

# Host-specificity and core taxa of seagrass leaf microbiome identified across tissue age and geographical regions | *Sanders-Smith, R. & Segovia, B.T.(joint contribution)*, Forbes, C., Hessing-Lewis, M., Morien, E., Lemay, M.A., O'Connor, M. I., Parfrey, L.W.

Bianca Trevizan Segovia

20/11/2020

## Table S4

```
### load packages ###
library(tidyverse)
library(reshape2)
library(dplyr)
library(stats)
library(ggplot2)
library(ggthemes)
library(Biostrings)
library(phyloseq)

### disabling scientific notation ###
options(scipen = 999)
```

### save Indval Table statistics for Table S4

```
### importing master table ###
All_multipatt <- read.csv("data/IndVal_taxonomy_relatabund.csv", header=T)
colnames(All_multipatt)

## [1] "OTU" "A.specificity." "B.fidelity."
## [4] "indval.stat" "p.value" "indval_group"
## [7] "Rank1" "Rank2" "Rank3"
## [10] "Rank4" "Rank5" "Rank6"
## [13] "Rank7" "artificial.choked" "seawater.choked"
## [16] "seawater.goose" "seawater.mcmullin" "seawater.triquet"
## [19] "zostera_new.choked" "zostera_new.goose" "zostera_new.mcmullin"
## [22] "zostera_new.triquet" "zostera_old.choked" "zostera_old.goose"
## [25] "zostera_old.mcmullin" "zostera_old.triquet"

### melting: rearrange data from wider to longer ###
All_multipatt_melt.RA <- melt(All_multipatt, id.vars = "OTU",
                             measure.vars = c("artificial.choked",
                                                "seawater.choked",
                                                "seawater.goose"),
```

```

        "seawater.mcmullin",
        "seawater.triquet",
        "zostera_new.choked",
        "zostera_new.goose",
        "zostera_new.mcmullin",
        "zostera_new.triquet",
        "zostera_old.choked",
        "zostera_old.goose",
        "zostera_old.mcmullin",
        "zostera_old.triquet"))

### melt automatically assigns column names variable and value ###
### change column names here ###
All_multipatt_melt.RA <- All_multipatt_melt.RA %>%
  dplyr::rename(relat_abund_host_region = variable, Abundance= value)

### add back the metadata and taxonomy info ###
All_multipatt_melt.RA$Rank4 <- All_multipatt$Rank4
All_multipatt_melt.RA$Rank5 <- All_multipatt$Rank5
All_multipatt_melt.RA$Rank6 <- All_multipatt$Rank6
All_multipatt_melt.RA$Rank7 <- All_multipatt$Rank7

### add back indval info while renaming indval_group to Core and stat to Indval ###
All_multipatt_melt.RA$Core <- All_multipatt$indval_group
All_multipatt_melt.RA$Indval <- All_multipatt$indval.stat
All_multipatt_melt.RA$p.value <- All_multipatt$p.value

### create column with only Host labels from the relat_abund_host_region column ###
### case_when and grepl allow to transform:
### i.e. the labels that have "new" in it, in the relat_abund_host_region column, into ~ "New leaf" ###
All_multipatt_melt.RA <- All_multipatt_melt.RA %>%
  dplyr::mutate(Host = case_when(grepl("new", relat_abund_host_region) ~ "New leaf",
                                grepl("old", relat_abund_host_region) ~ "Old leaf",
                                grepl("seawater", relat_abund_host_region) ~ "Seawater",
                                grepl("artificial", relat_abund_host_region) ~ "Artificial"))

### create new column with only Region labels from the relat_abund_host_region column ###
All_multipatt_melt.RA <- All_multipatt_melt.RA %>%
  dplyr::mutate(Region = case_when(grepl("choked", relat_abund_host_region) ~ "Choked",
                                grepl("triquet", relat_abund_host_region) ~ "Triquet",
                                grepl("goose", relat_abund_host_region) ~ "Goose",
                                grepl("mcmullin", relat_abund_host_region) ~ "McMullin"))

All_multipatt_melt.RA$OTU <- as.character(All_multipatt$OTU)
All_multipatt_melt.RA$A.specificity. <- as.character(All_multipatt$A.specificity.)
All_multipatt_melt.RA$B.fidelity. <- as.character(All_multipatt$B.fidelity.)
All_multipatt_melt.RA$Region <- as.character(All_multipatt_melt.RA$Region)
All_multipatt_melt.RA$Core <- as.character(All_multipatt_melt.RA$Core)
All_multipatt_melt.RA$Host <- as.character(All_multipatt_melt.RA$Host)

### LEAVE ALL TAXA SO READER CAN CHECK THE ONES LEFT OUT BY THE THRESHOLD CHOSEN ###
# ### filter indval stats higher than 0.7 ###

```

```

# All_multipatt_melt.RA <- dplyr::filter(All_multipatt_melt.RA, Indval > 0.7)
#
# ### filter prevalence to over 50%
# ### (taxa are present in over 50% of the samples from each indval group) ###
# All_multipatt_melt.RA <- dplyr::filter(All_multipatt_melt.RA, B.fidelity. > 0.5)

### filter for the indval_groups I am interested in ###
All_multipatt_melt.RA.subset <- dplyr::filter(All_multipatt_melt.RA,
                                              Core == "zostera_new" |
                                              Core == "zostera_old" |
                                              Core == "zostera_new+zostera_old" |
                                              Core == "seawater" |
                                              Core == "artificial" |
                                              Core == "artificial+zostera_old")

### create column Core2 with labels for the final graph: this will be the Y-axis of the bubble plot ###
All_multipatt_melt.RA.subset <- All_multipatt_melt.RA.subset %>%
  dplyr::mutate(Core2 = dplyr::recode(Core,
                                     "artificial" = "Artificial Core",
                                     "seawater" = "Seawater Core",
                                     "zostera_new" = "New leaf Core",
                                     "zostera_old" = "Old leaf Core",
                                     "zostera_new+zostera_old" = "Z. marina Core",
                                     "artificial+zostera_old" = "Marine Surfaces Core"))

### create object containing the order of the indval group I want to appear in the y-axis ###
x <- c("Z. marina Core", "New leaf Core",
       "Old leaf Core", "Seawater Core",
       "Artificial Core", "Marine Surfaces Core")

### arrange column according to the order established in x object ###
All_multipatt_melt.RA.subset <- All_multipatt_melt.RA.subset %>%
  dplyr::mutate(Core2 = factor(Core2, levels = x)) %>%
  dplyr::arrange(Core2)

### create column with lowest taxonomy level IDed ###
All_multipatt_melt.RA.subset$lowest_tax_level <- paste(All_multipatt_melt.RA.subset$Rank5,
All_multipatt_melt.RA.subset$Rank6, "_",
All_multipatt_melt.RA.subset$OTU, sep="")

### remove underlines from lowest_tax_level Rank5 labels ###
### remove first 2 characters and replace with nothing ###
All_multipatt_melt.RA.subset$lowest_tax_level <- gsub("^.{0,2}",
                                                    "",
All_multipatt_melt.RA.subset$lowest_tax_level)

## Refine labels ##
### relabel some of the taxa: ###
### add species level where we have it,
### fix some labels for order where family is not available
All_multipatt_melt.RA.subset <- All_multipatt_melt.RA.subset %>%
  dplyr::mutate(lowest_tax_level2 = dplyr::recode(lowest_tax_level,
                                                  "Oceanospirillaceae_Marinomonas_30173" = "Oceanospirillaceae",
                                                  "Oceanospirillaceae_Marinomonas_48134" = "Oceanospirillaceae",

```

"Flavobacteriaceae\_\_Tenacibaculum\_35344" = "Flavobacteriaceae\_\_Tenacibaculum\_35344",  
 "Sva0996\_marine\_group\_\_Sva0996\_marine\_group\_37101" = "Acidimicrobiales\_\_uncultured\_14831",  
 "LWSR.14\_\_LWSR.14\_27699" = "Rickettsiales\_LWSR.14\_27699",  
 "uncultured\_\_uncultured\_14831" = "Acidimicrobiales\_\_uncultured\_14831",  
 "uncultured\_\_uncultured\_14827" = "Acidimicrobiales\_\_uncultured\_14827",  
 "Piscirickettsiaceae\_\_Marine\_Methylotrophic\_Group\_3\_40696" = "Piscirickettsiaceae\_\_Marine\_Methylotrophic\_Group\_3\_40696",  
 "Piscirickettsiaceae\_\_Marine\_Methylotrophic\_Group\_3\_18005" = "Piscirickettsiaceae\_\_Marine\_Methylotrophic\_Group\_3\_18005",  
 "Piscirickettsiaceae\_\_Marine\_Methylotrophic\_Group\_3\_18003" = "Piscirickettsiaceae\_\_Marine\_Methylotrophic\_Group\_3\_18003",  
 "Piscirickettsiaceae\_\_Marine\_Methylotrophic\_Group\_3\_18006" = "Piscirickettsiaceae\_\_Marine\_Methylotrophic\_Group\_3\_18006",  
 "Piscirickettsiaceae\_\_Marine\_Methylotrophic\_Group\_3\_2701" = "Piscirickettsiaceae\_\_Marine\_Methylotrophic\_Group\_3\_2701",  
 "Simkaniaceae\_\_Candidatus\_Fritschea\_892" = "Simkaniaceae\_\_Candidatus\_Fritschea\_892",  
 "Flavobacteriaceae\_\_Polaribacter\_3\_52079" = "Flavobacteriaceae\_\_Polaribacter\_3\_52079",  
 "Halieaceae\_\_OM60(NOR5)\_clade\_44023" = "Halieaceae\_\_OM60(NOR5)\_clade\_44023",  
 "Flavobacteriaceae\_\_NS5\_marine\_group\_17409" = "Flavobacteriaceae\_\_NS5\_marine\_group\_17409",  
 "Flammeovirgaceae\_\_Candidatus\_Amoebophilus\_17904" = "Flammeovirgaceae\_\_Candidatus\_Amoebophilus\_17904",  
 "JTB255\_marine\_benthic\_group\_\_JTB255\_marine\_benthic\_group\_414" = "JTB255\_marine\_benthic\_group\_\_JTB255\_marine\_benthic\_group\_414",  
 "OCS116\_clade\_\_OCS116\_clade\_34534" = "Rhizobiales\_OCS116\_clade\_\_OCS116\_clade\_34534",  
 "FamilyI\_FamilyI\_20142" = "Cyanobacteria\_SubsectionI\_FamilyI\_20142",  
 "Planctomycetaceae\_\_Pir4\_lineage\_22023" = "Planctomycetaceae\_\_Pir4\_lineage\_22023",  
 "uncultured\_\_uncultured\_44981" = "Sphingobacteriales\_uncultured\_44981",  
 "OM190\_OM190\_41159" = "Planctomycetes\_OM190\_41159",  
 "NS9\_marine\_group\_\_NS9\_marine\_group\_47474" = "Flavobacteriales\_\_uncultured\_32551",  
 "RhodobacteraceaeNA\_32551" = "Rhodobacteraceae\_uncultured\_32551",  
 "RhodobacteraceaeNA\_18668" = "Rhodobacteraceae\_uncultured\_18668",  
 "NS9\_marine\_group\_\_NS9\_marine\_group\_17952" = "Flavobacteriales\_\_uncultured\_18668",  
 "Methylophilaceae\_\_OM43\_clade\_30773" = "Methylophilaceae\_\_OM43\_clade\_30773",  
 "Flavobacteriaceae\_\_NS5\_marine\_group\_33603" = "Flavobacteriaceae\_\_NS5\_marine\_group\_33603",  
 "OM1\_clade\_\_Candidatus\_Actinomarina\_3452" = "Acidimicrobiales\_\_Porticoccaceae\_SAR92\_clade\_17098",  
 "Porticoccaceae\_\_SAR92\_clade\_17098" = "Porticoccaceae\_\_SAR92\_clade\_17098",  
 "FamilyI\_\_Synechococcus\_29876" = "Cyanobacteria\_Synechococcus\_29876",  
 "Flavobacteriaceae\_\_NS3a\_marine\_group\_35569" = "Flavobacteriaceae\_\_NS3a\_marine\_group\_35569",  
 "FamilyI\_\_Synechococcus\_15906" = "Cyanobacteria\_Synechococcus\_15906",  
 "SAR86\_clade\_\_SAR86\_clade\_40537" = "Oceanospirillales\_SAR86\_clade\_\_SAR86\_clade\_40537",  
 "Flavobacteriaceae\_\_NS4\_marine\_group\_2487" = "Flavobacteriaceae\_\_NS4\_marine\_group\_2487",  
 "Porticoccaceae\_\_SAR92\_clade\_17017" = "Porticoccaceae\_\_SAR92\_clade\_17017",  
 "FlavobacteriaceaeNA\_35513" = "Flavobacteriaceae\_uncultured\_35513",  
 "Flavobacteriaceae\_\_NS5\_marine\_group\_17553" = "Flavobacteriaceae\_\_NS5\_marine\_group\_17553",  
 "Surface\_1\_Surface\_1\_45078" = "Alphaproteobacteria\_SAR11\_clade\_\_Surface\_1\_Surface\_1\_45078",  
 "Flavobacteriaceae\_\_NS4\_marine\_group\_46779" = "Flavobacteriaceae\_\_NS4\_marine\_group\_46779",  
 "Microbacteriaceae\_\_Candidatus\_Aquiluna\_48034" = "Microbacteriaceae\_\_Candidatus\_Aquiluna\_48034",  
 "Flavobacteriaceae\_\_NS4\_marine\_group\_40661" = "Flavobacteriaceae\_\_NS4\_marine\_group\_40661",  
 "Flavobacteriaceae\_\_NS5\_marine\_group\_33608" = "Flavobacteriaceae\_\_NS5\_marine\_group\_33608",  
 "FamilyI\_\_Synechococcus\_15943" = "Cyanobacteria\_Synechococcus\_15943",  
 "Flavobacteriaceae\_\_NS5\_marine\_group\_40600" = "Flavobacteriaceae\_\_NS5\_marine\_group\_40600",  
 "Microbacteriaceae\_\_Candidatus\_Aquiluna\_35850" = "Microbacteriaceae\_\_Candidatus\_Aquiluna\_35850",  
 "Flavobacteriaceae\_\_NS5\_marine\_group\_50346" = "Flavobacteriaceae\_\_NS5\_marine\_group\_50346",  
 "Flavobacteriaceae\_\_NS5\_marine\_group\_17554" = "Flavobacteriaceae\_\_NS5\_marine\_group\_17554",  
 "FamilyI\_\_Synechococcus\_15989" = "Cyanobacteria\_Synechococcus\_15989",  
 "Surface\_1\_Surface\_1\_45026" = "Alphaproteobacteria\_SAR11\_clade\_\_Surface\_1\_Surface\_1\_45026",

"NS9\_marine\_group\_\_NS9\_marine\_group\_17960" = "Flavobacteriales"  
 "Microbacteriaceae\_\_Candidatus\_Aquiluna\_48033" = "Microbacteriaceae"  
 "FamilyI\_\_Synechococcus\_15949" = "Cyanobacteria\_Synechococcus"  
 "Flavobacteriaceae\_\_NS5\_marine\_group\_40601" = "Flavobacteriaceae"  
  
 "RhodobacteraceaeNA\_49779" = "Rhodobacteraceae\_uncultured\_49779"  
 "Flavobacteriaceae\_\_NS5\_marine\_group\_33607" = "Flavobacteriaceae"  
 "PS1\_clade\_\_PS1\_clade\_46021" = "Rhizobiales\_PS1 clade\_46021",  
 "Flavobacteriaceae\_\_NS5\_marine\_group\_50345" = "Flavobacteriaceae"  
 "NS9\_marine\_group\_\_NS9\_marine\_group\_2296" = "Flavobacteriales"  
 "PS1\_clade\_\_PS1\_clade\_46024" = "Rhizobiales\_PS1 clade\_46024",  
 "Surface\_2\_\_Surface\_2\_44082" = "Alphaproteobacteria\_SAR11 clade\_44082",  
 "Halieaceae\_\_OM60(NOR5)\_clade\_44024" = "Halieaceae\_\_OM60(NOR5)\_clade\_44024",  
 "Porticoccaceae\_\_SAR92\_clade\_17016" = "Porticoccaceae\_\_SAR92 clade\_17016",  
 "Flavobacteriaceae\_\_NS4\_marine\_group\_46780" = "Flavobacteriaceae"  
 "Flavobacteriaceae\_\_NS5\_marine\_group\_40637" = "Flavobacteriaceae"  
 "Flavobacteriaceae\_\_NS2b\_marine\_group\_43365" = "Flavobacteriaceae"  
 "Flavobacteriaceae\_\_NS5\_marine\_group\_40640" = "Flavobacteriaceae"  
 "Flavobacteriaceae\_\_NS4\_marine\_group\_17607" = "Flavobacteriaceae"  
 "SAR116\_clade\_SAR116\_clade\_38203" = "Rickettsiales\_SAR116 clade\_38203",  
 "FlavobacteriaceaeNA\_35415" = "Flavobacteriaceae\_uncultured\_35415",  
 "uncultured\_\_uncultured\_6439" = "Sphingobacteriales\_uncultured\_6439",  
 "Flavobacteriaceae\_\_NS5\_marine\_group\_27262" = "Flavobacteriaceae"  
 "Spongiibacteraceae\_\_BD1.7\_clade\_24610" = "Spongiibacteraceae"  
 "Porticoccaceae\_\_SAR92\_clade\_17015" = "Porticoccaceae\_\_SAR92 clade\_17015",  
 "FlavobacteriaceaeNA\_51493" = "Flavobacteriaceae\_uncultured\_51493",  
 "KI89A\_clade\_\_KI89A\_clade\_16955" = "Gammaproteobacteria\_KI89A clade\_16955",  
 "Flavobacteriaceae\_\_NS5\_marine\_group\_17410" = "Flavobacteriaceae"  
 "SAR86\_clade\_\_SAR86\_clade\_30104" = "Oceanospirillales\_SAR86 clade\_30104",  
 "SAR116\_clade\_\_SAR116\_clade\_30265" = "Rickettsiales\_SAR116 clade\_30265",  
 "Flavobacteriaceae\_\_NS5\_marine\_group\_27264" = "Flavobacteriaceae"  
 "SAR86\_clade\_\_SAR86\_clade\_30105" = "Oceanospirillales\_SAR86 clade\_30105",  
 "Porticoccaceae\_\_SAR92\_clade\_17099" = "Porticoccaceae\_\_SAR92 clade\_17099",  
 "Flavobacteriaceae\_\_NS5\_marine\_group\_17455" = "Flavobacteriaceae"  
 "NS11.12\_marine\_group\_\_NS11.12\_marine\_group\_18042" = "Sphingobacteriales"  
 "E01.9C.26\_marine\_group\_\_E01.9C.26\_marine\_group\_39482" = "Gammaproteobacteria"  
 "Flavobacteriaceae\_\_NS5\_marine\_group\_27263" = "Flavobacteriaceae"  
 "Halieaceae\_\_OM60(NOR5)\_clade\_25329" = "Halieaceae\_\_OM60(NOR5)\_clade\_25329",  
 "Flavobacteriaceae\_\_NS4\_marine\_group\_40660" = "Flavobacteriaceae"  
 "Unknown\_Family\_\_uncultured\_22898" = "Gammaproteobacteria\_Incunabulum",  
 "FamilyI\_\_Synechococcus\_29875" = "Cyanobacteria\_Synechococcus"  
 "Bdellovibrionaceae\_\_OM27\_clade\_33887" = "Bdellovibrionaceae"  
 "PS1\_clade\_\_PS1\_clade\_4919" = "Rhizobiales\_PS1 clade\_4919",  
 "Comamonadaceae\_\_BAL58\_marine\_group\_39554" = "Comamonadaceae"  
 "Surface\_1\_\_Surface\_1\_29349" = "Alphaproteobacteria\_SAR11 clade\_29349",  
 "Surface\_1\_\_Surface\_1\_29343" = "Alphaproteobacteria\_SAR11 clade\_29343",  
 "SAR116\_clade\_\_SAR116\_clade\_30267" = "Rickettsiales\_SAR116 clade\_30267",  
 "Halieaceae\_\_OM60(NOR5)\_clade\_7879" = "Halieaceae\_\_OM60(NOR5)\_clade\_7879",  
 "FamilyI\_\_Synechococcus\_15910" = "Cyanobacteria\_Synechococcus"  
 "Surface\_2\_\_Surface\_2\_44081" = "Alphaproteobacteria\_SAR11 clade\_44081",  
 "FamilyI\_\_Synechococcus\_29877" = "Cyanobacteria\_Synechococcus"  
 "DEV007\_\_DEV007\_41701" = "Verrucomicrobiales\_DEV007\_41701",  
 "EctothiorhodospiraceaeNA\_38096" = "Ectothiorhodospiraceae\_uncultured\_38096"

"Sva0996\_marine\_group\_\_Sva0996\_marine\_group\_48418" = "Acidimicrobiales",  
 "Unknown\_Family\_\_uncultured\_38396" = "Gammaproteobacteria Incertae Sedis",  
 "[Caedibacter]\_caryophilus\_group\_\_[Caedibacter]\_caryophilus\_group\_37052" = "Acidimicrobiaceae",  
 "Acidimicrobiaceae\_\_Ilumatobacter\_37052" = "Acidimicrobiaceae",  
 "NS11.12\_marine\_group\_\_NS11.12\_marine\_group\_18293" = "Sphingobacteriales",  
 "Sva0996\_marine\_group\_\_Sva0996\_marine\_group\_37104" = "Acidimicrobiales",  
 "LWSR-14\_\_LWSR-14\_27699" = "Rickettsiales\_LWSR-14\_27699",  
 "DEV007\_\_DEV007\_21341" = "Verrucomicrobiales\_DEV007\_21341",  
 "KI89A\_clade\_\_KI89A\_clade\_36555" = "Gammaproteobacteria\_KI89A",  
 "LWSR.14\_\_LWSR.14\_27698" = "Rickettsiales\_LWSR.14\_27698",  
 "NA\_19759" = "Chromatiales\_uncultured\_19759",  
 "Sva0996\_marine\_group\_\_Sva0996\_marine\_group\_48494" = "Acidimicrobiales",  
 "NS9\_marine\_group\_\_NS9\_marine\_group\_34991" = "Flavobacteriales",  
 "FlavobacteriaceaeNA\_17637" = "Flavobacteriaceae\_uncultured\_17637",  
 "NS11.12\_marine\_group\_\_NS11.12\_marine\_group\_18292" = "Sphingobacteriales",  
 "uncultured\_\_uncultured\_27953" = "Sphingobacteriales\_uncultured\_27953",  
 "Sva0996\_marine\_group\_\_Sva0996\_marine\_group\_48419" = "Acidimicrobiales",  
 "RhodobacteraceaeNA\_18614" = "Rhodobacteraceae\_uncultured\_18614",  
 "DEV007\_\_DEV007\_41600" = "Verrucomicrobiales\_DEV007\_41600",  
 "Bdellovibrionaceae\_\_OM27\_clade\_33694" = "Bdellovibrionaceae",  
 "Sva0996\_marine\_group\_\_Sva0996\_marine\_group\_37083" = "Acidimicrobiales",  
 "SR1\_(Absconditabacteria)\_\_SR1\_(Absconditabacteria)\_1686" = "Absconditabacteria",  
 "SaprospiraceaeNA\_31206" = "Saprospiraceae\_uncultured\_31206",  
 "BacteriovoracaceaeNA\_38025" = "Bacteriovoracaceae\_uncultured\_38025",  
 "uncultured\_\_uncultured\_43340" = "Alphaproteobacteria Incertae Sedis",  
 "Acidimicrobiaceae\_\_Ilumatobacter\_37054" = "Acidimicrobiaceae",  
 "FlavobacteriaceaeNA\_43497" = "Flavobacteriaceae\_uncultured\_43497",  
 "SR1\_(Absconditabacteria)\_\_SR1\_(Absconditabacteria)\_16876" = "Absconditabacteria",  
 "NS11.12\_marine\_group\_\_NS11.12\_marine\_group\_18196" = "Sphingobacteriales",  
 "SaprospiraceaeNA\_264" = "Saprospiraceae\_uncultured\_264",  
 "Candidatus\_Campbellbacteria\_\_Candidatus\_Campbellbacteria\_855" = "Candidatus\_Campbellbacteria",  
 "NS9\_marine\_group\_\_NS9\_marine\_group\_47478" = "Flavobacteriales",  
 "SR1\_(Absconditabacteria)\_\_SR1\_(Absconditabacteria)\_16877" = "Absconditabacteria",  
 "Unknown\_Family\_\_uncultured\_4229" = "Gammaproteobacteria Incertae Sedis",  
 "KI89A\_clade\_\_KI89A\_clade\_36557" = "Gammaproteobacteria\_KI89A",  
 "NS9\_marine\_group\_\_NS9\_marine\_group\_47476" = "Flavobacteriales",  
 "SR1\_(Absconditabacteria)\_\_SR1\_(Absconditabacteria)\_1676" = "Absconditabacteria",  
 "Blfdi19\_\_Blfdi19\_12501" = "Myxococcales\_Blfdi19\_12501",  
 "uncultured\_\_uncultured\_28102" = "Alphaproteobacteria Incertae Sedis",  
 "Sva0996\_marine\_group\_\_Sva0996\_marine\_group\_37084" = "Acidimicrobiales",  
 "RhodobacteraceaeNA\_49050" = "Rhodobacteraceae\_uncultured\_49050",  
 "Sva0996\_marine\_group\_\_Sva0996\_marine\_group\_48492" = "Acidimicrobiales",  
 "Sva0996\_marine\_group\_\_Sva0996\_marine\_group\_36981" = "Acidimicrobiales",  
 "KI89A\_clade\_\_KI89A\_clade\_36558" = "Gammaproteobacteria\_KI89A",  
 "OM190\_\_OM190\_41199" = "Planctomycetes\_OM190\_41199",  
 "MD3.55\_\_MD3.55\_8538" = "Rickettsiales\_MD3.55\_8538",  
 "DEV007\_\_DEV007\_21339" = "Verrucomicrobiales\_DEV007\_21339",  
 "SaprospiraceaeNA\_31073" = "Saprospiraceae\_uncultured\_31073",  
 "RhodobacteraceaeNA\_51778" = "Rhodobacteraceae\_uncultured\_51778",  
 "KI89A\_clade\_\_KI89A\_clade\_13114" = "Gammaproteobacteria\_KI89A",  
 "DEV007\_\_DEV007\_5336" = "Verrucomicrobiales\_DEV007\_5336",  
 "RhodobacteraceaeNA\_51853" = "Rhodobacteraceae\_uncultured\_51853",  
 "uncultured\_\_uncultured\_44983" = "Sphingobacteriales\_uncultured\_44983"



```

"FamilyI_Phormidium_25998" = "Cyanobacteria_Phormidium_25998",
"PhyllobacteriaceaeNA_18789" = "Phyllobacteriaceae_uncultured",
"uncultured_uncultured_39291" = "Alphaproteobacteria_Incertae",
"OM190__OM190_3822" = "Planctomycetes_OM190_3822",
"OM190__OM190_41206" = "Planctomycetes_OM190_41206",

"OM190__OM190_41159" = "Planctomycetes_OM190_41159",
"NS11-12_marine_group__NS11-12_marine_group_18201" = "Sphingol",
"Flavobacteriaceae_Polaribacter_1_35413" = "Flavobacteriaceae",
"Rhodobacteraceae_Roseobacter_clade_NAC11.7_lineage_31818" =
"Methylophilaceae_OM43_clade_30774" = "Methylophilaceae_OM43",
"Rhodobacteraceae_Roseobacter_clade_NAC11-7_lineage_10635" =
"Spongiibacteraceae_BD1.7_clade_24991" = "Spongiibacteraceae",
"Simkaniaceae_Candidatus_Fritschea_13330" = "Simkaniaceae_Cand",
"NS11.12_marine_group__NS11.12_marine_group_18201" = "Sphingol",
"NS9_marine_group__NS9_marine_group_17743" = "Flavobacteriales",
"NS11.12_marine_group__NS11.12_marine_group_18199" = "Sphingol",
"JL.ETNP.Y6__JL.ETNP.Y6_47274" = "Oceanospirillales_JL.ETNP.Y6",
"FamilyI_FamilyI_20142" = "Cyanobacteria_FamilyI_20142",
"SAR116_clade__SAR116_clade_12953" = "Rickettsiales_SAR116 cl",
"SAR116_clade__SAR116_clade_38203" = "Rickettsiales_SAR116 cl",
"Porticoccaceae__SAR92_clade_17018" = "Porticoccaceae__SAR92",
"Oceanospirillaceae_Marinomonas_48134" = "Oceanospirillaceae__",
"Flavobacteriaceae_Tenacibaculum_35344" = "Flavobacteriaceae__",
"Sva0996_marine_group__Sva0996_marine_group_37101" = "Acidimicrobiales",
"LWSR.14__LWSR.14_27699" = "Rickettsiales__LWSR.14_27699",
"uncultured_uncultured_14831" = "Acidimicrobiales_uncultured",
"unculture_uncultured_14827" = "Acidimicrobiales_uncultured",
"Piscirickettsiaceae_Marine_Methylophilic_Group_3_40696" = "Piscirickettsiaceae",
"Piscirickettsiaceae_Marine_Methylophilic_Group_3_18005" = "Piscirickettsiaceae",
"Piscirickettsiaceae_Marine_Methylophilic_Group_3_18003" = "Piscirickettsiaceae",
"Simkaniaceae_Candidatus_Fritschea_892" = "Simkaniaceae_Cand",
"Flavobacteriaceae_Polaribacter_3_52079" = "Flavobacteriaceae",
"Haliaceae_OM60(NOR5)_clade_44023" = "Haliaceae_OM60(NOR5)",
"Flavobacteriaceae_NS5_marine_group_17409" = "Flavobacteriaceae",
"Flammeovirgaceae_Candidatus_Amoebophilus_17904" = "Flammeovirgaceae",
"Haliaceae_OM60(NOR5)_clade_44023" = "Haliaceae_OM60(NOR5)",
"Flavobacteriaceae_NS5_marine_group_17409" = "Flavobacteriaceae",

))

### remove double-underlines and spaces from labels ###
All_multipatt_melt.RA.subset$lowest_tax_level2 <- gsub("__", "_",
All_multipatt_melt.RA.subset$lowest_tax_level2)

core_taxa_indval_stats <- All_multipatt_melt.RA.subset %>%
  dplyr::select(-c(relat_abund_host_region, Abundance, Core, Host, Region, lowest_tax_level)) %>%
  distinct() # this will remove any duplicate OTUs

core_taxa_indval_stats$OTU <- as.character(core_taxa_indval_stats$OTU)

```

## load fasta file sequences

```
fasta <- readDNASTringSet("data/NODE-REPRESENTATIVES.DOWNSTREAM.fasta")
OTU <- names(fasta)
sequence <- paste(fasta)
df_sequences <- data.frame(OTU, sequence)
df_sequences$OTU <- as.character(df_sequences$OTU)
```

## combine sequences of fasta file with indval taxonomy table

```
Table_S4 <- inner_join(core_taxa_indval_stats, df_sequences, by = "OTU")
write.csv(Table_S4, "data/TableS4.csv", row.names = FALSE)
```