

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)

### 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",
                                                  "Sva0996_marine_group__Sva0996_marine_group_37101" = "Acidimi

```

"LWSR.14_LWSR.14_27699" = "Rickettsiales_LWSR.14_27699",
 "uncultured_uncultured_14831" = "Acidimicrobiales_uncultured",
 "uncultured_uncultured_14827" = "Acidimicrobiales_uncultured",
 "Piscirickettsiaceae_Marine_Methylotrophic_Group_3_40696" = "
 "Piscirickettsiaceae_Marine_Methylotrophic_Group_3_18005" = "
 "Piscirickettsiaceae_Marine_Methylotrophic_Group_3_18003" = "
 "Piscirickettsiaceae_Marine_Methylotrophic_Group_3_18006" = "
 "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",
 "OM190_OM190_41159" = "Planctomycetes_OM190_41159",
 "NS9_marine_group_NS9_marine_group_47474" = "Flavobacteriales_NS9_marine_group_NS9_marine_group_47474",
 "RhodobacteraceaeNA_32551" = "Rhodobacteraceae_uncultured_32551",
 "RhodobacteraceaeNA_18668" = "Rhodobacteraceae_uncultured_18668",
 "NS9_marine_group_NS9_marine_group_17952" = "Flavobacteriales_NS9_marine_group_NS9_marine_group_17952",
 "Methylophilaceae_OM43_clade_30773" = "Methylophilaceae_OM43_clade_30773",
 "Flavobacteriaceae_NS5_marine_group_33603" = "Flavobacteriaceae_NS5_marine_group_33603",
 "OM1_clade_Candidatus_Actinomarina_3452" = "Acidimicrobiales_OM1_clade_Candidatus_Actinomarina_3452",
 "Porticoccaceae_SAR92_clade_17098" = "Porticoccaceae_SAR92_clade_17098",
 "FamilyI_Synechococcus_29876" = "Cyanobacteria_Synechococcus_FamilyI_Synechococcus_29876",
 "Flavobacteriaceae_NS3a_marine_group_35569" = "Flavobacteriaceae_NS3a_marine_group_35569",
 "FamilyI_Synechococcus_15906" = "Cyanobacteria_Synechococcus_FamilyI_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_FamilyI_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_FamilyI_Synechococcus_15989",
 "Surface_1_Surface_1_45026" = "Alphaproteobacteria_SAR11_clade_Surface_1_Surface_1_45026",
 "NS9_marine_group_NS9_marine_group_17960" = "Flavobacteriales_NS9_marine_group_NS9_marine_group_17960",

"Microbacteriaceae__Candidatus_Aquiluna_48033" = "Microbacteriaceae__Candidatus_Aquiluna_48033",
 "FamilyI__Synechococcus_15949" = "Cyanobacteria_Synechococcus_15949",
 "Flavobacteriaceae__NS5_marine_group_40601" = "Flavobacteriaceae__NS5_marine_group_40601",
 "RhodobacteraceaeNA_49779" = "Rhodobacteraceae_uncultured_49779",
 "Flavobacteriaceae__NS5_marine_group_33607" = "Flavobacteriaceae__NS5_marine_group_33607",
 "PS1_clade__PS1_clade_46021" = "Rhizobiales_PS1 clade_46021",
 "Flavobacteriaceae__NS5_marine_group_50345" = "Flavobacteriaceae__NS5_marine_group_50345",
 "NS9_marine_group__NS9_marine_group_2296" = "Flavobacteriales__NS9_marine_group_2296",
 "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__NS4_marine_group_46780",
 "Flavobacteriaceae__NS5_marine_group_40637" = "Flavobacteriaceae__NS5_marine_group_40637",
 "Flavobacteriaceae__NS2b_marine_group_43365" = "Flavobacteriaceae__NS2b_marine_group_43365",
 "Flavobacteriaceae__NS5_marine_group_40640" = "Flavobacteriaceae__NS5_marine_group_40640",
 "Flavobacteriaceae__NS4_marine_group_17607" = "Flavobacteriaceae__NS4_marine_group_17607",
 "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__NS5_marine_group_27262",
 "Spongiibacteraceae__BD1.7_clade_24610" = "Spongiibacteraceae__BD1.7_clade_24610",
 "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__NS5_marine_group_17410",
 "SAR86_clade__SAR86_clade_30104" = "Oceanospirillales_SAR86 clade_30104",
 "SAR116_clade__SAR116_clade_30265" = "Rickettsiales_SAR116 clade_30265",
 "Flavobacteriaceae__NS5_marine_group_27264" = "Flavobacteriaceae__NS5_marine_group_27264",
 "SAR86_clade__SAR86_clade_30105" = "Oceanospirillales_SAR86 clade_30105",
 "Porticoccaceae__SAR92_clade_17099" = "Porticoccaceae__SAR92_clade_17099",
 "Flavobacteriaceae__NS5_marine_group_17455" = "Flavobacteriaceae__NS5_marine_group_17455",
 "NS11.12_marine_group__NS11.12_marine_group_18042" = "Sphingobacteriales_NS11.12_marine_group_18042",
 "E01.9C.26_marine_group__E01.9C.26_marine_group_39482" = "Gammaproteobacteria_E01.9C.26_marine_group_39482",
 "Flavobacteriaceae__NS5_marine_group_27263" = "Flavobacteriaceae__NS5_marine_group_27263",
 "Halieaceae__OM60(NOR5)_clade_25329" = "Halieaceae__OM60(NOR5)_clade_25329",
 "Flavobacteriaceae__NS4_marine_group_40660" = "Flavobacteriaceae__NS4_marine_group_40660",
 "Unknown_Family__uncultured_22898" = "Gammaproteobacteria_Incunabulum_22898",
 "FamilyI__Synechococcus_29875" = "Cyanobacteria_Synechococcus_29875",
 "Bdellovibrionaceae__OM27_clade_33887" = "Bdellovibrionaceae__OM27_clade_33887",
 "PS1_clade__PS1_clade_4919" = "Rhizobiales_PS1 clade_4919",
 "Comamonadaceae__BAL58_marine_group_39554" = "Comamonadaceae__BAL58_marine_group_39554",
 "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_15910",
 "Surface_2__Surface_2_44081" = "Alphaproteobacteria_SAR11 clade_44081",
 "FamilyI__Synechococcus_29877" = "Cyanobacteria_Synechococcus_29877",
 "DEV007__DEV007_41701" = "Verrucomicrobiales_DEV007_41701",
 "EctothiorhodospiraceaeNA_38096" = "EctothiorhodospiraceaeNA_38096",
 "Sva0996_marine_group__Sva0996_marine_group_48418" = "Acidimicrobiales_Sva0996_marine_group_48418",

"Unknown_Family_uncultured_38396" = "Gammaproteobacteria_Incertae
 "[Caedibacter]_caryophilus_group__[Caedibacter]_caryophilus_g
 "Acidimicrobiaceae_Ilumatobacter_37052" = "Acidimicrobiaceae
 "NS11.12_marine_group__NS11.12_marine_group_18293" = "Sphingol
 "Sva0996_marine_group__Sva0996_marine_group_37104" = "Acidimi
 "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" = "Acidimi
 "NS9_marine_group__NS9_marine_group_34991" = "Flavobacteriales
 "FlavobacteriaceaeNA_17637" = "Flavobacteriaceae_uncultured_1
 "NS11.12_marine_group__NS11.12_marine_group_18292" = "Sphingol
 "uncultured_uncultured_27953" = "Sphingobacteriales_uncultur
 "Sva0996_marine_group__Sva0996_marine_group_48419" = "Acidimi
 "RhodobacteraceaeNA_18614" = "Rhodobacteraceae_uncultured_186
 "DEV007_DEV007_41600" = "Verrucomicrobiales_DEV007_41600",
 "Bdellovibrionaceae_OM27_clade_33694" = "Bdellovibrionaceae_
 "Sva0996_marine_group__Sva0996_marine_group_37083" = "Acidimi
 "SR1_(Absconditabacteria)_SR1_(Absconditabacteria)_1686" = "
 "SaprospiraceaeNA_31206" = "Saprospiraceae_uncultured_31206",
 "BacteriovoracaceaeNA_38025" = "Bacteriovoracaceae_uncultured
 "uncultured_uncultured_43340" = "Alphaproteobacteria_Incertae
 "Acidimicrobiaceae_Ilumatobacter_37054" = "Acidimicrobiaceae
 "FlavobacteriaceaeNA_43497" = "Flavobacteriaceae_uncultured_4
 "SR1_(Absconditabacteria)_SR1_(Absconditabacteria)_16876" = "
 "NS11.12_marine_group__NS11.12_marine_group_18196" = "Sphingol
 "SaprospiraceaeNA_264" = "Saprospiraceae_uncultured_264",
 "Candidatus_Campbellbacteria_Candidatus_Campbellbacteria_855
 "NS9_marine_group__NS9_marine_group_47478" = "Flavobacteriales
 "SR1_(Absconditabacteria)_SR1_(Absconditabacteria)_16877" = "
 "Unknown_Family_uncultured_4229" = "Gammaproteobacteria_Ince
 "KI89A_clade_KI89A_clade_36557" = "Gammaproteobacteria_KI89A
 "NS9_marine_group__NS9_marine_group_47476" = "Flavobacteriales
 "SR1_(Absconditabacteria)_SR1_(Absconditabacteria)_1676" = "
 "Blfdi19_Blfdi19_12501" = "Myxococcales_Blfdi19_12501",
 "uncultured_uncultured_28102" = "Alphaproteobacteria_Incertae
 "Sva0996_marine_group__Sva0996_marine_group_37084" = "Acidimi
 "RhodobacteraceaeNA_49050" = "Rhodobacteraceae_uncultured_490
 "Sva0996_marine_group__Sva0996_marine_group_48492" = "Acidimi
 "Sva0996_marine_group__Sva0996_marine_group_36981" = "Acidimi
 "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_517
 "KI89A_clade_KI89A_clade_13114" = "Gammaproteobacteria_KI89A
 "DEV007_DEV007_5336" = "Verrucomicrobiales_DEV007_5336",
 "RhodobacteraceaeNA_51853" = "Rhodobacteraceae_uncultured_518
 "uncultured_uncultured_44983" = "Sphingobacteriales_uncultur
 "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__OM4",
"Rhodobacteraceae__Roseobacter_clade_NAC11-7_lineage_10635" =
"Spongiibacteraceae__BD1.7_clade_24991" = "Spongiibacteraceae",
"Simkaniaceae__Candidatus_Fritschea_13330" = "Simkaniaceae__C",
"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.Y",
"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 c

))

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

### create an object with top 10 core taxa of each sample type ###
OTUs_top_10 <- c("Methylophilaceae__Methylotheria_45266",
  "Methylophilaceae__Methylotheria_45304",
  "Rhodobacteraceae__uncultured_32477",
  "Methylophilaceae__Methylotheria_51108",
  "Methylophilaceae__Methylotheria_29491",
  "Methylophilaceae__Methylotheria_45135",
  "Alteromonadaceae__Paraglaciecola_3808",
  "Oceanospirillaceae__Marinomonas_48192",
  "Methylophilaceae__Methylotheria_45179",
  "Alteromonadaceae__Alteromonas_25375",
  "Saprospiraceae__Rubidimonas_39641",
  "Piscirickettsiaceae__Marine_Methylotheria_Group_3_40696",
  "Rhizobiaceae__Rhizobium_19725",
  "Oceanospirillaceae__Marinomonas_30173",
  "Piscirickettsiaceae__uncultured_42467",
  "Piscirickettsiaceae__uncultured_42432",
  "Oceanospirillaceae__Marinomonas_48134",
  "Piscirickettsiaceae__Marine_Methylotheria_Group_3_18005",
  "Methylophilaceae__Methylotheria_29493",
  "Methylophilaceae__Methylotheria_29578",
  "Piscirickettsiaceae__Marine_Methylotheria_Group_3_18003",
  "Granulosicoccaceae__Granulosicoccus_15155",
  "Alteromonadaceae__Glaciecola_7867",
  "Granulosicoccaceae__Granulosicoccus_46431",
  "Rhodobacteraceae__Octadecabacter_40749",

```

```

"Spongiibacteraceae__BD1.7_clade_24991",
"Haliaceae__Pseudohalaea_25282",
"Granulosicoccaceae__Granulosicoccus_15168",
"Simkaniaceae__Candidatus_Fritschea_892",
"Planctomycetaceae__Blastopirellula_22112",
"Saprospiraceae__Lewinella_33348",
"Flavobacteriaceae__Ulvibacter_24786",
"Flavobacteriaceae__Polaribacter_3_52079",
"Haliaceae__OM60(NOR5)_clade_44023",
"Flavobacteriaceae__NS5_marine_group_17409",
"Flavobacteriaceae__Tenacibaculum_35344",
"Rhodobacteraceae__Ascidiahabitans_31970",
"Haliaceae__Luminiphilus_7857",
"Cryomorphaceae__uncultured_34172",
"Flavobacteriaceae__Formosa_43620",
"Flavobacteriaceae__Flavicella_27016",
"Saprospiraceae__uncultured_2307",
"Thiotrichaceae__Cocleimonas_45935",
"Thiotrichaceae__Cocleimonas_32070",
"Oceanospirillaceae__Oleibacter_36424",
"Flammeovirgaceae__Candidatus_Amoebophilus_17904",
"Sva0996_marine_group__Sva0996_marine_group_37101",
"Rhodobacteraceae__uncultured_37802",
"Thiotrichaceae__Cocleimonas_32145",
"LWSR.14__LWSR.14_27699",
"Aeromonadaceae__uncultured_43869",
"Thiotrichaceae__Cocleimonas_51144",
"Saprospiraceae__uncultured_44888",
"uncultured__uncultured_14831",
"uncultured__uncultured_14827",
"Flavobacteriaceae__Maribacter_47051",
"Saprospiraceae__Rubidimonas_39402",
"Saprospiraceae__Lewinella_10383",
"Erythrobacteraceae__Erythrobacter_415",
"Granulosicoccaceae__Granulosicoccus_49894",
"Granulosicoccaceae__Granulosicoccus_49861")

```

```

multipatt_top_10 <- All_multipatt_melt.RA.subset %>%
  dplyr::filter(lowest_tax_level %in% OTUs_top_10 )

```

```
### relabel some of the taxa:
```

```
### add species level where we have it,
```

```
### fix some labels for order where family is not available ###
```

```

multipatt_top_10 <- multipatt_top_10 %>%

```

```
  dplyr::mutate(lowest_tax_level2= dplyr::recode(lowest_tax_level,
```

```
    "Oceanospirillaceae__Marinomonas_30173" = "Oceanospirillaceae__
```

```
    "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__Marine_Methylophilic_Group_3_18005"
"Piscirickettsiaceae__Marine_Methylophilic_Group_3_18003" = "Piscirickettsiaceae__Marine_Methylophilic_Group_3_18003"
"Simkaniaceae__Candidatus_Fritschea_892" = "Simkaniaceae__Candidatus_Fritschea_892"
"Flavobacteriaceae__Polaribacter_3_52079" = "Flavobacteriaceae__Polaribacter_3_52079"
"Haliaceae__OM60(NOR5)_clade_44023" = "Haliaceae__OM60(NOR5)_clade_44023"
"Flavobacteriaceae__NS5_marine_group_17409" = "Flavobacteriaceae__NS5_marine_group_17409"
"Flammeovirgaceae__Candidatus_Amoebophilus_17904" = "Flammeovirgaceae__Candidatus_Amoebophilus_17904"
"Haliaceae__OM60(NOR5)_clade_44023" = "Haliaceae__OM60(NOR5)_clade_44023"
"Flavobacteriaceae__NS5_marine_group_17409" = "Flavobacteriaceae__NS5_marine_group_17409"
))

### remove doubleunderlines and spaces from labels ###
multipatt_top_10$lowest_tax_level2 <- gsub("__", "_",
multipatt_top_10$lowest_tax_level2)

multipatt_top_10 <- multipatt_top_10 %>%
  dplyr::select(-c(relat_abund_host_region, Abundance, Core, Host, Region, lowest_tax_level)) %>%
  distinct() # this will remove any duplicate OTUs

multipatt_top_10$OTU <- as.character(multipatt_top_10$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(multipatt_top_10, df_sequences, by = "OTU")
write.csv(Table_S4, "data/TableS4.csv", row.names = FALSE)

```