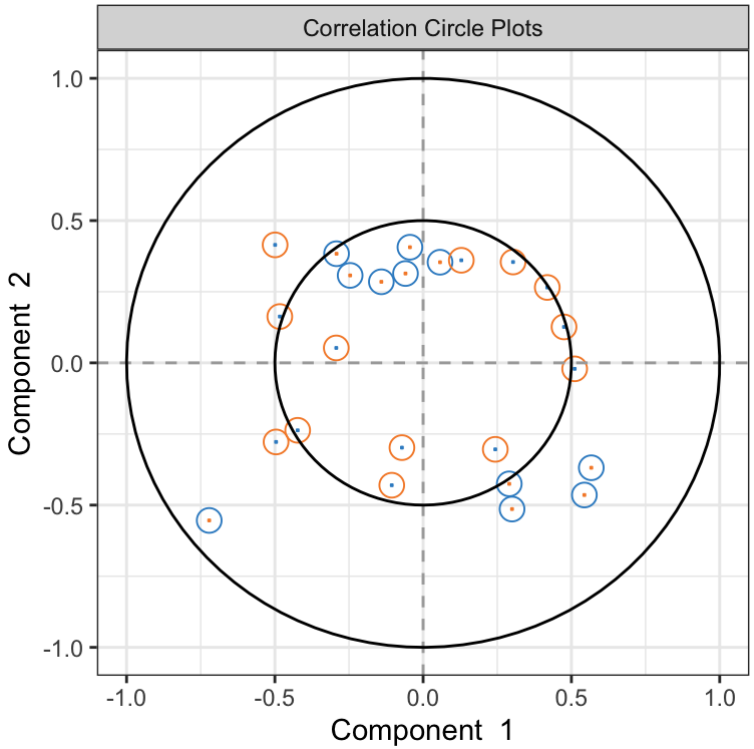
Run the first model; use the first time point; account for within site variation

```
set.seed(861)
MyResult.diablo <- block.splsda(Xa, Y1,    scale=TRUE,   near.zero.var=TRUE, scheme="centroid", ncomp = 4)
plotIndiv(MyResult.diablo,
          ind.names = FALSE,
          ellipse = TRUE,
          star=TRUE,
          legend=TRUE,
          abline = TRUE)
```

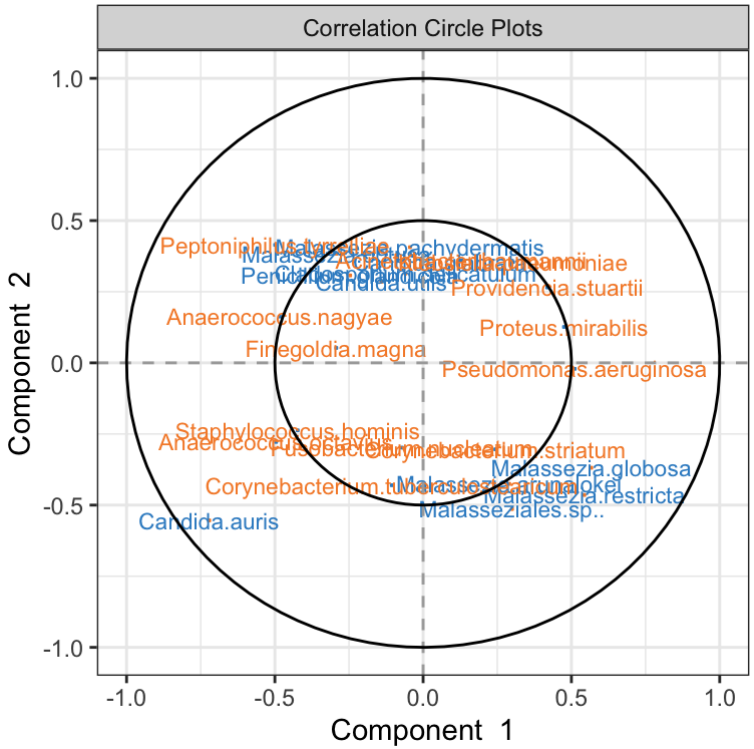


We plot the variables that segregate samples based on culture positive/negative status. Notably, we get an error from the model system indicating the direction of the correlations between variables may not be accurately represented by the correlation circle.

```
plotVar(MyResult.diablo,
      comp = c(1, 2),
      cutoff = 0.25,
      rad.in = 0.5,
      cex=c(4, 4),
      var.names = FALSE,
      style="ggplot2",
      overlap = TRUE,
      axes.box = "all",
      label.axes.box = "both")
```



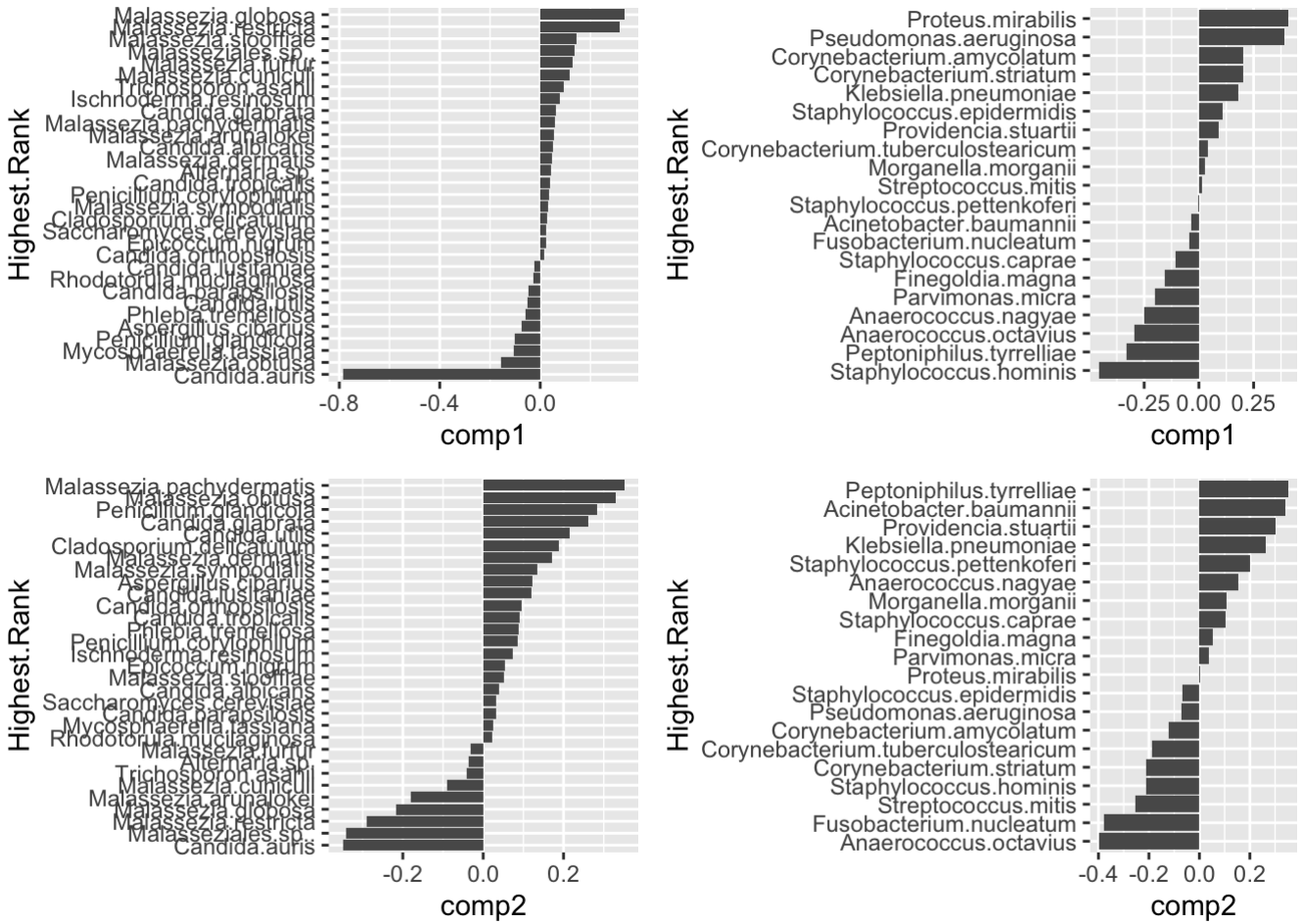

```
plotVar(MyResult.diablo,
      comp = 1:2,
      cutoff = 0.25,
      rad.in = 0.5,
      cex=c(3, 3),
      var.names = TRUE,
      style="ggplot2",
      overlap = TRUE,
      axes.box = "all",
      label.axes.box = "both")
```



Let's instead look at the variables that are driving segregation across component 1 just by looking at the variable loadings. We see that the bacterial and fungal results are similar to those identified by the single table analysis. For fungi, we see that M. globosa and M. restrica are most strongly associated with culture negativity. Candida auris, as expected, is most strongly associated with culture positivity.

```
tax1 = data.frame(bac_match@tax_table@.Data) %>%
  dplyr::select(., c("Kingdom", "Class", "Order", "Family", "Genus", "Highest.Rank"))
tax1$Highest.Rank = str_replace_all(tax1$Highest.Rank, " ", ".")
loadings.16s = data.frame(MyResult.diablo$loadings$bacteria)
loadings.16s$Highest.Rank= rownames(loadings.16s)
loadings.16s$Highest.Rank = str_replace_all(loadings.16s$Highest.Rank, " ", ".")
loadings.16s = plyr::join(loadings.16s, tax1)

loadings.16s = loadings.16s[with(loadings.16s, order(comp1)),]
ordering16s = as.vector(loadings.16s$Highest.Rank)
loadings.16s$Highest.Rank  <- factor(loadings.16s$Highest.Rank, levels = ordering16s)


tax1 = data.frame(its_match@tax_table@.Data) %>%
  dplyr::select(., c("Kingdom", "Class", "Order", "Family", "Genus", "Highest.Rank"))
tax1$Highest.Rank = str_replace_all(tax1$Highest.Rank, " ", ".")
loadings.its = data.frame(MyResult.diablo$loadings$fungi)
loadings.its$Highest.Rank= rownames(loadings.its)
loadings.its$Highest.Rank = str_replace_all(loadings.its$Highest.Rank, " ", ".")
loadings.its = plyr::join(loadings.its, tax1)
loadings.its = loadings.its[with(loadings.its, order(comp1)),]
orderingITS = as.vector(loadings.its$Highest.Rank)
loadings.its$Highest.Rank  <- factor(loadings.its$Highest.Rank, levels = orderingITS)


### Coordinate 1
p1 = ggplot() +
  geom_col(data=loadings.its, aes(Highest.Rank, comp1)) +
  coord_flip()

p2 = ggplot() +
  geom_col(data=loadings.16s, aes(Highest.Rank, comp1)) +
  coord_flip()

### Coordinate 2
loadings.16s = loadings.16s[with(loadings.16s, order(comp2)),]
ordering16s = as.vector(loadings.16s$Highest.Rank)
loadings.16s$Highest.Rank  <- factor(loadings.16s$Highest.Rank, levels = ordering16s)

loadings.its = loadings.its[with(loadings.its, order(comp2)),]
```

```
orderingITS = as.vector(loadings.its$Highest.Rank)
loadings.its$Highest.Rank  <- factor(loadings.its$Highest.Rank, levels = orderingITS)



p3 = ggplot() +
  geom_col(data=loadings.its, aes(Highest.Rank, comp2)) +
  coord_flip()

p4 = ggplot() +
  geom_col(data=loadings.16s, aes(Highest.Rank, comp2)) +
  coord_flip()

grid.arrange(p1, p2, p3, p4, ncol=2)
```



Let's look at explicit correlations

Since the multiple table model threw an error saying the direction of the correlation may not be accurately represented in the model let's just look at straight up correlations between each taxa's distribution and the distribution of Candida auris across samples. We will use a linear mixed effects model to account for multiple measures within an indidivudal

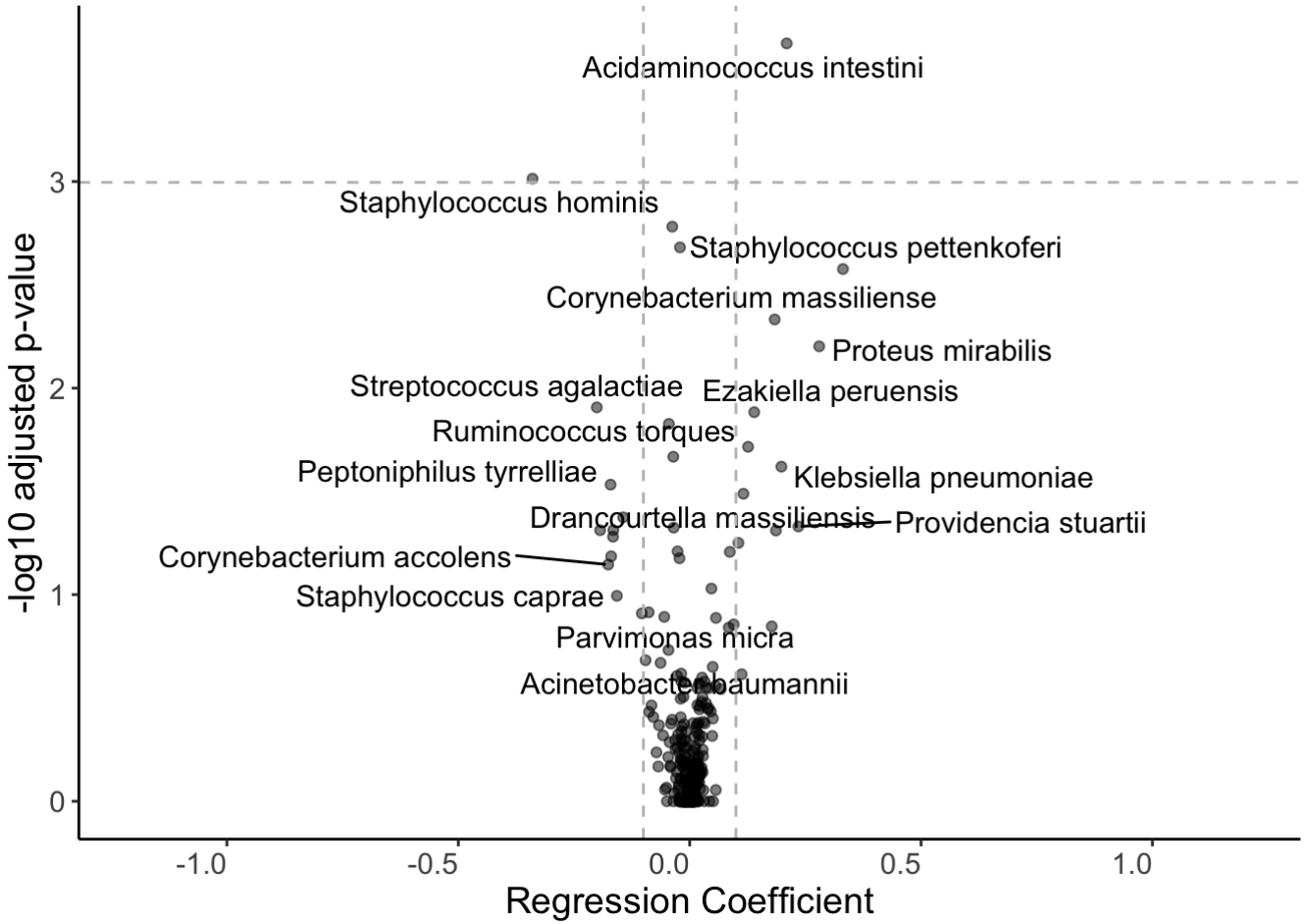
Merge the its and bacterial tables

```
#get rid of the tree - fungal
otus = otu_table(its_match)
map = sample_data(its_match)
tax = its_match@tax_table@.Data
its_match = merge_phyloseq(otus, map, tax_table(tax))


#subset the its table to eliminate noisy taxa; otherwise model fails
library(DESeq2);library(genefilter)
filtergroup = genefilter::filterfun(genefilter::kOverA(k=20, A=10)) #k = number of samples; A = abundance
#filter  taxa
        filtPhy = filter_taxa(its_match, filtergroup, prune=TRUE)
        filtPhy = prune_taxa(taxa_sums(filtPhy) > 0, filtPhy)
        filt_its = subset_samples(filtPhy, Unique_ptid != 32)

#generate a combined fungal/bacterial table so we can adjust pvalues in the regression appropriately
phy = merge_phyloseq(bac_match, filt_its)

#transform to centered log ratio
set.seed(78927)
phy = transform_sample_counts(phy, function(x) compositions::clr(x))

#make a map for the regression
map = data.frame(sample_data(phy)) %>%
  dplyr::select(., c("sqrt_cauris", "Unique_ptid", "SiteID"))

#convert site and survey period to numeric
map$SiteID = as.numeric(factor(map$SiteID))
map = data.frame(scale(map))

#get the otu table of the centered log ratio table
otus = data.frame(otu_table(phy))

attach(map)
all=data.frame(cbind(otus, map))

#set up empty lists
mod <- list()
out <- list()
adjp <- list()
```

```
#https://stackoverflow.com/questions/57590176/adjust-p-values-obtained-with-lmertestlmer-for-multiple-comparisons
adjMC <- function( model_name ) {
  model_glht <- glht(model_name)
  model_MCadj <- summary(model_glht, test = adjusted('holm')) # Bonferroni-Holm
  return(model_MCadj)
}

library(multcomp)
for(i in names(otus)[-1]){
  mod[[i]] <- lmerTest::lmer(get(i) ~ sqrt_cauris   + SiteID +
                               (1 | Unique_ptid  ),
      data = all)
  adjp[[i]]  = adjMC(mod[[i]])
  out[[i]] = broom.mixed::tidy(adjp[[i]], conf.int = TRUE, .name_repair = "unique")
}
tax = data.frame(phy@tax_table@.Data)
out = out %>% map_dfr(~ .x %>% as_tibble(), .id = "Highest.Rank")
out$Highest.Rank = str_replace_all(out$Highest.Rank, "([.])", " ")

df = data.frame(out) %>%
  plyr::join(tax) %>%
  subset(., contrast=="sqrt_cauris")

#make a volcano plot
library(RColorBrewer)
pal <- brewer.pal(n = 4, name = 'Set1')
df$adj.p.value = as.numeric(as.character(df$adj.p.value))
my.annotation = subset(df, adj.p.value  < 0.05 & estimate > 0.1 | estimate <= -0.1)
anno2 = subset(df, Highest.Rank %in% c("Acinetobacter baumannii", "Pseudomonas aeruginosa"))
my.annotation = data.frame(rbind(my.annotation, anno2))
df$Highest.Rank = as.character(as.factor(df$Highest.Rank))
df$neglog = -log10(df$adj.p.value)
df$neglog  = ifelse(df$neglog=="Inf", 12, df$neglog)

## Create a column to indicate which genes to label
df$species.label = ifelse(df$Highest.Rank %in% my.annotation$Highest.Rank, "TRUE", "FALSE")
myannotations = subset(df, species.label==TRUE)
```

let's see a volcano plot of taxa whose distributions are significantly associated with Candida auris abundance. We see that the taxa identified by this model are similar to those identified by Mixomics.

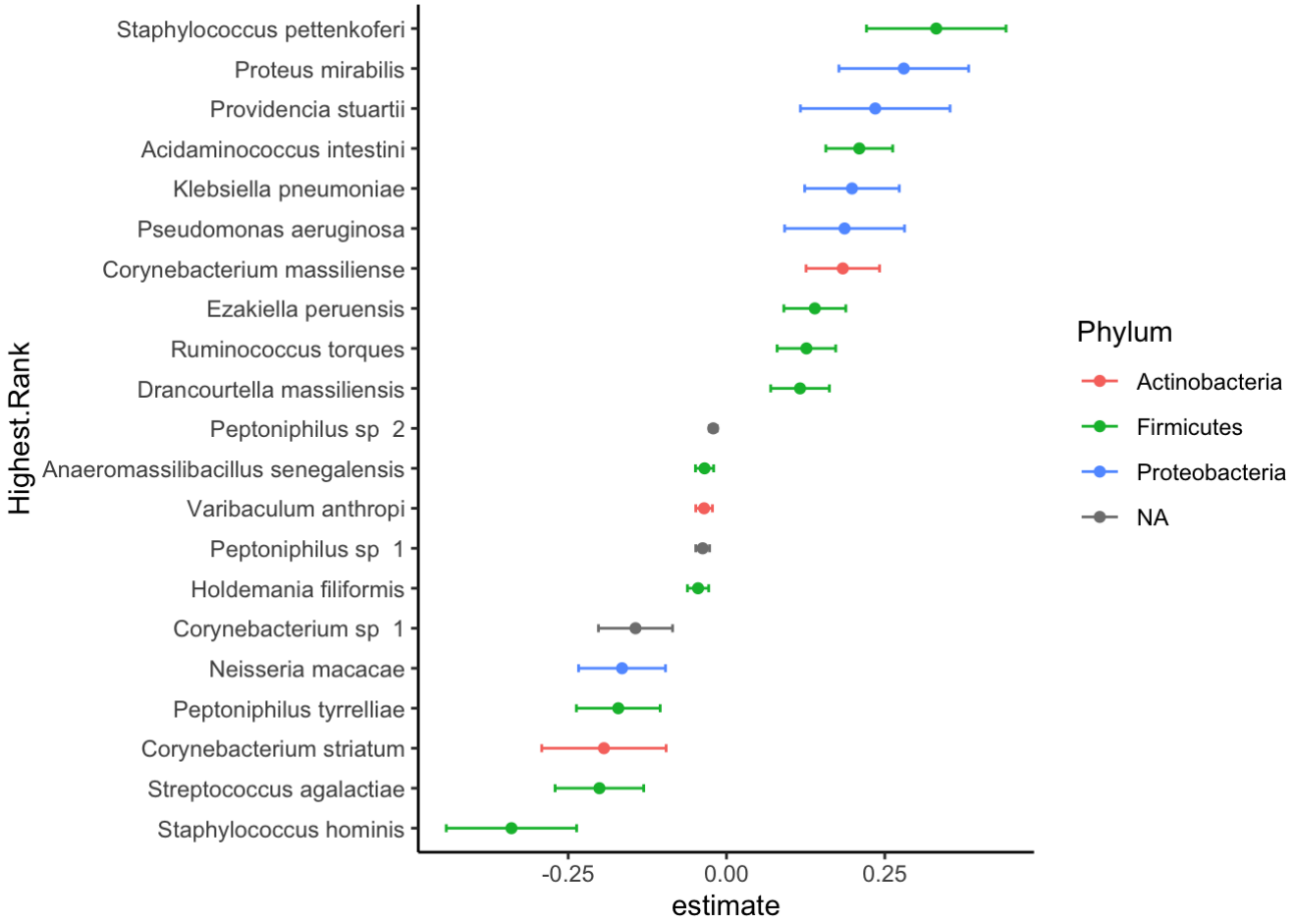
```
Figure3C = ggplot(df) +
  geom_point(aes(x = estimate, y = neglog), alpha=0.5) +
  geom_text_repel(data=myannotations, aes(x = estimate, y = neglog, label = Highest.Rank),size=4) +
  xlab("Regression Coefficient") +
  ylab("-log10 adjusted p-value")  +
  xlim(-1.2, 1.2)  +
  geom_vline(xintercept = -0.1, linetype='dashed', color="gray") +
  geom_vline(xintercept = 0.1, linetype='dashed', color="gray") +
  geom_hline(yintercept = -log(0.05),  linetype='dashed', color="gray") +
  theme_classic() +
  scale_color_manual(values=c("#377EB8" ,"#4DAF4A"))+
    theme(text = element_text(size=14),
        axis.text.x = element_text(angle=0, hjust=1)) + theme(legend.position = "none")

Figure3C
```



print the table

```
myannotations = subset(df, adj.p.value  < 0.05 )
myannotations = myannotations[with(myannotations, order(estimate)),]

myannotations$upper = myannotations$std.error + myannotations$estimate
myannotations$lower = myannotations$estimate- myannotations$std.error
ordering = as.vector(myannotations$Highest.Rank)
myannotations$Highest.Rank  <- factor(myannotations$Highest.Rank, levels = ordering)
p = ggplot(myannotations, aes(Highest.Rank, estimate, color=Phylum)) + geom_point() +
        geom_errorbar(aes(ymin=upper, ymax=lower), width=.2,
                 position=position_dodge(.9)) + coord_flip() +
  theme_classic()

p
```

Figure 3: Underlying skin microbiome (fungal and bacterial communities) integrated with C. auris colonization status



Supplementary Data 5

Figure 4: fungal communities, dominated by Malassezia and Candida species, of surveilled body sites, have differential stability, resilience and likelihood of invasion by Candida auris.

Diana Proctor

05/12/2020

last updated: April 22 2021

draft: submission to Nature Medicine October 2020, Revised submission Feb 2021

Manuscript Title: Integrated genomic, epidemiologic investigation of Candida auris skin colonization in a skilled nursing facility

Authors: Diana M Proctor1, Thelma Dangana2, D. Joseph Sexton3, Christine Fukuda2, Rachel D Yelin2, Mary Stanley2, Pamela B Bell2, Sangeetha Baskaran2, Clay Deming1, Qiong Chen1, Sean Conlan1, Morgan Park4, NISC Comparative Sequencing Program4, Rory M Welsh3, Snigdha Vallabhaneni3,5, Tom Chiller3, Kaitlin Forsberg3, Stephanie R. Black6, Massimo Pacilli6, Heidi H Kong7, Michael Y Lin2, Michael E Schoeny8, Anastasia P Litvintseva3, Julia A Segre1+, Mary K Hayden2+

Affiliations:

- 1Microbial Genomics Section, Translational and Functional Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA
- 2 Department of Internal Medicine, Division of Infectious Diseases, Rush University Medical Center, Chicago, IL 60612, USA.
- 3Mycotic Diseases Branch, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA.
- 4NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA.
- 5Division of Healthcare Quality Promotion, NCEZID, CDC, USA
- 6Dermatology Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD 20892, USA
- 7Department of Nursing, Rush University Medical Center, Chicago, , IL 60612, USA.
- +Contributed equally

Description of the dataset

Here, we examine the structure and dynamics of the mycobiome, using the ITS1 data set. We filter to remove ASVs that have fewer than 10 reads, as well as species that cannot be assigned to a taxonomic rank beyond the division level. The entire data object has 134 species distributed across 1115 samples.

For the CST specific analysis, we subset on the sites that have the highest sequencing depth, as well as the most paired samples for individual subjects. A total of 17 subjects have T1 and T2 samples for the sites: c("Fg", "Tw", "N", "Ic")).

To accomplish this, we read in the following data:

1. its_match.rds

Here, we render the following figures:

1. Figure 4
2. Supplementary Figure 7
3. Supplementary Figure 8

Load libraries and set plotting preferences

```
library(knitr)
opts_chunk$set(fig.width=12, fig.height=8,
               echo=TRUE, warning=FALSE, message=FALSE, error = FALSE)
```

```
library(phyloseq);
library(ggplot2);
library(magrittr);
library(reshape2);
library(gridExtra);
library(viridis);
library(gridExtra);
library(stringr);
library(cowplot);
library(ggpubr);
library(ggalt);
library(vegan);
library(scales)
library(markovchain)
library(knitr)
library(gridGraphics)
library(wesanderson)

#set theme
theme_set(theme_bw())


#cluster analysis, takes two arguments
#x: is the distance matrix
#k: is the number of clusters
mypam = function(x, k) {
        pam(x, k, diss=TRUE, stand=FALSE)
}
```

read in the data set

```
phy = readRDS(file="~/Desktop/candida_auris_rush/its_match_CDI-out.rds") %>%
      subset_taxa(., Highest.Rank != "Less_than_10_per_ASV Less_than_10_per_ASV") %>%
      subset_taxa(., Highest.Rank != "Fungal sp.")

tax = data.frame(phy@tax_table@.Data)
taxa_names(phy) = tax$Seq
write.csv(tax, "~/Desktop/candida_auris_rush/its_match_CDI-out-tax_uncorrected.csv")

tax = read.csv("~/Desktop/candida_auris_rush/its_match_CDI-out-tax_corrected.csv")
rownames(tax) = tax$X
tax = as.matrix(tax) %>%
  tax_table(.)
tax_table(phy) = tax

#update taxa names to highest rank
tax = data.frame(phy@tax_table@.Data)
new.names = tax$Highest.Rank
taxa_names(phy) = new.names
```

Create a data subset for subjects who have matched samples for the following 4 sites:

- identify subjects who have 4 samples at t1 and t2 with: c("Fg", "Tw", "N", "Ic"))
- 17 subjects and if we require An then this drops
- minimum number of samples is 459, median 22,294 and Max is 278K.

```
### get the list of subjects who have paired samples at t1 and t2 with: c("Fg", "Tw", "N", "Ic"))
test = phy %>%
      subset_taxa(., Highest.Rank != "Less_than_10_per_ASV Less_than_10_per_ASV") %>%
      subset_taxa(., Highest.Rank != "Fungal sp.") %>%
      subset_samples(., SiteID %in% c("Fg", "Tw", "N", "Ic")) %>%
      sample_data(.) %>%
      subset(., Survey_Period %in% c(1, 2)) %>%
      as.data.frame(.)

#get the list of subjects
counts = data.frame(table(test$Unique_ptid)) %>%
  subset(., Freq ==8) %>%
  dplyr::select(., Var1)


#subset the phyloseq object on these subjects
subjects2keep = as.vector(counts$Var1)
time  =  subset_samples(phy, Unique_ptid %in% subjects2keep)


#initial analysis was done with all genera
limphy = phy %>%
      subset_taxa(.,  Genus %in% c("g__Malassezia", "g__Candida")) %>%
      tax_glom(., "Species") %>%
      transform_sample_counts(., function(x) x/sum(x))  %>%
      subset_samples(., SiteID %in% c("Fg", "Tw", "N", "Ic"))
```

Figure 4A: Principal coordinates analysis of the Weighted Unifrac metric of the fungal community at each body site (Tw, Fg, Ic, N).

Samples are shaded according to cluster identity, as revealed by partition around medoids analysis. Cluster 1 tends to be dominated by Malassezia restricta (N=252, 52.2%), Cluster 2 by Malassezia slooffiae (N=60, 12.4%), Cluster 3 by divers Candida species (N=61, 12.6%), and Cluster 4 by Candida auris (N=110, 22.8%). Community state types (CST) are identified by the gap statistic computed on a partition around medoids analysis. Segregation of Malassezia and Candida species across the first axis explains 62% of the variance. Candida species segregate across the second major axis, which accounts for ~15% of the variance.

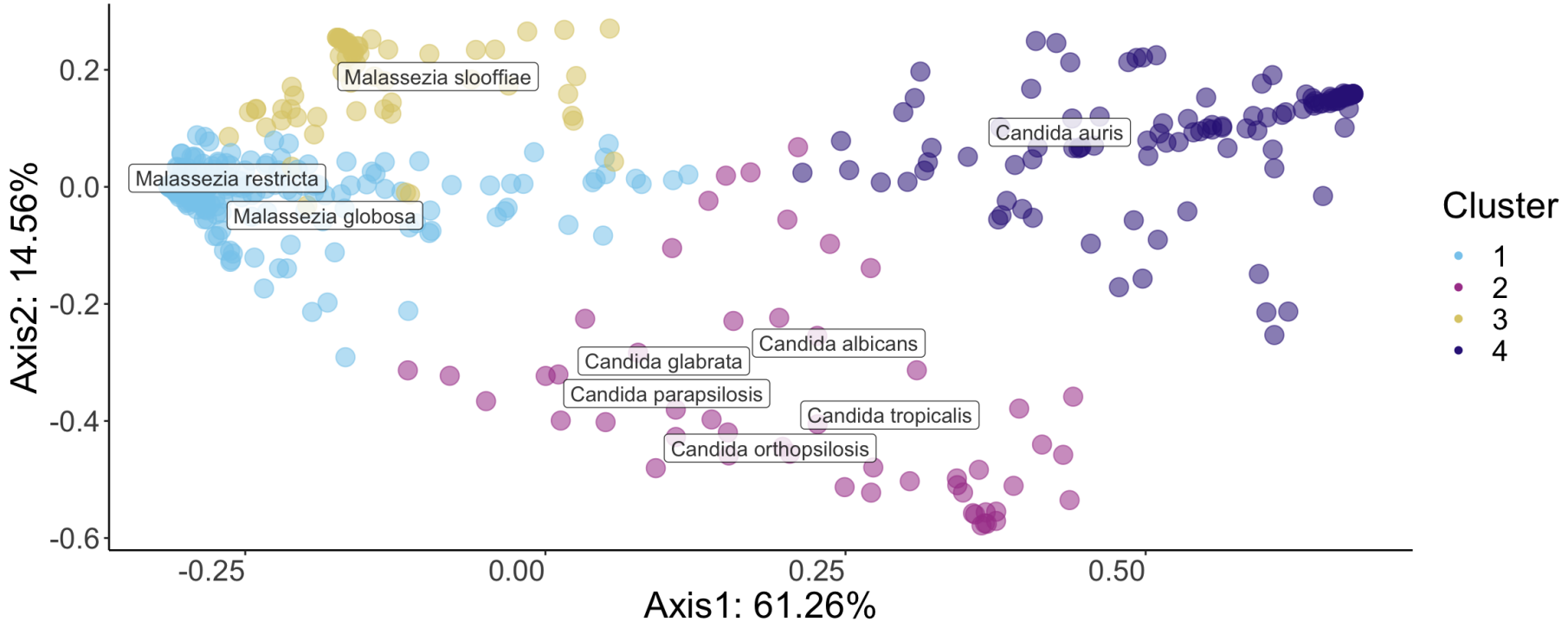
```
#plot a pcoa of the correlation matrix
set.seed(987)
wuf =  phyloseq::distance(limphy, method="wunifrac")
wuf  = as.matrix(wuf)

library(RColorBrewer)
k=4
CSTColors <- brewer.pal(5,"Paired")[2:5]
names(CSTColors) <- 1:k
pm <- cluster::pam(wuf, diss=TRUE, k=k)

matPCOA <- ape::pcoa(wuf)
pcoa.df <- data.frame(pm$clustering, matPCOA$vectors, sample_data(limphy))
colnames(pcoa.df)[1] <- "Cluster"
pcoa.df$Cluster <- as.factor(pcoa.df$Cluster)

pcoa.df$SiteID = plyr::revalue(pcoa.df$SiteID,
                        c("An" = "Peri-anal",
                          "Ax"= "Axilla",
                          "Bu/To" = "Buccal Mucosa / Tongue",
                          "Ea" = "Ear Canal",
                          "Fg" = "Fingertips",
                          "Ic" = "Inguinal crease",
                          "N" = "Anterior nares",
                          "Ne" = "Neck",
                          "Tc" = "Tracheostomy site",
                          "Tw" = "Toeweb"))

colorBlindGrey8   <- c("#999999", "#E69F00", "#56B4E9", "#009E73",
                       "#F0E442", "#0072B2", "#D55E00", "#CC79A7")
scales::show_col(colorBlindGrey8)
```


```
#how much of the variance does each axis explain?
eig <- matPCOA$values
ExplainedVariance = 100*(eig$Eigenvalues/sum(eig$Eigenvalues))
ExplainedVariance[1:10]
```

```
##  [1] 61.2590061 14.5612544 13.7641462  7.0787404  2.1316565  1.5166844
##  [7]  1.1882370  1.1058587  1.0261814  0.9010484
```

```
replacement.map <- data.frame(pm$clustering,sample_data(limphy))
colnames(replacement.map)[1] <- "Cluster"

###### plot the pcoa coloring samples by the cluster number vs. sites
replacement.map$Cluster <- as.factor(replacement.map$Cluster)
replacement.map = sample_data(replacement.map)
sample_data(limphy) = replacement.map
sample_data(limphy) = replacement.map
sample_data(phy) = replacement.map

#### unifrac PCoA
ord = transform_sample_counts(limphy, function(x) x/sum(x)) %>%
      ordinate(.,  method="PCoA", distance="wunifrac")
evals = 100*(ord$values$Eigenvalues/sum(ord$values$Eigenvalues))

### Plot ordination
p = plot_ordination(limphy, ord, type="biplot", color="Cluster")

myannotations = subset(p$data, Highest.Rank %in%
                         c("Malassezia globosa", "Malassezia restricta",
                         "Malassezia slooffiae", "Candida tropicalis", "Candida parapsilosis",
                         "Candida albicans", "Candida orthopsilosis",    "Candida glabrata",
                         "Candida auris"  ))

safe_colorblind_palette <- c("#88CCEE",  "#AA4499","#DDCC77",  "#332288",
                             "#44AA99", "#999933", "#882255", "#661100", "#6699CC", "#888888")

Figure4a = ggplot()   +
      theme_classic()  +
      geom_point(data=subset(p$data, Cluster != "Taxa"),
                 aes(x=Axis.1, y=Axis.2, size=4, alpha=0.8, color=Cluster)) +
      scale_color_manual(values=safe_colorblind_palette) +
      ggrepel::geom_label_repel(data=myannotations,
                                mapping = aes(x=Axis.1, y=Axis.2, label = Highest.Rank), size = 4,
                                box.padding=0.05, alpha=0.8, color="black") +
      coord_fixed(sqrt(evals[2] / evals[1]))  +
      xlab(paste0("Axis1: ", round(evals[[1]], 2), "%"))+
      ylab(paste0("Axis2: ", round(evals[[2]], 2), "%"))+
      theme(plot.title = element_text(size=20),
      axis.title.x = element_text(size=20),
```

```
      axis.title.y = element_text(size=20),
      text = element_text(size=20), axis.text.x = element_text(angle=0, hjust=1))  +
  guides(shape=FALSE, alpha=FALSE, size=FALSE)

Figure4a
```



```
ggsave(Figure4a, file="~/Desktop/Figure4a.pdf", device="pdf", width = 8, height = 5)
```

Figure 4B: Heatmaps of the relative abundance of the top 20 species in each sample, clustered by CST.

B. Heatmaps of the relative abundance of the top 20 species in each sample, clustered by CST. Shading is based on the relative abundance of taxa within each sample.

```
keep = names(sort(taxa_sums(phy), TRUE)[1:20])

top20=prune_taxa(keep, phy) %>%
    prune_samples(sample_sums(.)>0, .)


taxa_names(top20) = data.frame(top20@tax_table@.Data)$Highest.Rank

italicsy <- element_text(face = "italic", size = 20)



#set up a loop to get OTU abundances across sites within cluters
clusts <- 1:k
clust.list <- vector("list", length(clusts))
names(clust.list) <- clusts
clust.phys <- vector("list", length(clusts))
names(clust.phys) <- clusts
taxa.order <- names(sort(taxa_sums(phy)))
plot_heatmap_by_cluster <- function(clustDf, phy, clustN, label) {
        clustn = subset(clustDf, Cluster==clustN)
        CS1 = as.vector(rownames(clustn))
        CS1.phys <- prune_samples(CS1, phy) %>%
                  prune_taxa(taxa_sums(.) > 0 , .) %>%
                  prune_samples(sample_sums(.) > 0 , .) %>%
                  transform_sample_counts(., function(x) 100*round(x/sum(x), 2))
        p  =  plot_heatmap(CS1.phys, taxa.label="Highest.Rank", taxa.order=taxa.order) +
              ggtitle(paste("CST ", clustN, ": ", label)) +
              theme(axis.text.x=element_blank(),
              axis.ticks.x=element_blank(),
              axis.title.y = element_text(color="black", size=16, face="bold"),
              text = element_text(size=12), axis.text.y = element_text(angle=0, hjust=1)) +
              ylab("") + theme(legend.position="none") + xlab("")
}


c1 = plot_heatmap_by_cluster(clustDf=pcoa.df,
                             phy=top20, clustN=1, label="Malassezia diverse")  +
   theme( axis.text.y = italicsy)
c2 = plot_heatmap_by_cluster(clustDf=pcoa.df,
```

```
                              phy=top20, clustN=2, label="Candida diverse   ") +
        theme(axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank())
c3 = plot_heatmap_by_cluster(clustDf=pcoa.df,
                              phy=top20, clustN=3, label="Malassezia sloofiae") +
        theme(axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank())

c4 = plot_heatmap_by_cluster(clustDf=pcoa.df,
                              phy=top20, clustN=4, label="Candida auris    ") +
        theme(axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank())

plot_legend <- function(clustDf, phy, clustN, label) {
        clustn = subset(clustDf, Cluster==clustN)
        CS1 = as.vector(rownames(clustn))
        CS1.phys <- prune_samples(CS1, phy) %>%
                  prune_taxa(taxa_sums(.) > 0 , .) %>%
                  prune_samples(sample_sums(.) > 0 , .) %>%
                  transform_sample_counts(., function(x) 100*round(x/sum(x), 2))
        p  =  plot_heatmap(CS1.phys, taxa.label="Highest.Rank", taxa.order=taxa.order) +
              ggtitle(paste("CST ", clustN, ": ", label)) +
              theme(axis.text.x=element_blank(),
              axis.ticks.x=element_blank(),
              axis.title.y = element_text(color="black", size=16, face="bold"),
              text = element_text(size=12), axis.text.y = element_text(angle=0, hjust=1)) +
              ylab("") + xlab("")
}

c5 = plot_legend(clustDf=pcoa.df, phy=top20, clustN=2, label="Candida auris    ") +
        theme(axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank())

fig4blegend <- cowplot::get_legend(c5)

Figure4B <- ggarrange((c1 + theme(legend.position = "none")),
                      (c2 + theme(legend.position = "none")),
```

```
                        (c3 + theme(legend.position = "none")),
                        (c4 + theme(legend.position = "none")), ncol=4, widths = c(1.75, 1, 1, 1))




Figure4b <- cowplot::plot_grid(Figure4B, fig4blegend, nrow = 1, rel_widths = c(1,.1))
ggsave(Figure4b, file="~/Desktop/proctor_manuscript/Figure4/Figure4b.png",  width = 12, height = 8, units ="in",d
pi = 300, device = "png")
ggsave(Figure4b, file="~/Desktop/proctor_manuscript/Figure4/Figure4b.pdf",  width = 14, height = 8, units ="in",d
pi = 300, device = "pdf")
ggsave(Figure4b, file="~/Desktop/proctor_manuscript/Figure4/Figure4b.png",  width = 18, height = 8, units ="in",d
pi = 300, device = "png")

Figure4b
```

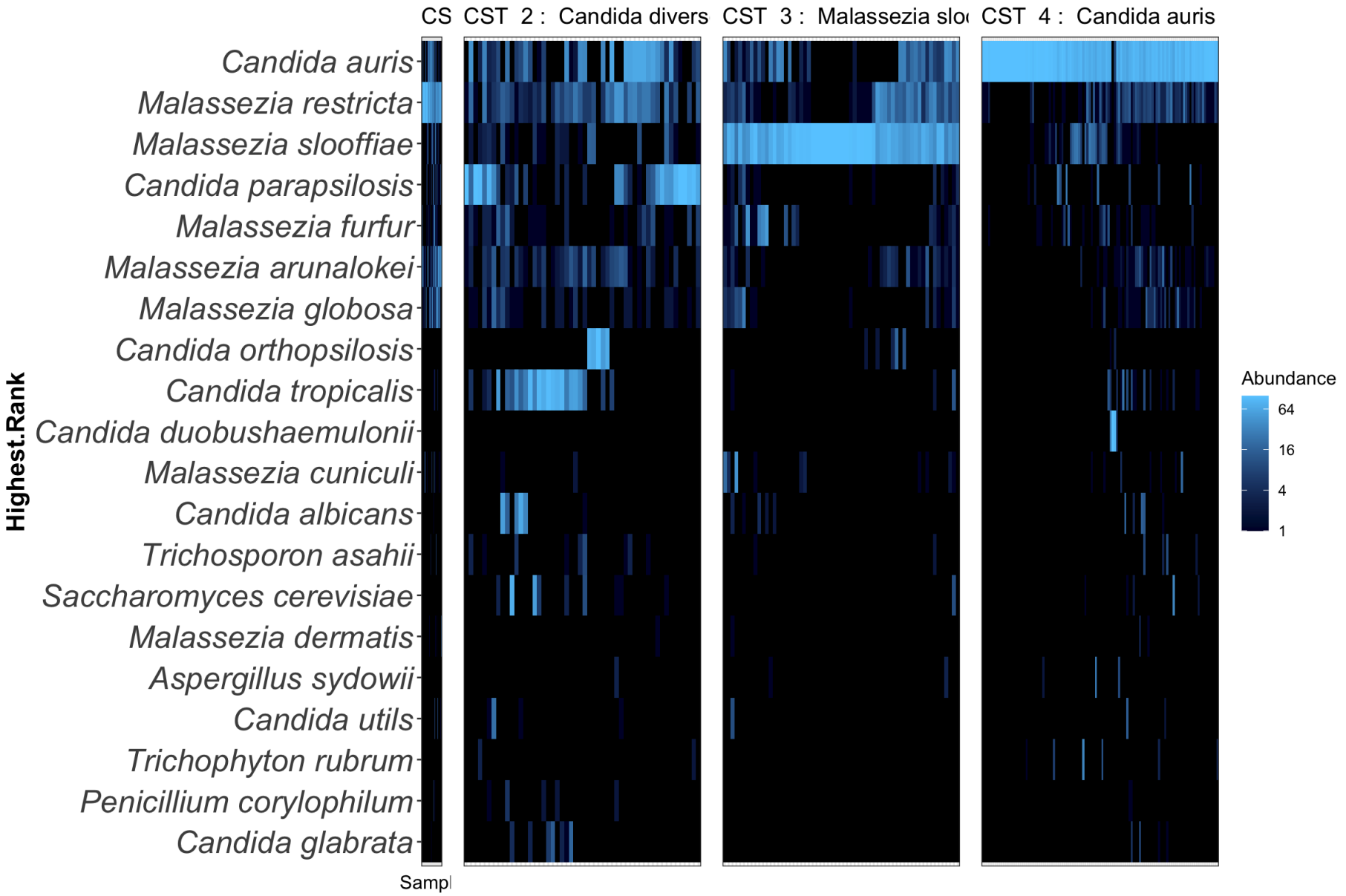


Figure 4C

C. Self and inter-state transition probabilities inferred for samples of the toeweb, fingertips, inguinal crease, and nares.

```
sample_data(limphy)$CST = paste0("CST", sample_data(limphy)$Cluster)


t1p = subset_samples(limphy, Survey_Period !="3") %>%
  subset_samples(., Unique_ptid %in% subjects2keep)
counts = data.frame(table(sample_data(t1p)$Unique_ptid))
colnames(counts) = c("Unique_ptid", "NSamplesPerSubject")
sample_data(t1p)$NSamplesPerSubject = 8

t1 = subset_samples(t1p, Survey_Period==1)
t2 = subset_samples(t1p, Survey_Period==2)
states_by_subject_t1 = table(sample_data(t1)$Cluster, sample_data(t1)$Unique_ptid)
states_by_subject_t2 = table(sample_data(t2)$Cluster, sample_data(t2)$Unique_ptid)


#we'll look here at what samples are in each cst
p = data.frame(sample_data(phy)) %>%
    ggplot(., aes(as.factor(Survey_Period), SiteID, color=Cluster)) + geom_point() +
            theme_bw()+ facet_wrap(~Unique_ptid)+
            scale_colour_manual(name = "CST", values = CSTColors[1:5])

#p

#let's get the counts of samples within each cst
tab = table(p$data$Site.Extended, p$data$Cluster)
Coltots = 100*(colSums(tab) / sum(colSums(tab)))
tots = colSums(tab)
```

Calculate transition probabilities and initial states for the complete dataset

```
###generate CSTS
set.seed(87)
k=4
CSTs = c("CST1", "CST2", "CST3", "CST4")

df  = data.frame(sample_data(t1p))  %>%
  dplyr::select(Unique_ptid, Cluster,  Survey_Period, SiteID) %>%
  dcast(., Unique_ptid+SiteID~Survey_Period,value.var="Cluster")


colnames(df)[3:4] = c("t1", "t2")
df = dplyr::select(df, c("t1", "t2"))

#save this df to dfSummary
dfSummary = df
tab = table(df$t1, df$t2) # t1=row, t2=columns
tab = as.matrix(tab)
all.trans <- matrix(tab, nrow=k, ncol=k)
all.trans =  all.trans/rowSums(all.trans)  # Normalize row sums to 1; divide by prior state
all.trans[is.nan(all.trans)] = 0

all.init = table(df$`t1`)
all.init = c(all.init) #total 92
all.two = table(df$`t2`) #total 92
```

print the probability transition matrix

```
#print(all.trans)
```

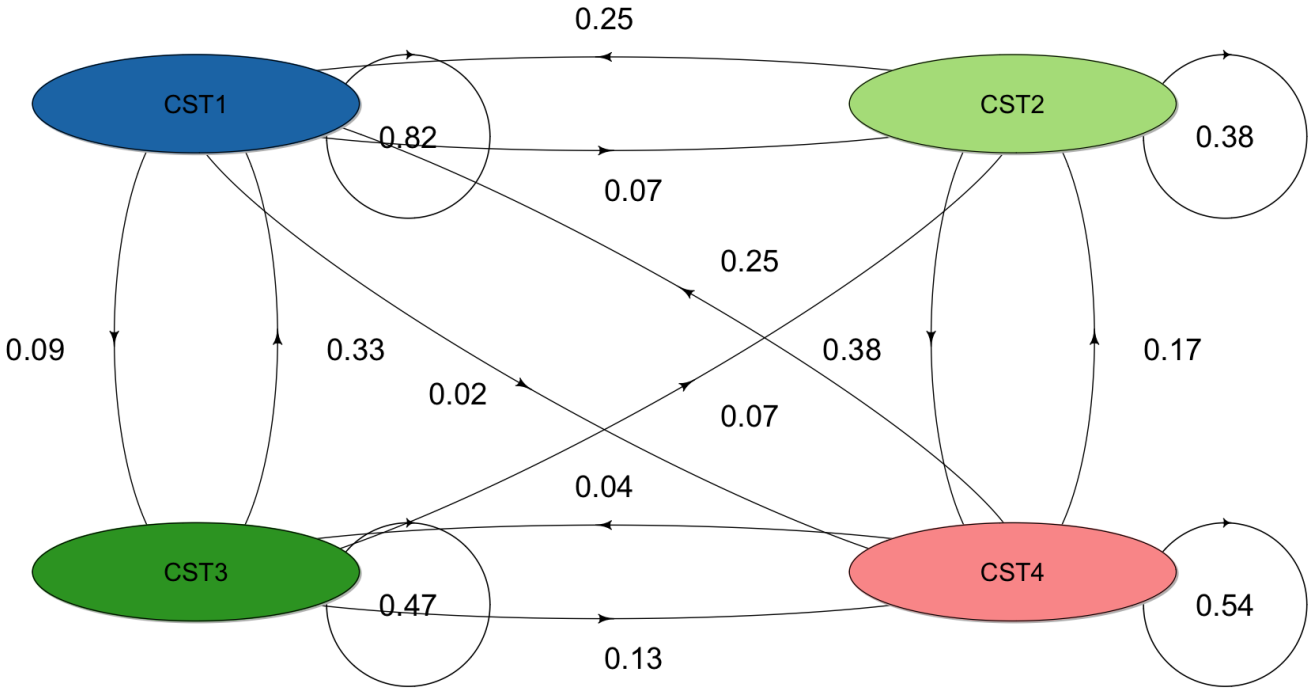
print the initial matrix, the number of samples in each state from t1 to t2

- 92 total

```
#all.init
```

Figure 4C: Self and inter-state transition probabilities inferred for samples of the toeweb, fingertips, inguinal crease, and nares.

```
rownames(all.trans)=CSTs
colnames(all.trans)=CSTs

# Make Markov chain object
set.seed(789)
mcAll <- new("markovchain", states = CSTs, byrow = TRUE,
             transitionMatrix = all.trans, name = "All Sites")


plot_transition_matrix <- function(M, site){
    RoundedM = apply(M, 1, function(x) round(x, 2))
    diagram::plotmat(RoundedM , pos=c(2, 2),
        lwd = 0.7, box.lwd = 0.6,
        shadow.size = 0.001, shadow.col = "grey", dr = 0.008,
        name=CSTs,
        cex.txt = 1.2,
        box.size = 0.1,
        box.type = "circle",
        box.prop = 0.3,
        box.col = CSTColors,
        arr.length=.2,
        arr.width=0.1,
        self.lwd = .8,
        self.cex = .5,
        self.shifty = -.02,
        self.shiftx = .13,
        main = paste0(site), cex=3)

}

plot_transition_matrix(M=all.trans, site="")
```



save the plot to eps

```
setEPS()
postscript("~/Desktop/proctor_manuscript/Figure4/Figure4c.eps")
plot_transition_matrix(M=all.trans, site="")
dev.off()
```

```
## quartz_off_screen
##                 2
```

Figure 4D: Scatterplot of the predicted numbers of samples in each CST at 3, 6, and 12 months post sample collection compared to the actual number of samples at 3 months.

Predictions were generated using the Markov chain in panel c.

Note: we have 77 samples in the third time point, so we've lost some samples. We'll normalize the projected and actual values to account for this disparity to see how well the model projects the number of samples in each cst.

```
# the predicted number of subjects in each state at time point 1
after2Months <-   all.init * (t(mcAll) * t(mcAll))
after3Months <-   all.init *  (mcAll ^ 3)
after3Months
```

```
##          CST1     CST2     CST3     CST4
## [1,] 54.00826 11.36882 10.44705 16.17587
```

```
after7Months <-   all.init *  (mcAll ^ 7)
after12Months <-   all.init *  (mcAll ^ 12)
after12Months
```

```
##          CST1     CST2     CST3     CST4
## [1,] 55.78669 11.00471 10.44714 14.76146
```

```
after24Months <-   all.init *  (mcAll ^ 24)
after24Months
```

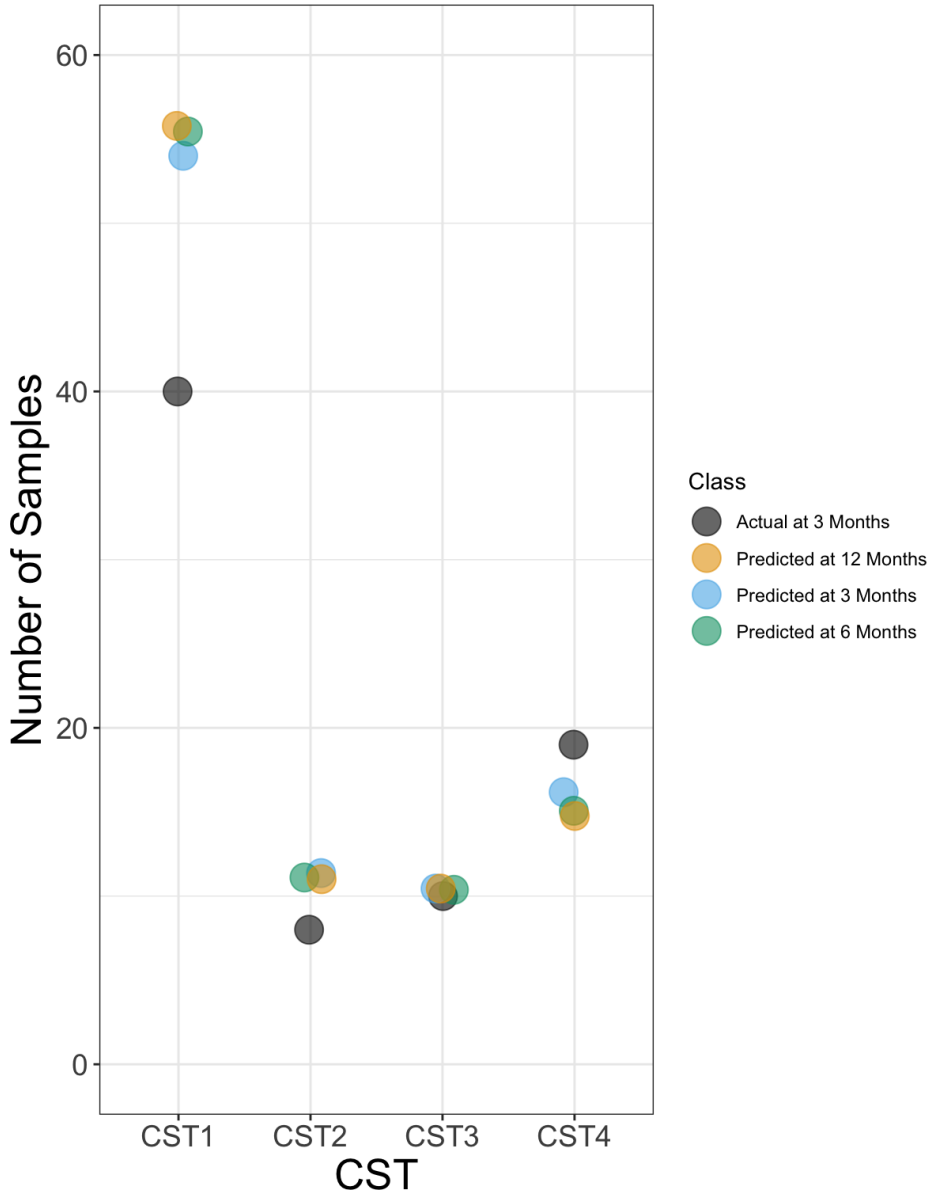
```
##          CST1    CST2    CST3     CST4
## [1,] 55.80238 10.9994 10.4524 14.74581
```

```
after36Months <-   all.init *  (mcAll ^ 36)
after36Months
```

```
##          CST1     CST2     CST3     CST4
## [1,] 55.80242 10.99939 10.45242 14.74578
```

The predictions are off because we have 77 samples at time point 3 for these sites, but we have 92 samples for the projections

```
#how many samples are there in each state at time 3
#drop the sites that are missing alot of samples across subjects

#drop specific subjects
t3 = subset_samples(limphy, Unique_ptid %in% subjects2keep)
df = data.frame(sample_data(t3))
t3 = subset(df, Survey_Period=="3")
actual =  table(t3$CST)



# the predicted number of subjects in each state at time point 1
after3Months <-   all.init *  (mcAll ^ 3)
after6Months <-   all.init *  (mcAll ^ 6)
after12Months <-   all.init *  (mcAll ^ 12)

df = data.frame(rbind(actual, after3Months,  after6Months, after12Months))
rownames(df) = c("Actual at 3 Months", "Predicted at 3 Months",
                 "Predicted at 6 Months" ,"Predicted at 12 Months")

df$Class = rownames(df)
dfm = melt(df, id.vars = "Class")

cbbPalette <- c("#000000", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")

df$Class = rownames(df)
dfm = melt(df, id.vars = "Class")

#figure 4d
Figure4d = ggplot(dfm, aes(variable, value, color=Class)) +
  geom_jitter(size=6, alpha=0.6, width = 0.09) +
  scale_color_manual(values=cbbPalette)+ theme(axis.text = element_text(size = 14))  +
  ylab("Number of Samples") + xlab("CST") + theme(axis.title = element_text(size = 20)) + ylim(0, 60)
Figure4d+
      theme(aspect.ratio = 2/1)
```



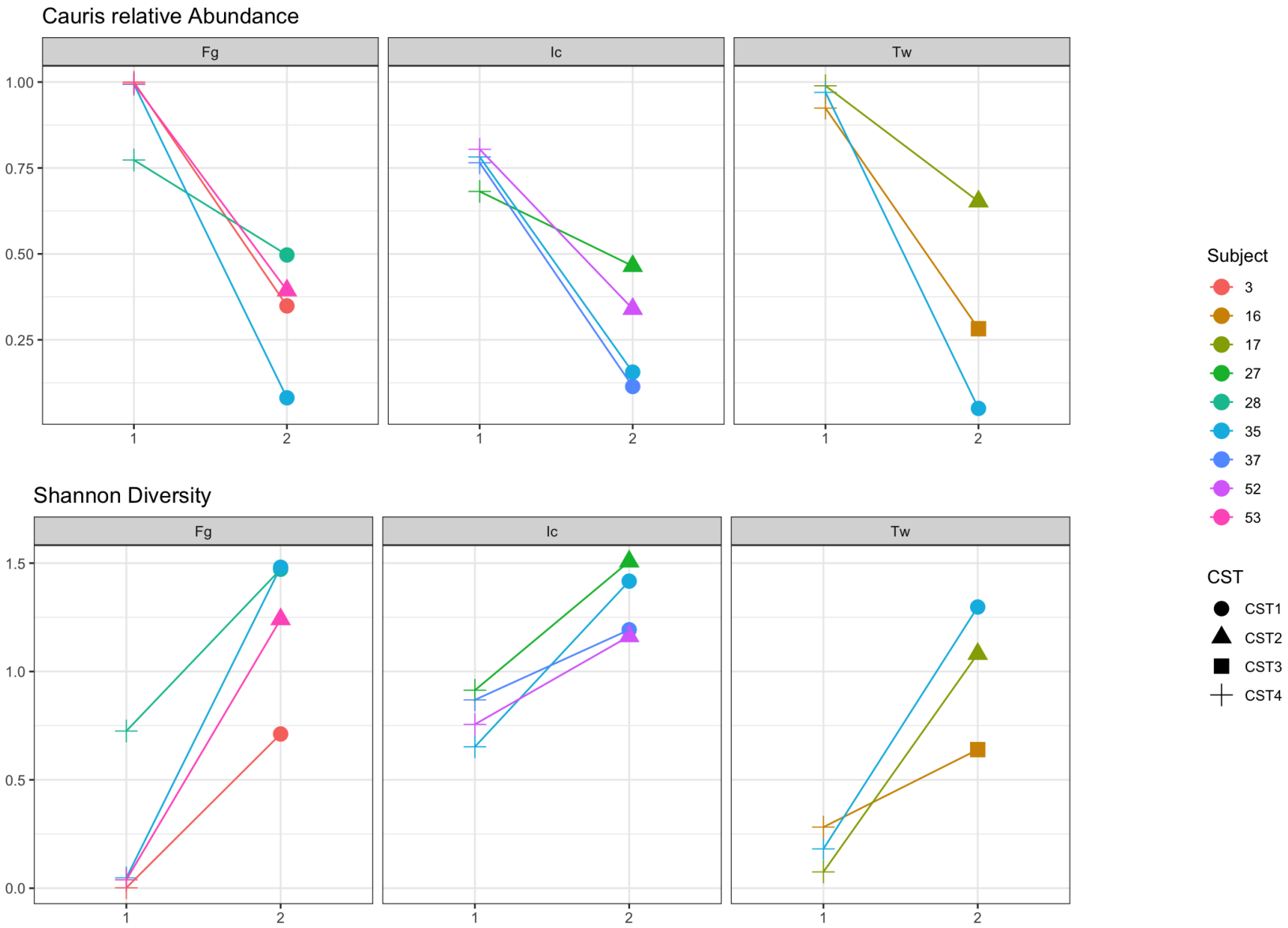

```
ggsave(Figure4d, file="~/Desktop/proctor_manuscript/Figure4/Figure4d.pdf", device="pdf", width = 5, height = 5)
```

Supplementary Figure 8: Look at transitions from CST1

Supplementary Figure 8: Examination of the mycobiome at sites that transition away from Candida auris domination between survey 1 and survey 2. Colors correspond to unique subjects. Shapes correspond to the mycobiome CST. Across all panels survey 1 or 2 is shown on the x-axis. On the y-axes, the relative abundance of C. auris (top panel) or Shannon diversity (middle panel) is depicted. For this analysis we looked exclusively at sites that transitioned away from domination by Candida auris at the first survey (Survey 1) towards domination by another species at the second time point (Survey 2).

```
df  = data.frame(sample_data(t1p))  %>%
  dplyr::select(Unique_ptid, Cluster,  Survey_Period, SiteID) %>%
  dcast(., Unique_ptid+SiteID~Survey_Period,value.var="Cluster")

colnames(df)[3:4] = c("t1", "t2")
unstable.df =  subset(df, t1=="4" & t2  !="4")
unstable.df$Table.MatcherT1 = paste0(unstable.df$Unique_ptid,";" ,unstable.df$SiteID,";" ,1)
unstable.df$Table.MatcherT2 = paste0(unstable.df$Unique_ptid,";" ,unstable.df$SiteID,";" ,2 )

keep1 = unstable.df$Table.MatcherT1
keep2 = unstable.df$Table.MatcherT2
keep =  c(keep1,  keep2)

sample_data(t1p)$Table.MatcherT1 = paste0(sample_data(t1p)$Unique_ptid,";" ,
                                          sample_data(t1p)$SiteID,";" ,
                                          sample_data(t1p)$Survey_Period)
test = subset_samples(t1p, Table.MatcherT1 %in% keep)
    df = data.frame(otu_table(test), sample_data(test))

    p1 = ggplot(df, aes(Survey_Period, Candida.auris,
                        group=Unique_ptid, color=as.factor(Unique_ptid), shape=CST)) +
          geom_point(size=4)  + geom_line() +
          facet_wrap(~SiteID, ncol=5) +
          scale_x_discrete(breaks=c(1, 2), labels=c(1,2))  + ylab("") +  xlab("") +
           theme(legend.position="bottom") + ggtitle("Cauris relative Abundance")


#get a plot of  the malassezia/candida  index
tax = data.frame(test@tax_table@.Data)
    tax = dplyr::select(tax, c("Family", "Genus"))
    tax = as.matrix(tax)
    tax = tax_table(tax)
    tax_table(test) = tax
    genusTest = tax_glom(test, "Genus")
    df = data.frame(otu_table(genusTest), sample_data(genusTest))
    df$index  = df$Malassezia.restricta/df$Candida.auris


    p2 = ggplot(df, aes(Survey_Period, index,
                        group=Unique_ptid, color=as.factor(Unique_ptid), shape=CST)) +
```

```
          geom_point(size=4)  + geom_line() +
          facet_wrap(~SiteID, ncol=5) +
          scale_x_discrete(breaks=c(1, 2), labels=c(1,2))   + ylab("") + xlab("") +
          ggtitle("Malassezia / Candida Ratio")




#look at diversity
df = data.frame(sample_data(test), estimate_richness(test, measures="Shannon"))

    p4 = ggplot(df, aes(Survey_Period, Shannon,
                        group=Unique_ptid, color=as.factor(Unique_ptid), shape=CST)) +
          geom_point(size=4)  + geom_line() +
          facet_wrap(~SiteID, ncol=5) +
          scale_x_discrete(breaks=c(1, 2), labels=c(1,2))  + ylab("Shannon  Diversity")+
          ylab("") + xlab("") +
          ggtitle("Shannon Diversity") +
          scale_color_discrete(name = "Subject")


SupplementalFigureX1 <- ggarrange((p1 + theme(legend.position = "none")),
                      (p4 + theme(legend.position = "none")), ncol = 1)

SupplementalFigureX1_legend <- cowplot::get_legend(p4)

SupplementalFigureX1 <- cowplot::plot_grid(SupplementalFigureX1, SupplementalFigureX1_legend, nrow = 1, rel_width
s = c(1,.3))
SupplementalFigureX1
```

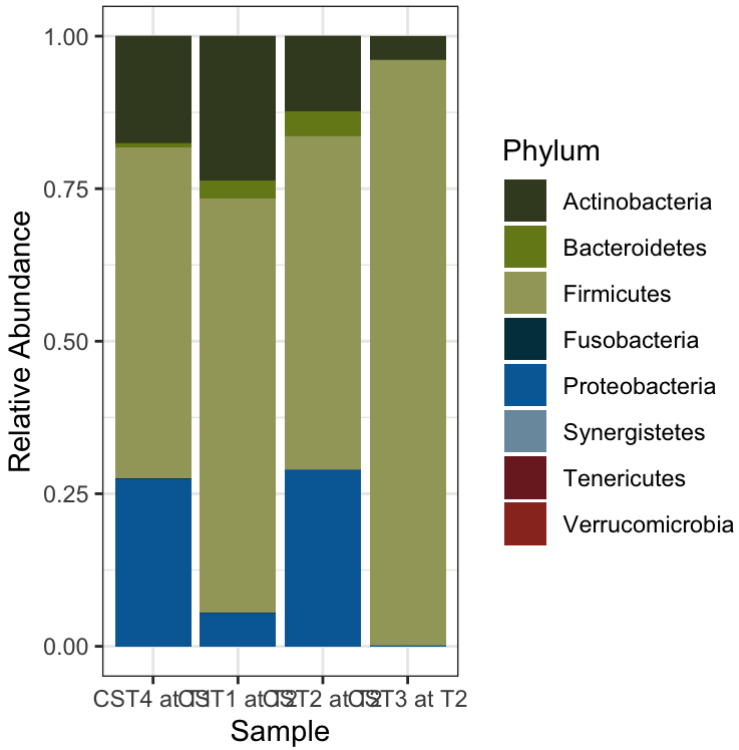


what happens to the bacterial community when you transition away from cauris domination

- we see that communities dominated by Candida auris at T1 are enriched in proteobacteria. As these samples shift from domination by Candida auris at T1 to domination by Malassezia spp at T2 (CST1 at T2; CST3 at T2) there is a reduction in Proteobacteria. Samples shifting from domination by Candida auris towards domination by diverse Candida species maintain high levels of Proteobacteria.

```
t1.keep = paste0("Subject_", unstable.df$Unique_ptid, ";",
                 "Site_", unstable.df$SiteID, ";",
                 "Survey_1")
t2.keep = paste0("Subject_", unstable.df$Unique_ptid, ";",
                 "Site_", unstable.df$SiteID, ";",
                 "Survey_2")

samples2keep = c(t1.keep, t2.keep)

bac_match = readRDS(file="~/Desktop/candida_auris_rush/merged_16s_bac_match_cauris_clinical_map_withsqrt_withtree
_withcoo_2020-03-19.rds") %>%
  subset_samples(., Fungal.Matcher %in% samples2keep) %>%
  prune_taxa(taxa_sums(.) > 0, .)

map = data.frame(sample_data(test)) %>%
  dplyr::select(., c("Fungal.Matcher", "Cluster"))

new.map = plyr::join(map, data.frame(sample_data(bac_match)))
rownames(new.map) = new.map$Fungal.Matcher
sample_data(bac_match) = sample_data(new.map)

df  = data.frame(sample_data(bac_match))  %>%
  dplyr::select(Unique_ptid, Cluster,  Survey_Period, SiteID) %>%
  dcast(., Unique_ptid+SiteID~Survey_Period,value.var="Cluster")

colnames(df)[3:4] = c("t1", "t2")
unstable.df =  subset(df, t1=="4" & t2  !="4")
unstable.df$Table.MatcherT1 = paste0(unstable.df$Unique_ptid,";" ,unstable.df$SiteID,";" ,1)
unstable.df$Table.MatcherT2 = paste0(unstable.df$Unique_ptid,";" ,unstable.df$SiteID,";" ,2 )

keep1 = unstable.df$Table.MatcherT1
keep2 = unstable.df$Table.MatcherT2
keep =  c(keep1,  keep2)

sample_data(bac_match)$Table.MatcherT1 = paste0(sample_data(bac_match)$Unique_ptid,";" ,
                                                sample_data(bac_match)$SiteID,";" ,
                                                sample_data(bac_match)$Survey_Period)
test2 = subset_samples(bac_match, Table.MatcherT1 %in% keep)
    df = data.frame(otu_table(test2), sample_data(test2))
```

```
#look at diversity
df = data.frame(sample_data(test2), estimate_richness(test2, measures="Shannon"))

    p4 = ggplot(df, aes(Survey_Period, Shannon,
                        group=Unique_ptid, color=as.factor(Unique_ptid), shape=Cluster)) +
          geom_point(size=4)  + geom_line() +
          facet_wrap(~SiteID, ncol=5) +
          scale_x_discrete(breaks=c(1, 2), labels=c(1,2))  + ylab("Shannon  Diversity")+
          ylab("") + xlab("") +
          ggtitle("Shannon Diversity") +
          scale_color_discrete(name = "Subject")

sample_data(test2)$glom_var = paste0(sample_data(test2)$Cluster,";" ,
                                             sample_data(test2)$Survey_Period)

foo = merge_samples(test2, "glom_var") %>%
  transform_sample_counts(., function(x) x/sum(x))
p = plot_bar(foo, "Sample", "Abundance", fill="Phylum")
dat = p$data
ordering = c("4;1", "1;2", "2;2", "3;2")
ISU_secondary_palette <- c("#3E4827", "#76881D", "#A2A569",
                           "#003D4C", "#006BA6", "#7A99AC",
                           "#7C2529", "#9A3324", "#BE531C",
                           "#8B5B29", "#B9975B", "#EED484",
                           "#6E6259", "#707372", "#ACA39A", "#C8102E")

dat$Sample  <- factor(dat$Sample, levels = ordering)
      p = ggplot(dat) +
            geom_col(aes(x=Sample, y=Abundance, fill=Phylum), position="stack") +
            ylab("Relative Abundance") +   scale_color_manual(values=ISU_secondary_palette) +
            scale_fill_manual(values=ISU_secondary_palette) + ylim(0, 1)+
        scale_x_discrete(labels=c("4;1" = "CST4 at T1",
                                  "1;2" = "CST1 at T2",
                                  "2;2" = "CST2 at T2",
                                  "3;2" = "CST3 at T2"))

  p
```



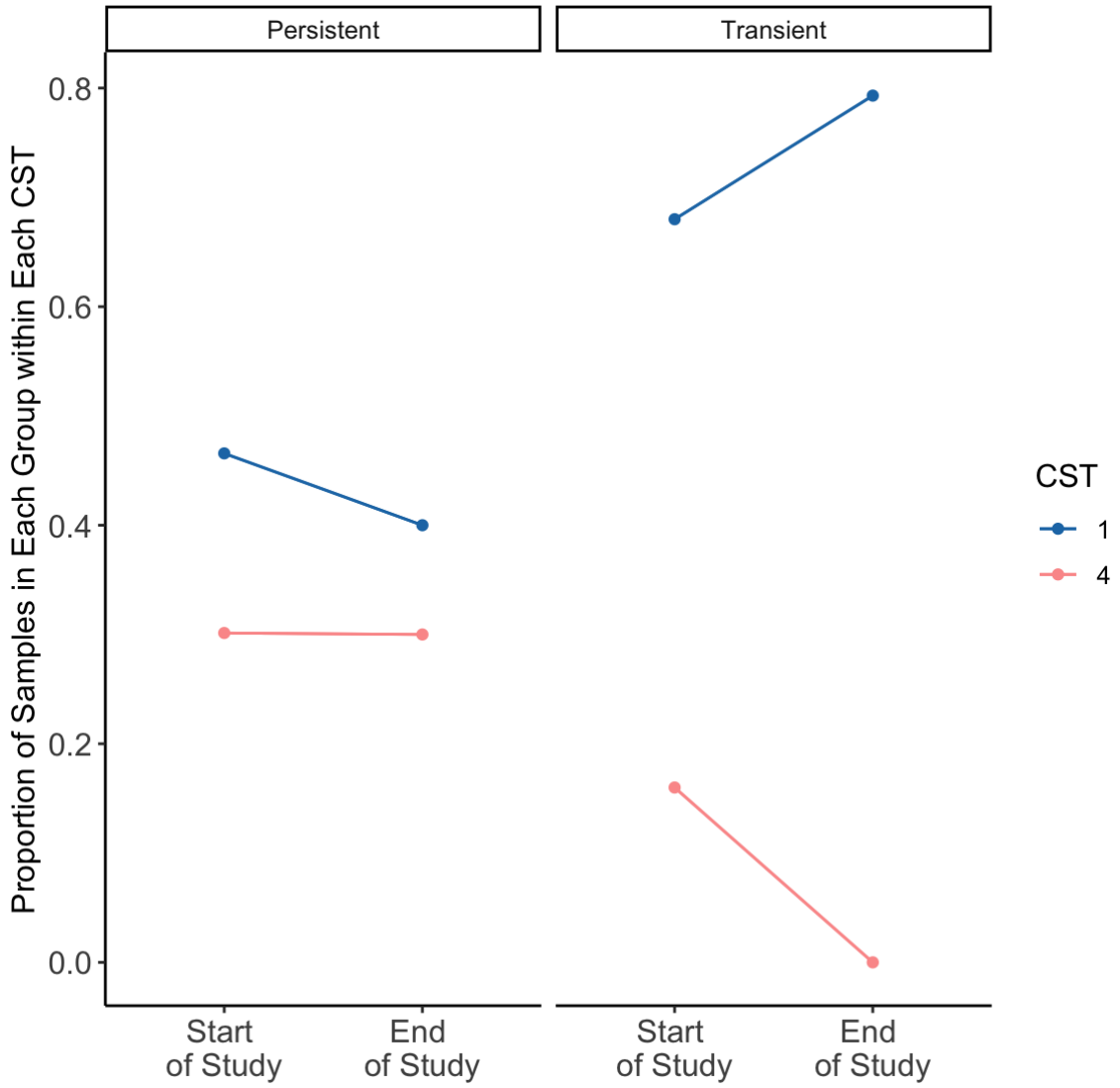
Supplementary Figure 7: What happens to the individuals who are only transiently colonized in terms of CST status?

- 25 samples t1
- 29 samples t2
- 29 samples t3

Supplementary Figure 8: Proportion of sites within each CST over time for individuals who were either transiently or persistently colonized. The proportion of samples dominated by C. auris (CST4) remained roughly constantly (~30%) for individuals who were persistently colonized (Left). In contrast, the proportion of sites dominated by C. auris dropped from 16% to 0% from the first to the third time point in transiently colonized individuals (Right). Among those transiently colonized, the reduction in sites dominated by C. auris was accompanied by a concomitant increase in the proportion of sites dominated by commensal Malassezia species. Of special interest, the proportion of sites dominated by Malassezia species was higher across all time points for those who were transiently colonized compared to those persistently colonized.

```
transient = subset_samples(limphy, outcome== "Transient Colonization")


df  = data.frame(sample_data(transient))  %>%
  dplyr::select(Unique_ptid, Cluster,  Survey_Period, SiteID) %>%
  dcast(., Unique_ptid+SiteID~Survey_Period,value.var="Cluster")

colnames(df)[3:5] = c("t1", "t2", "t3")
dfm = melt(df, id.vars = c("Unique_ptid", "SiteID"))
dfm$complete = complete.cases(dfm$value)
dfm = subset(dfm, complete==TRUE)

p = ggplot(dfm, aes(variable, value, group=Unique_ptid, color=as.factor(Unique_ptid))) + facet_wrap(~SiteID) + ge
om_point() + geom_path()

trans.tab = data.frame(table(dfm$value, dfm$variable))
trans.tab1 = subset(trans.tab, Var2=="t1")
trans.tab1$Prop = trans.tab1$Freq/25

trans.tab2 = subset(trans.tab, Var2=="t2")
trans.tab2$Prop = trans.tab2$Freq/29

trans.tab3 = subset(trans.tab, Var2=="t3")
trans.tab3$Prop = trans.tab3$Freq/29

transDF = data.frame(rbind(trans.tab1, trans.tab2, trans.tab3))
transDF$Group = "Transient"
p1 = ggplot(transDF, aes(Var1, Prop)) + geom_col() + facet_wrap(~Var2) + ggtitle("Transient") + ylim(0, 1)

################### Now let's look at those who are peristently coloniized
transient = subset_samples(limphy, outcome== "Persistent Colonization")


df  = data.frame(sample_data(transient))  %>%
  dplyr::select(Unique_ptid, Cluster,  Survey_Period, SiteID) %>%
  dcast(., Unique_ptid+SiteID~Survey_Period,value.var="Cluster")

colnames(df)[3:5] = c("t1", "t2", "t3")
dfm = melt(df, id.vars = c("Unique_ptid", "SiteID"))
dfm$complete = complete.cases(dfm$value)
```

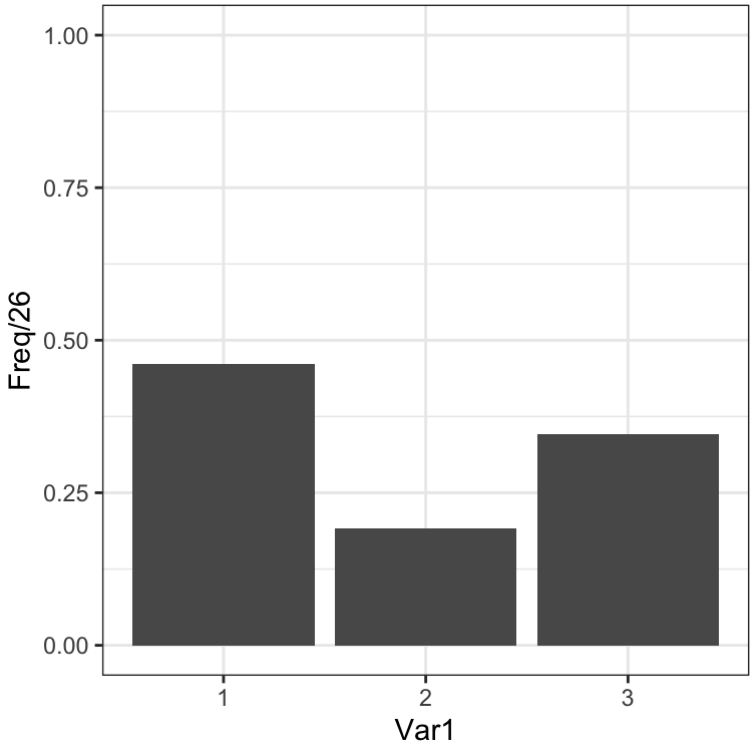
```
dfm = subset(dfm, complete==TRUE)



persist.tab = data.frame(table(dfm$value, dfm$variable))
persist.tab1 = subset(persist.tab, Var2=="t1")
persist.tab1$Prop = persist.tab1$Freq/73

persist.tab2 = subset(persist.tab, Var2=="t2")
persist.tab2$Prop = persist.tab2$Freq/81

persist.tab3 = subset(persist.tab, Var2=="t3")
persist.tab3$Prop = persist.tab3$Freq/80

persistDF = data.frame(rbind(persist.tab1, persist.tab2, persist.tab3))
persistDF$Group = "Persistent"
p2= ggplot(persistDF, aes(Var1, Prop)) + geom_col() + facet_wrap(~Var2)+ ggtitle("Persistent") + ylim(0, 1)

######################### Now let's make a graph for the supplement
newdf = data.frame(rbind(transDF, persistDF)) %>%
  subset(., Var2 %in% c("t1", "t3")) %>%
  subset(., Var1 %in% c(1, 4))


p = ggplot(newdf, aes(Var2, Prop, group=Var1, color=Var1))  + facet_wrap(~Group) + geom_point() + geom_line() +
  ylab("Proportion of Samples in Each Group within Each CST") + geom_line()+
  scale_color_manual(values=CSTColors, name = "CST") +
  scale_x_discrete(labels=c("t1" = "Start \nof Study",
                            "t3" = "End \nof Study"))+
  theme_classic() + xlab("")+
    theme(text = element_text(size=12),
        axis.text.x = element_text(size = 12),
        axis.text.y = element_text(size = 12),
        axis.title.x = element_text( size = 12),
        axis.title.y = element_text( size = 12))
p
```



Are the fingertips stably colonized over time? Or is there really just transient colonization at this site?

We see that a substantial proption of individuals have colonization at just one survey, but an equal fraction are colonized at 2 or more surveys.

```
site_codes = read.csv("~/Desktop/candida_auris_rush/sitecode_to_factored_sites.csv")
data = read.csv("~/Desktop/candida_auris_rush/manuscript/data/Cauris_Analytic_2020-5-20.csv") %>%
  dplyr::select(., c("Unique_ptid", "Survey_Period", "Cauris_Result", "site")) %>%
  plyr::join(., site_codes)

fingers = subset(data, SiteID=="Fg")
#look at the complete data
p = ggplot(fingers, aes(Survey_Period, Cauris_Result, group=Unique_ptid)) + facet_wrap(~Unique_ptid) + geom_point
() + geom_line()
#drop the single time point subjects N=11
fingers1 = subset(fingers, !(Unique_ptid %in% c(6:10, 25, 30, 32, 49, 51, 56)))
p = ggplot(fingers1, aes(Survey_Period, Cauris_Result, group=Unique_ptid)) + facet_wrap(~Unique_ptid) + geom_poin
t() + geom_line()
#drop those who werren't colonized at any timepoint (either 2 or 3) N=17
fingers2 = subset(fingers1, !(Unique_ptid %in% c(1, 5, 11, 13, 20:22, 24, 29, 31, 34, 36, 41, 43:45, 50)))
p = ggplot(fingers2, aes(Survey_Period, Cauris_Result, group=Unique_ptid)) + facet_wrap(~Unique_ptid) + geom_poin
t() + geom_line()
#ddrop those with only 2 tinmepoints
fingers3 = subset(fingers2, !(Unique_ptid %in% c(19, 23, 52)))
p = ggplot(fingers3, aes(Survey_Period, Cauris_Result, group=Unique_ptid)) + facet_wrap(~Unique_ptid) + geom_poin
t() + geom_line()



mylength=function(x) length(x)
df = doBy::summaryBy(Cauris_Result~Unique_ptid, data=fingers3, FUN=c(sum, length))
df$Percent = df$Cauris_Result.sum / df$Cauris_Result.length

df = data.frame(table(df$Cauris_Result.sum))
p = ggplot(df, aes(Var1, Freq/26 )) + geom_col() + ylim(0, 1)
p
```



What are the version numbers of all packages and ultilies used in this script?

```
sessionInfo()
```

```
## R version 4.0.2 (2020-06-22)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Mojave 10.14.6
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/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] RColorBrewer_1.1-2  wesanderson_0.3.6   gridGraphics_0.5-1
##  [4] markovchain_0.8.5-4 scales_1.1.1        vegan_2.5-7
##  [7] lattice_0.20-41     permute_0.9-5       ggalt_0.4.0
## [10] ggpubr_0.4.0        cowplot_1.1.1       stringr_1.4.0
## [13] viridis_0.5.1       viridisLite_0.3.0   gridExtra_2.3
## [16] reshape2_1.4.4      magrittr_2.0.1      ggplot2_3.3.3
## [19] phyloseq_1.32.0     knitr_1.30
##
## loaded via a namespace (and not attached):
##  [1] colorspace_2.0-0    ggsignif_0.6.0      ellipsis_0.3.1
##  [4] rio_0.5.16          XVector_0.28.0      Deriv_4.1.2
##  [7] farver_2.1.0        matlab_1.0.2        ggrepel_0.9.1
## [10] fansi_0.4.2         codetools_0.2-18    splines_4.0.2
## [13] extrafont_0.17      doBy_4.6.8          ade4_1.7-16
## [16] jsonlite_1.7.2      broom_0.7.3         Rttf2pt1_1.3.8
## [19] cluster_2.1.0       compiler_4.0.2      backports_1.2.1
## [22] assertthat_0.2.1    Matrix_1.3-2        htmltools_0.5.1
## [25] prettyunits_1.1.1   tools_4.0.2         igraph_1.2.6
## [28] gtable_0.3.0        glue_1.4.2          dplyr_1.0.5
## [31] maps_3.3.0          Rcpp_1.0.6          carData_3.0-4
## [34] Biobase_2.48.0      cellranger_1.1.0    vctrs_0.3.6
## [37] Biostrings_2.56.0   multtest_2.44.0     ape_5.4-1
## [40] nlme_3.1-151        extrafontdb_1.0     iterators_1.0.13
## [43] xfun_0.20           openxlsx_4.2.3      diagram_1.6.5
```

```
## [46] lifecycle_1.0.0     rstatix_0.6.0       zlibbioc_1.34.0
## [49] MASS_7.3-53         hms_1.0.0           parallel_4.0.2
## [52] biomformat_1.16.0   proj4_1.0-10        rhdf5_2.32.4
## [55] expm_0.999-6        yaml_2.2.1          curl_4.3
## [58] stringi_1.5.3       S4Vectors_0.26.1    foreach_1.5.1
## [61] BiocGenerics_0.34.0 zip_2.1.1           shape_1.4.5
## [64] rlang_0.4.10        pkgconfig_2.0.3     evaluate_0.14
## [67] purrr_0.3.4         Rhdf5lib_1.10.1     labeling_0.4.2
## [70] tidyselect_1.1.0    plyr_1.8.6          R6_2.5.0
## [73] IRanges_2.22.2      generics_0.1.0      DBI_1.1.1
## [76] pillar_1.5.1        haven_2.3.1         foreign_0.8-81
## [79] withr_2.4.1         mgcv_1.8-33         survival_3.2-7
## [82] abind_1.4-5         ash_1.0-15          tibble_3.1.0
## [85] crayon_1.4.1        car_3.0-10          KernSmooth_2.23-18
## [88] utf8_1.1.4          rmarkdown_2.6       progress_1.2.2
## [91] readxl_1.3.1        data.table_1.13.6   forcats_0.5.0
## [94] digest_0.6.27       tidyr_1.1.2         RcppParallel_5.0.2
## [97] stats4_4.0.2        munsell_0.5.0
```