Dormancy and dispersal structure bacterial communities across ecosystem boundaries

Nathan I. Wisnoski, Mario E. Muscarella, Megan L. Larsen, and Jay T. Lennon

04 November, 2019

# Initial Setup

First, we’ll load the packages we’ll need for the analysis, as well as some other functions.

# Import Required Packages  
library("png")  
library("grid")  
library("tidyverse")   
library("vegan")  
library("viridis")  
library("cowplot")  
library("ggrepel")  
library("iNEXT")  
library("broom")  
library("ggpmisc")  
library("pander")  
library("lubridate")  
library("betapart")  
library("adespatial")  
library("VennDiagram")  
  
source("bin/mothur\_tools.R")  
se <- function(x, ...){sd(x, na.rm = TRUE)/sqrt(length(na.omit(x)))}

Next, we’ll set the aesthetics of the figures we will produce.

my.cols <- RColorBrewer::brewer.pal(n = 4, name = "Greys")[3:4]  
  
# Set theme for figures in the paper  
theme\_set(theme\_classic() +   
 theme(axis.title = element\_text(size = 16),  
 axis.title.x = element\_text(margin = margin(t = 15, b = 15)),  
 axis.title.y = element\_text(margin = margin(l = 15, r = 15)),  
 axis.text = element\_text(size = 14),  
 axis.text.x = element\_text(margin = margin(t = 5)),  
 axis.text.y = element\_text(margin = margin(r = 5)),  
 #axis.line.x = element\_line(size = 1),  
 #axis.line.y = element\_line(size = 1),  
 axis.line.x = element\_blank(),  
 axis.line.y = element\_blank(),  
 axis.ticks.x = element\_line(size = 1),  
 axis.ticks.y = element\_line(size = 1),  
 axis.ticks.length = unit(.1, "in"),  
 panel.border = element\_rect(color = "black", fill = NA, size = 1.5),  
 legend.title = element\_blank(),  
 legend.text = element\_text(size = 14),  
 strip.text = element\_text(size = 14),  
 strip.background = element\_blank()  
 ))

## Import Data

Here, we read in the processed sequence files from mothur (shared and taxonomy) and a design of the sampling. We also load in the environmental data. We then remove the mock community from the dataset and ensure the the design and OTU table are aligned by row.

# Define Inputs  
# Design = general design file for experiment  
# shared = OTU table from mothur with sequence similarity clustering  
# Taxonomy = Taxonomic information for each OTU  
design <- "data/UL.design.txt"  
shared <- "data/ul\_resgrad.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.opti\_mcc.shared"  
taxon <- "data/ul\_resgrad.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.pick.opti\_mcc.0.03.cons.taxonomy"  
  
# Import Design  
design <- read.delim(design, header=T, row.names=1)  
  
# Import Shared Files  
OTUs <- read.otu(shared = shared, cutoff = "0.03") # 97% Similarity  
  
# Import Taxonomy  
OTU.tax <- read.tax(taxonomy = taxon, format = "rdp")  
  
# Load environmental data  
env.dat <- read.csv("data/ResGrad\_EnvDat.csv", header = TRUE)  
env.dat <- env.dat[-c(16,17,18),]  
  
# Subset to just the reservoir gradient sites  
OTUs <- OTUs[str\_which(rownames(OTUs), "RG"),]  
OTUs <- OTUs[-which(rownames(OTUs) == "RGMockComm"),]  
  
# make sure OTU table matches up with design order  
design <- design[-c(34:39),]  
OTUs <- OTUs[match(rownames(design), rownames(OTUs)),]  
design$distance <- max(na.omit(design$distance)) - design$distance  
env.dat$distance <- max(na.omit(env.dat$dist.dam)) - env.dat$dist.dam

## Clean and transform OTU table

Here, we remove OTUs with low incidence across sites, we remove any samples with low coverage, and we standardize the OTU table by log-transforming the abundances and relativizing by site.

# Remove OTUs with less than two occurences across all sites  
#OTUs <- OTUs[, which(colSums(OTUs) >= 2)]  
  
# Sequencing Coverage  
coverage <- rowSums(OTUs)  
  
# Remove Low Coverage Samples (This code removes two sites: Site 5DNA, Site 6cDNA)  
lows <- which(coverage < 10000)  
OTUs <- OTUs[-which(coverage < 10000), ]  
design <- design[-which(coverage < 10000), ]  
otus.for.inext <- t(OTUs)  
# Remove OTUs with < 2 occurences across all sites  
OTUs <- OTUs[, which(colSums(OTUs) >= 2)]  
coverage <- rowSums(OTUs)  
  
# Rarify the community, nest RNA in DNA, and reorganize OTU table  
set.seed(47405)  
OTUs <- rrarefy(OTUs, min(coverage))  
OTUs.w.dna <- OTUs[which(design$type == "water" & design$molecule == "DNA"),]  
rowSums((OTUs.w.dna > 1))

## RGD01 RGD02 RGD03 RGD04 RGD06 RGD07 RGD08 RGD09 RGD10 RGD11 RGD12 RGD13   
## 319 405 468 372 415 693 545 704 687 1050 1387 515   
## RGD14 RGD15   
## 548 1313

OTUs.w.rna <- OTUs[which(design$type == "water" & design$molecule == "RNA"),]  
rowSums((OTUs.w.rna > 1))

## RGc01 RGc02 RGc03 RGc04 RGc05 RGc07 RGc08 RGc09 RGc10 RGc11 RGc12 RGc13   
## 130 142 69 283 142 56 101 162 462 159 185 163   
## RGc14 RGc15   
## 108 107

OTUs.w.dna <- OTUs.w.dna + as.matrix(decostand(OTUs.w.rna, method = "pa"))  
rowSums((OTUs.w.dna > 1))

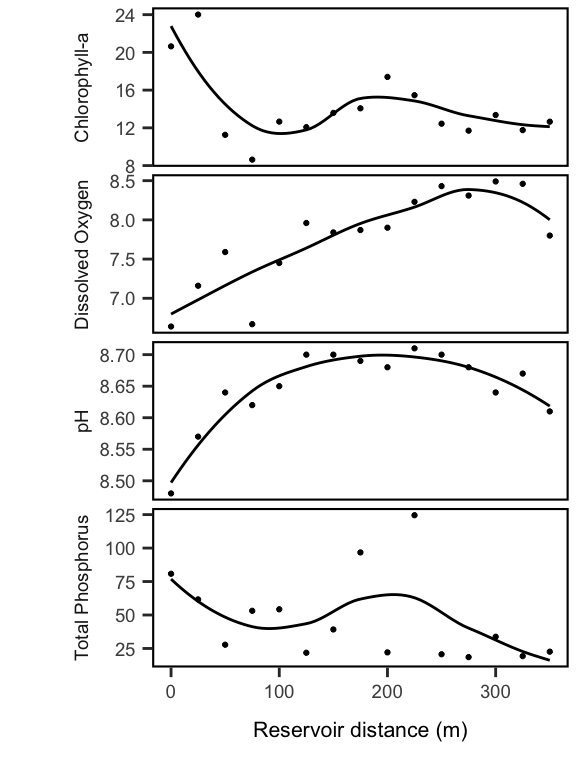
## RGD01 RGD02 RGD03 RGD04 RGD06 RGD07 RGD08 RGD09 RGD10 RGD11 RGD12 RGD13   
## 325 412 472 385 429 699 554 712 741 1065 1396 531   
## RGD14 RGD15   
## 566 1321

OTUs <- rbind(OTUs[1:3,],  
 OTUs.w.dna,  
 OTUs.w.rna)  
OTUs <- OTUs[match(rownames(design), rownames(OTUs)),]  
  
# Make Relative Abundance Matrices  
OTUsREL <- decostand(OTUs, method = "total")  
  
# Log Transform Relative Abundances  
OTUsREL.log <- decostand(OTUs, method = "log")

# Reservoir environmental gradients

Just to see if there are any strong underlying resource or nutrient gradients in the reservoir, we’ll plot them along the distance of the reservoir.

facet.labs <- c(`chla` = "Chlorophyll-a",  
 `color` = "Color",  
 `DO` = "Dissolved Oxygen",  
 `pH` = "pH",  
 `TP` = "Total Phosphorus")  
  
env.dat %>% select(distance, DO, pH, TP, chla) %>%   
 gather(variable, value, -distance) %>%   
 ggplot(aes(x = distance, y = value)) +   
 geom\_point() +   
 geom\_smooth(method = "loess", color = "black", se = F) +   
 facet\_grid(variable ~., scales = "free", switch = "y",   
 labeller = as\_labeller(facet.labs)) +   
 theme(strip.background = element\_blank(),   
 strip.text = element\_text(size = 14),  
 strip.placement = "outside") +   
 labs(x = "Reservoir distance (m)",  
 y = "") +  
 scale\_y\_continuous() +  
 ggsave("figures/env\_vars.pdf", height = 3/4\*4\*3, width = 4, units = "in")



So, there are some weak gradients, but nothing too prevailing.

# Analyze Diversity

Now, we will analyze the bacterial diversity in the reservoir and nearby soils to figure out how well they support different mechanisms of community assembly.

## How does -diversity vary along the reservoir?

First, we use the method of rarefaction and extrapolation developed by Chao et al. in the iNEXT package.

# Observed Richness  
S.obs <- rowSums((OTUs > 0) \* 1)  
  
# Simpson's Evenness  
SimpE <- function(x = ""){  
 x <- as.data.frame(x)  
 D <- diversity(x, "inv")  
 S <- sum((x > 0) \* 1)   
 E <- (D)/S   
 return(E)  
}  
simpsE <- round(apply(OTUs, 1, SimpE), 3)  
shan <- diversity(OTUs, index = "shannon")  
exp.shan <- exp(shan)  
alpha.div <- cbind(design, S.obs, simpsE, shan, exp.shan)  
  
# define singleton estimator from Chiu and Chao 2016 PeerJ  
source("bin/Chao\_functions.R")  
  
# # estimate richness  
singleton.apply <- function(x){  
 singleton.Est(x, "abundance")$corrected.data  
}  
  
# otus.for.inext <- apply(otus.for.inext, MARGIN = 2, singleton.apply)  
# divestim <- estimateD(otus.for.inext, datatype = "abundance",  
# base = "size", conf = 0.95)  
# saveRDS(divestim, file = "intermediate-data/inext-output.rda")  
divestim <- readRDS("intermediate-data/inext-output.rda")  
divestim.df <- divestim %>%   
 mutate(habitat = str\_to\_title(design[as.character(site),"type"]))

Here is the resulting curve, showing the higher diversity in soil samples relative to the lake samples.

# divestim.df %>%  
# ggplot(aes(x = x, y = y,  
# ymin = y.lwr, ymax = y.upr,  
# color = habitat, fill = habitat, group = site)) +  
# geom\_ribbon(data=subset(divestim.df, method == "extrapolated"), alpha = 0.3) +  
# geom\_line(data=subset(divestim.df, method == "interpolated"), size = 1, alpha = .8) +  
# geom\_line(alpha = 1, linetype = "dashed") +  
# scale\_x\_continuous(labels = scales::comma, limits = c(0, 90000)) +  
# labs(x = "Sample size", y = "Estimated richness") +  
# theme(legend.position = "none") +  
# #theme(legend.position = c(.88,.5)) +  
# annotate(label = "Soil", size = 6, geom = "text", x = 85000, y = 5000) +  
# annotate(label = "Water", size = 6, geom = "text", x = 85000, y = 1500) +  
# scale\_color\_grey(end = .7) +  
# scale\_fill\_grey(end = .7)

Next, we’ll extract the estimates for the Hill numbers at different levels of q, which differentially weight common versus rare species.

# hill.estim <- divestim$AsyEst %>% filter(Diversity == "Species richness") %>%   
# left\_join(rownames\_to\_column(alpha.div), by = c("Observed" = "S.obs")) %>%   
# select(Site, rowname, station, molecule, type, distance) %>%   
# left\_join(divestim$AsyEst, by = "Site")  
  
hill.water <- divestim.df %>%   
 filter(site %in% rownames(OTUs)) %>%   
 left\_join(rownames\_to\_column(alpha.div, var = "site")) %>%   
 filter(habitat == "Water")

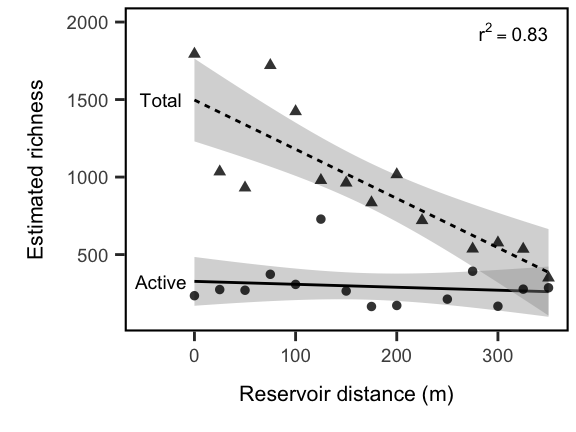
## Warning: Column `site` joining factor and character vector, coercing into  
## character vector

hill.water.rich <- subset(hill.water, order == 0)  
hill.water.shan <- subset(hill.water, order == 1)  
hill.water.simp <- subset(hill.water, order == 2)  
  
hill.water.mod.rich <- lm(qD ~ distance \* molecule, data = hill.water.rich)  
hill.water.mod.shan <- lm(qD ~ distance \* molecule, data = hill.water.shan)  
hill.water.mod.simp <- lm(qD ~ distance \* molecule, data = hill.water.simp)  
  
# summary(hill.water.mod.rich)  
# summary(hill.water.mod.shan)  
# summary(hill.water.mod.simp)  
  
# tidy up the model output  
hill.water.mods <- as\_tibble(rbind.data.frame(  
 tidy(hill.water.mod.rich) %>% add\_column(Diversity = "Richness"),  
 tidy(hill.water.mod.shan) %>% add\_column(Diversity = "Shannon"),  
 tidy(hill.water.mod.simp) %>% add\_column(Diversity = "Simpson")  
))

# Summary table of the model results.   
hill.water.mods %>%   
 group\_by(Diversity) %>%   
 rename("Term" = term,   
 "Estimate" = estimate,   
 "Std. Error" = std.error,   
 "Statistic" = statistic,   
 "p-value" = p.value) %>%   
 select(Diversity, everything()) %>%   
 pander(round = 4)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Diversity | Term | Estimate | Std. Error | Statistic | p-value |
| Richness | (Intercept) | 1497 | 100.6 | 14.88 | 0 |
| Richness | distance | -3.176 | 0.4976 | -6.381 | 0 |
| Richness | moleculeRNA | -1170 | 142.3 | -8.222 | 0 |
| Richness | distance:moleculeRNA | 2.985 | 0.7003 | 4.263 | 3e-04 |
| Shannon | (Intercept) | 153.7 | 19.41 | 7.921 | 0 |
| Shannon | distance | -0.2941 | 0.096 | -3.062 | 0.0053 |
| Shannon | moleculeRNA | -123.9 | 27.46 | -4.513 | 1e-04 |
| Shannon | distance:moleculeRNA | 0.2457 | 0.1352 | 1.818 | 0.0815 |
| Simpson | (Intercept) | 55.44 | 6.47 | 8.57 | 0 |
| Simpson | distance | -0.0783 | 0.032 | -2.446 | 0.0221 |
| Simpson | moleculeRNA | -36.78 | 9.151 | -4.019 | 5e-04 |
| Simpson | distance:moleculeRNA | 0.0402 | 0.045 | 0.8918 | 0.3813 |

# hill.estim %>% filter(type == "water") %>%   
# mutate(molecule = ifelse(molecule == "DNA", "Total", "Active")) %>%   
# ggplot(aes(x = distance, y = Estimator,   
# ymin = LCL, ymax = UCL,  
# color = molecule, fill = molecule, shape = molecule)) +   
# geom\_point(size =3) +   
# # geom\_errorbar(size = .5, aes(ymin = Estimator - s.e., ymax = Estimator + s.e.),   
# # width = 10, alpha = 0.5) +  
# geom\_smooth(method = "lm", aes(linetype = molecule)) +  
# labs(x = "Reservoir distance (m)",  
# y = "") +  
# scale\_color\_manual(values = my.cols) +  
# scale\_fill\_manual(values = my.cols) +   
# theme(legend.position = c(.88,.95), strip.placement = "outside",  
# strip.text = element\_text(size = 16)) +  
# scale\_x\_reverse() +  
# facet\_grid(Diversity ~ ., scales = "free", switch = "y") +  
# guides(fill = guide\_legend(override.aes=list(fill=NA)))  
 #facet\_grid(Diversity ~ ., scales = "free")  
  
# postitions for labels  
xpos = max((na.omit(hill.water$distance)))  
yposDNA = predict(hill.water.mod.rich, newdata = data.frame(distance = 0, molecule = "DNA"))  
yposRNA = predict(hill.water.mod.rich, newdata = data.frame(distance = 0, molecule = "RNA"))  
alpha.fig <- hill.water %>% filter(type == "water", order == 0) %>%   
 mutate(molecule = ifelse(molecule == "DNA", "Total", "Active")) %>%   
 ggplot(aes(x = distance, y = qD,   
 ymin = qD.LCL, ymax = qD.UCL,  
 shape = molecule)) +   
 # geom\_errorbar(size = .5, width = 10, alpha = 0.5) +  
 geom\_smooth(method = "lm", aes(linetype = molecule), color = "black") +  
 geom\_point(size =3, alpha = 0.8) +   
 labs(x = "Reservoir distance (m)",  
 y = "Estimated richness") +  
 scale\_y\_continuous(breaks = seq(0, 2000, by = 500)) +  
 scale\_x\_continuous(limits = c(-49, 350)) +  
 theme(legend.position = "none") +  
 guides(fill = guide\_legend(override.aes=list(fill=NA))) +  
 annotate("text", x = -33, y = yposRNA ,   
 label = "Active", size = 5) +  
 annotate("text", x = -33, y = yposDNA ,   
 label = "Total", size = 5) +  
 annotate(geom = "text", x = xpos, y = 2000, hjust = 1, vjust = 1, size = 5,  
 label = paste0("r^2== ",round(summary(hill.water.mod.rich)$r.squared, 2)), parse = T) +  
 ggsave("figures/alpha\_fig.pdf")  
alpha.fig

 So, from the basis of these results, we can make the following conclusions. First, we note that diversity in the total community decays from the stream inlet to the dam of the reservoir. That is, all the lines have a negative slope. However, we do not see this decay in the metabolically active community. Second, we note that the metabolically actively community has much lower diversity than the total community near the soils, but this difference decreases toward the dam. Last, because we quantified diversity across three orders of Hill numbers (q = 0, 1, and 2), we can also say something about the relative importance of rare versus common taxa along the reservoir transect. We see the the significance of the distance-by-molecule interaction term decrease as rare taxa are downweighted in favor of common taxa. This suggests that the differences between the active and total communities along the transect is driven primarily by rare taxa. However, the general trend of higher Simpson diversity across the whole transect suggests that low-activity, but relatively common, taxa are maintained in the reservoir.

## Similarity To Terrestrial Habitat Across Gradient (Terrestrial Influence)

Here, we fit a linear model to the similarity of the aquatic community to the soil community.

# Similarity to Soil Sample  
UL.bray <- 1-as.matrix(vegdist(OTUsREL.log, method="bray"))  
UL.bray.lake <- UL.bray[-c(1:3), 1:3]   
bray.mean <- round(apply(UL.bray.lake, 1, mean), 3)  
bray.se <- round(apply(UL.bray.lake, 1, se), 3)  
UL.sim <- cbind(design[-c(1:3), ], bray.mean, bray.se)  
  
# Calculate Linear Model  
model.terr <- lm(bray.mean ~ distance \* molecule, data = UL.sim)  
predict(model.terr, newdata = data.frame(distance = 0, molecule = c("RNA", "DNA")))

## 1 2   
## 0.03090104 0.17193131

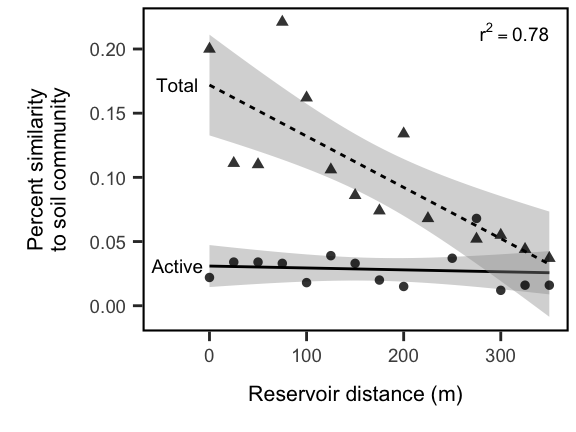
pander(model.terr)

Fitting linear model: bray.mean ~ distance \* molecule

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| **(Intercept)** | 0.1719 | 0.0138 | 12.46 | 5.707e-12 |
| **distance** | -0.0003988 | 6.827e-05 | -5.841 | 5.045e-06 |
| **moleculeRNA** | -0.141 | 0.01952 | -7.226 | 1.821e-07 |
| **distance:moleculeRNA** | 0.0003839 | 9.608e-05 | 3.996 | 0.0005324 |

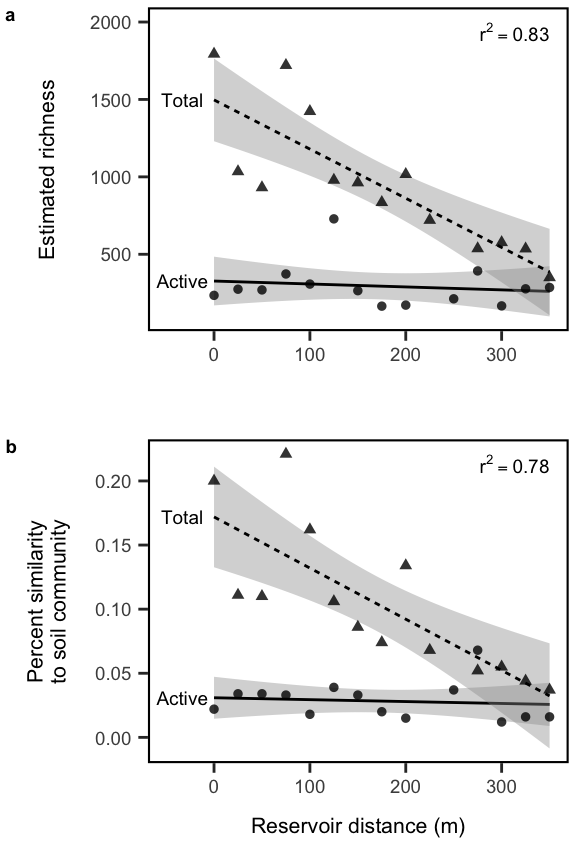
# # Calculate Confidance Intervals of Model  
# newdata.terr <- data.frame(cbind(UL.sim$molecule, UL.sim$distance))  
# conf95.terr <- predict(model.terr, newdata.terr, interval="confidence")  
#   
# # Dummy Variables Regression Model ("Terrestrial Influence")  
# D2 <- (UL.sim$molecule == "RNA")\*1  
# fit.Fig.3b <- lm(UL.sim$bray.mean ~ UL.sim$distance + D2 + UL.sim$distance\*D2)  
# D2.R2 <- round(summary(fit.Fig.3b)$r.squared, 2)  
# summary(fit.Fig.3b)  
#   
#   
# DNA.int.3b <- fit.Fig.3b$coefficients[1]  
# DNA.slp.3b <- fit.Fig.3b$coefficients[2]  
# RNA.int.3b <- DNA.int.3b + fit.Fig.3b$coefficients[3]  
# RNA.slp.3b <- DNA.slp.3b + fit.Fig.3b$coefficients[4]

ypred.act <- predict(model.terr, newdata = data.frame(distance = 0, molecule = "RNA"))  
ypred.tot <- predict(model.terr, newdata = data.frame(distance = 0, molecule = "DNA"))  
similarity.plot <- UL.sim %>%   
 mutate(molecule = ifelse(UL.sim$molecule == "DNA", "Total", "Active")) %>%   
 ggplot(aes(x = distance, y = bray.mean, shape = molecule)) +  
 geom\_smooth(method = "lm", aes(linetype = molecule), color = "black", show.legend = T) +   
 geom\_point(alpha = 0.8, size = 3, show.legend = T) +   
 labs(y = str\_wrap("Percent similarity to soil community", width = 20),   
 x = "Reservoir distance (m)") +   
 theme(legend.position = "none") +  
 scale\_x\_continuous(limits = c(-49,350)) +   
 annotate(geom = "text", x = 350, y = max(UL.sim$bray.mean), hjust = 1, vjust = 1, size = 5,  
 label = paste0("r^2== ",round(summary(model.terr)$r.squared, 2)), parse = T) +  
 annotate("text", x = -33, y = ypred.act, label = "Active", size = 5) +  
 annotate("text", x = -33, y = ypred.tot, label = "Total", size = 5) +  
 ggsave("figures/similarity\_fig.pdf")  
  
similarity.plot

 We find that our model captures most of the variation in community structure . We note a significant influence of distance on community similarity and the presence of a significant interaction between distance and whether the comparison is for active or total bacterial communities. This indicates that total communities decay faster with distance to soils than active communities do, which might be explained by the large difference in initial intercept. Active communities are always highly dissimilar to soil communities and remain so across the lake, while total lake communities are initially similar to soils, but this influence dissipates with distance into the reservoir.

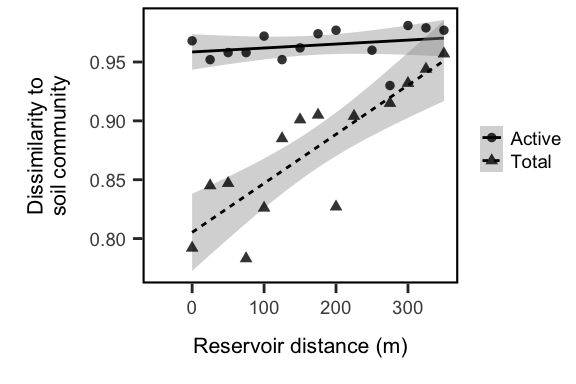
### Create combined figure

plot\_grid(alpha.fig + labs(x = ""), similarity.plot,   
 align = "hv",  
 labels = "auto", ncol = 1) +  
 ggsave("figures/alpha\_similarity\_paneled.pdf")

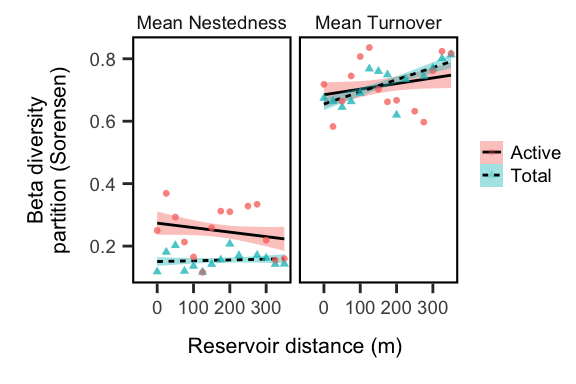


## Are the aquatic samples nested subsets of the soil?

betapart.sor <- beta.pair(decostand(OTUs, method = "pa"), "sorensen")  
  
nest.lake <- as.matrix(betapart.sor$beta.sne)[-c(1:3), 1:3]   
nest.mean <- round(apply(nest.lake, 1, mean), 3)  
nest.se <- round(apply(nest.lake, 1, se), 3)  
UL.nest <- cbind(design[-c(1:3), ], nest.mean, nest.se)  
  
turn.lake <- as.matrix(betapart.sor$beta.sim)[-c(1:3), 1:3]   
turn.mean <- round(apply(turn.lake, 1, mean), 3)  
turn.se <- round(apply(turn.lake, 1, se), 3)  
UL.turn <- cbind(design[-c(1:3), ], turn.mean, turn.se)  
  
sor.lake <- as.matrix(betapart.sor$beta.sor)[-c(1:3), 1:3]   
sor.mean <- round(apply(sor.lake, 1, mean), 3)  
sor.se <- round(apply(sor.lake, 1, se), 3)  
UL.sor <- cbind(design[-c(1:3), ], sor.mean, sor.se)  
  
left\_join(UL.nest, UL.turn) %>% left\_join(UL.sor) %>%   
 mutate(molecule = ifelse(molecule == "DNA", "Total", "Active")) %>%   
 ggplot(aes(x = distance, y = sor.mean, shape = molecule)) +  
 geom\_smooth(method = "lm", aes(linetype = molecule), color = "black", show.legend = T) +   
 geom\_point(alpha = 0.8, size = 3, show.legend = T) +   
 labs(y = str\_wrap("Dissimilarity to soil community", width = 20),   
 x = "Reservoir distance (m)") +   
 scale\_x\_continuous(limits = c(-49,350))



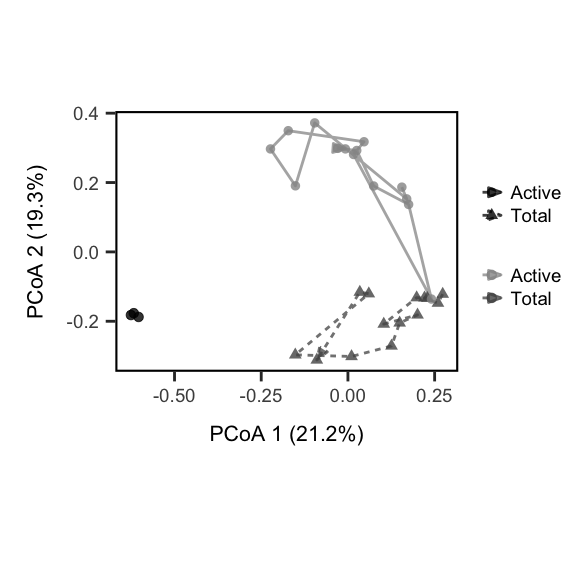
betadivcomp.sor <- beta.div.comp(mat = OTUsREL.log, coef = "S", quant = FALSE, save.abc = FALSE)  
  
rich.lake <- as.matrix(betadivcomp.sor$rich)[-c(1:3), 1:3]   
rich.se <- round(apply(rich.lake, 1, se), 3)  
rich.mean <- round(apply(rich.lake, 1, mean), 3)  
UL.rich <- cbind(design[-c(1:3), ], rich.mean, rich.se)  
  
repl.lake <- as.matrix(betadivcomp.sor$repl)[-c(1:3), 1:3]   
repl.mean <- round(apply(repl.lake, 1, mean), 3)  
repl.se <- round(apply(repl.lake, 1, se), 3)  
UL.repl <- cbind(design[-c(1:3), ], repl.mean, repl.se)  
  
UL\_betapartitions <- left\_join(UL.nest, UL.turn) %>% left\_join(UL.rich) %>% left\_join(UL.repl) %>%   
 gather(nest.se, turn.se, rich.se, repl.se, key = "partition", value = "se") %>%   
 gather(nest.mean, turn.mean, rich.mean, repl.mean, key = "partition", value = "beta")  
   
UL\_betapartitions %>%   
 mutate(molecule = ifelse(molecule == "DNA", "Total", "Active")) %>%   
 mutate(family = ifelse(partition %in% c("nest.mean", "turn.mean"), "Baselga", "Podani")) %>%   
 filter(family == "Baselga") %>%   
 mutate(partition = ifelse(partition == "nest.mean", "Mean Nestedness", "Mean Turnover")) %>%   
 ggplot(aes(x = distance, y = beta, shape = molecule, color = molecule, fill = molecule)) +  
 geom\_smooth(method = "lm", aes(linetype = molecule), color = "black", show.legend = T) +   
 geom\_point(alpha = 0.3, size = 2, show.legend = T) +   
 #geom\_errorbar(aes(ymax = beta + se, ymin = beta - se), width = 10) +  
 facet\_wrap(.~partition) +  
 labs(y = str\_wrap("Beta diversity partition (Sorensen)", width = 20),   
 x = "Reservoir distance (m)") +   
 scale\_x\_continuous(limits = c(-49,350)) +  
 ggsave("figures/nestedness.pdf", width = 8, height = 4)



## How does community structure change along the gradient?

First, we’ll just get an overview of how the communities look along the aquatic transect.

ul.pcoa <- cmdscale(vegdist(OTUsREL.log, method="bray"), 2, eig = T, add = T)  
explainvars <- round(eigenvals(ul.pcoa)[c(1,2)]/sum(eigenvals(ul.pcoa)),3) \*100  
water.pcvals <- data.frame(scores(ul.pcoa)) %>%   
 rownames\_to\_column("name") %>%   
 left\_join(rownames\_to\_column(design, "name")) %>%   
 arrange(desc(distance)) %>% filter(type == "water")  
soil.pcvals <- data.frame(scores(ul.pcoa)) %>%   
 rownames\_to\_column("name") %>%   
 left\_join(rownames\_to\_column(design, "name")) %>%   
 arrange(desc(distance)) %>% filter(type == "soil")  
pc\_dists <- tibble(  
 DNA\_dim1 = subset(water.pcvals, molecule == "DNA")$Dim1,  
 DNA\_dim2 = subset(water.pcvals, molecule == "DNA")$Dim2,  
 RNA\_dim1 = subset(water.pcvals, molecule == "RNA")$Dim1,  
 RNA\_dim2 = subset(water.pcvals, molecule == "RNA")$Dim2)  
  
pcoa.fig <- data.frame(scores(ul.pcoa)) %>%   
 rownames\_to\_column("name") %>%   
 left\_join(rownames\_to\_column(design, "name")) %>%   
 arrange(desc(distance)) %>% filter(type == "water") %>%   
 mutate(molecule = ifelse(molecule == "DNA", "Total", "Active")) %>%   
 ggplot(aes(x = Dim1, y = Dim2)) +  
 geom\_path(size = 1, alpha = 0.75, arrow = arrow(angle = 20,  
 length = unit(0.35, "cm"),  
 type = "closed"), aes(color = molecule, linetype = molecule)) +  
 geom\_point(size = 3, alpha = 0.8, aes(color = molecule, shape = molecule)) +   
 geom\_point(data = select(soil.pcvals, Dim1, Dim2), col = "black", alpha = .8, size = 3) +  
 scale\_color\_manual("Community Subset", values = my.cols) +  
 geom\_segment(data = pc\_dists,  
 aes(x = DNA\_dim1, y = DNA\_dim2,  
 xend = RNA\_dim1, yend = RNA\_dim2),  
 alpha = 0) +  
 coord\_fixed(ratio = 1) +  
 labs(x = paste0("PCoA 1 (", explainvars[1],"%)"),  
 y = paste0("PCoA 2 (", explainvars[2],"%)")) +  
 # theme(legend.position = "none") +  
 # annotate(geom = "text", x = .2, y = 0, label = "Active", size = 5) +  
 # annotate(geom = "text", x = -.25, y = -.3, label = "Total", size = 5) +   
 # annotate(geom = "text", x = .3, y = -.4, label = "Soils", size = 5) +  
 ggsave("figures/pcoa.pdf")  
pcoa.fig



So, it appears that there is convergence in community structure along the path from stream inlet to the dam. This could reflect a loss of soil-derived taxa in the aquatic samples. To test this, we’ll look at -diversity along the gradient with respect to the soil samples. If we see a decay in similarity to soils, this suggests soil taxa are having a comparatively lower influence with distance from the inlet.

# Identifying the Soil Bacteria

Now, we wish to determine whether soil-derived taxa are driving this pattern, and then ask who these influential soil bacteria are.

To classify soil bacteria, we take an incidence-based approach and classify OTUs as:  
- present in the soil and present, but never active, in the reservoir  
- present in the soil and active in the reservoir

# separate lake and soil samples  
lake.total <- OTUs[which(design$molecule == "DNA", design$type == "water"),]  
soil.total <- OTUs[which(design$molecule == "DNA", design$type == "soil"),]  
  
# which otus are present in both lake and soil samples  
lake.and.soil.total <- OTUs[which(design$molecule == "DNA", design$type == "water"),  
 which(colSums(lake.total) > 0 & colSums(soil.total) > 0)]  
  
# isolate just the dna and rna lake communities  
w.dna <- OTUs[which(design$molecule == "DNA" & design$type == "water"), ]  
w.rna <- OTUs[which(design$molecule == "RNA" & design$type == "water"), ]  
  
# pull out the lake rna counts for otus found in lake and soil  
lake.and.soil.act <- w.rna[,colnames(lake.and.soil.total)]  
  
# of these lake and soil taxa, which are never active? active?  
nvr.act <- which(colSums(lake.and.soil.act) == 0)  
yes.act <- which(colSums(lake.and.soil.act) != 0)  
  
# how many otus are active relative to the total number of otus   
length(nvr.act) / ncol(lake.and.soil.total) # 88% of soil-derived bac never active

## [1] 0.8210454

length(yes.act) / ncol(soil.total) # 8% of all soil taxa were active in lake

## [1] 0.1327096

# of taxa who were never active, what fraction of the total community did they represent?  
sum(rowSums(w.dna[,names(nvr.act)]))

## [1] 23585

sum(rowSums(w.dna[,names(yes.act)]))

## [1] 499388

sum(rowSums(w.dna[,names(nvr.act)])) / sum(rowSums(w.dna))

## [1] 0.04509793

# of taxa who became active, what fraction of the dna community did they represent?  
sum(rowSums(w.dna[,names(yes.act)])) / sum(rowSums(w.dna))

## [1] 0.9549021

prop.nvr.act <- rowSums(w.dna[,nvr.act]) / rowSums(w.dna)  
# cbind.data.frame(design.dna, inactive = prop.nvr.act) %>%   
# ggplot(aes(x = distance, y = inactive)) +  
# geom\_point() +   
# geom\_line(stat = "smooth", method = "lm", formula = y ~ x, se = F) +  
# labs(x = "Reservoir transect (m)", y = "Rel. abundance of taxa\n that are never active") +  
# scale\_x\_reverse()

We calculate the richness of the soil taxa that are never active in the lake. We calculate richness from the DNA-based samples.

# pull out their dna abundances and calculate richness  
terr.lake <- w.dna[ , c(names(nvr.act))]  
terr.rich <- rowSums((terr.lake > 0) \* 1)  
terr.REL <- rowSums(terr.lake) / rowSums(w.dna)   
design.dna <- design[which(design$molecule == "DNA" & design$type == "water"), ]  
terr.rich.log <- log10(terr.rich)  
terr.REL.log <- log10(terr.REL)  
  
terr.mod1 <- lm(terr.rich.log ~ design.dna$distance)  
summary(terr.mod1)

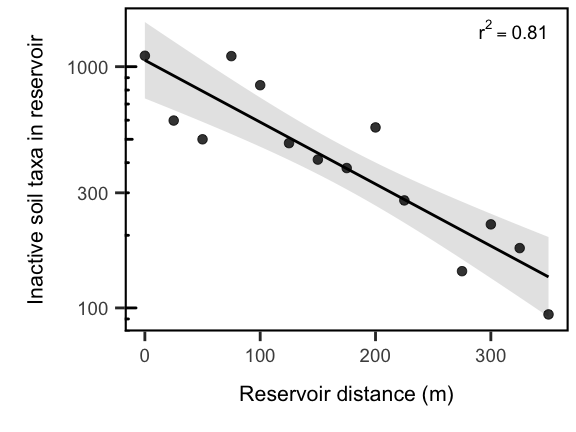
##   
## Call:  
## lm(formula = terr.rich.log ~ design.dna$distance)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.199417 -0.123300 -0.000783 0.080926 0.234711   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.0266909 0.0726577 41.657 2.37e-14 \*\*\*  
## design.dna$distance -0.0025661 0.0003595 -7.138 1.18e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1478 on 12 degrees of freedom  
## Multiple R-squared: 0.8094, Adjusted R-squared: 0.7935   
## F-statistic: 50.95 on 1 and 12 DF, p-value: 1.184e-05

T1.R2 <- round(summary(terr.mod1)$r.squared, 2)  
T1.int <- terr.mod1$coefficients[1]  
T1.slp <- terr.mod1$coefficients[2]  
pander(terr.mod1)

Fitting linear model: terr.rich.log ~ design.dna$distance We find distance is a highly significant predictor of the richness of these soil-derived taxa (on a log-scale).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| **(Intercept)** | 3.027 | 0.07266 | 41.66 | 2.374e-14 |
| **design.dna$distance** | -0.002566 | 0.0003595 | -7.138 | 1.184e-05 |

transient.plot <- tibble(transient\_rich = terr.rich, distance = design.dna$distance) %>%   
 ggplot(aes(x = distance, y = transient\_rich)) +   
 geom\_smooth(method = "lm", color = "black", fill = "grey") +  
 geom\_point(size = 3, alpha = .8, color = "black") +   
 scale\_y\_log10() +  
 annotation\_logticks(sides = "l", size = 1) +  
 labs(x = "Reservoir distance (m)",  
 y = "Inactive soil taxa in reservoir") +  
 annotate("text", x = 350, y = max(terr.rich) + 200, hjust = 1, vjust = 0, size = 5,  
 label = paste0("r^2== ",T1.R2), parse = T) +  
 ggsave("figures/transients.pdf")  
transient.plot



# plot\_grid(alpha.fig,   
# similarity.plot,   
# pcoa.fig + ,   
# transient.plot,  
# align = "hv", axis = "tlbr",  
# labels = "auto", ncol = 2) +  
# ggsave("figures/large\_panel.pdf", width = 12, height = 8)

# See which taxa are shared between habitats

OTUs.PA <- decostand(OTUsREL, method = "pa")  
soil <- names(which(colSums(OTUs.PA[design$type == "soil",]) > 0))  
water.dna <- names(which(colSums(OTUs.PA[design$type == "water" & design$molecule == "DNA",]) > 0))  
water.rna <- names(which(colSums(OTUs.PA[design$type == "water" & design$molecule == "RNA",]) > 0))  
  
sum(water.rna %in% water.dna)

## [1] 2085

nsoil <- length(soil)  
nwdna <- length(water.dna)  
nwrna <- length(water.rna)  
otus.by.habitat <- list("Soil" = soil, "Total Aquatic" = water.dna, "Active Aquatic" = water.rna)  
  
venn.diagram(otus.by.habitat, "figures/venn\_by\_habitat.png",   
 imagetype = "png",   
 fontfamily = "sans",   
 cat.fontfamily = "sans",  
 alpha = .25)

## [1] 1

# What is the fate of soil-derived taxa in the reservoir?

So, we observe that most soil-derived taxa appear to decay once they enter the reservoir. Do any soil-derived taxa persist in the active bacterial community of the reservoir and do they rise to high relative abundances?

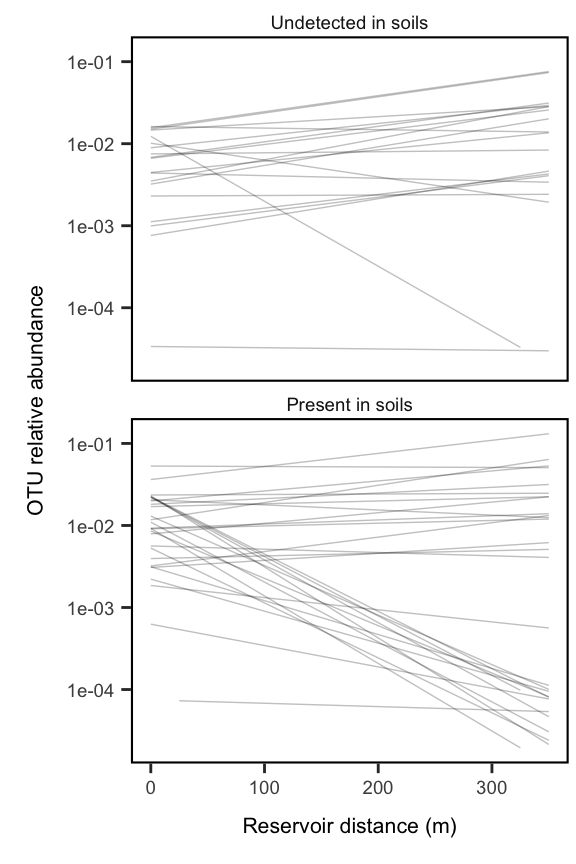
# identify otus in soil samples and lake samples  
in.soil <- OTUs[, which(colSums(OTUs[c(1:3),]) > 0 )]  
#in.lake <- OTUs[, which(colSums(OTUs[-c(1:3),]) > 0)]  
  
# isolate just the rna water samples and convert to presence-absence  
in.lake.rna <- OTUs[which(design$molecule == "RNA" & design$type == "water"), ]  
in.lake.rna.pa <- (in.lake.rna > 0) \* 1  
  
# define the 'core' taxa as otus present in 50% of samples  
in.lake.core <- w.dna[, which((colSums(in.lake.rna.pa) / nrow(in.lake.rna.pa)) >= 0.75)]  
  
# of the core, how many are also in the soil samples?  
in.lake.core.from.soils <- in.lake.core[, intersect(colnames(in.lake.core), colnames(in.soil))]  
  
# of the core which are not in the soil samples  
in.lake.core.not.soils <- in.lake.core[, setdiff(colnames(in.lake.core), colnames(in.soil))]  
  
# Find the relative abundance of the core taxa and prepare data frame to plot  
in.lake.core.from.soils.REL <- in.lake.core.from.soils / rowSums(w.dna)  
  
in.soil.to.plot <- as.data.frame(in.lake.core.from.soils.REL) %>%   
 rownames\_to\_column("sample\_ID") %>%   
 gather(otu\_id, rel\_abundance, -sample\_ID) %>%   
 left\_join(rownames\_to\_column(design.dna, "sample\_ID")) %>%   
 add\_column(found = "soils")  
  
in.lake.core.not.soils.REL <- in.lake.core.not.soils / rowSums(w.dna)  
  
in.lake.to.plot <- as.data.frame(in.lake.core.not.soils.REL) %>%   
 rownames\_to\_column("sample\_ID") %>%   
 gather(otu\_id, rel\_abundance, -sample\_ID) %>%   
 left\_join(rownames\_to\_column(design.dna, "sample\_ID")) %>%   
 add\_column(found = "lake")

Now, lets plot the abundances of the OTUs across the reservoir and split them up into whether they were recovered in soils or not.

bind\_rows(in.soil.to.plot, in.lake.to.plot) %>%   
 ggplot(aes(x = distance, y = rel\_abundance, group = otu\_id)) +   
 labs(x = "Reservoir distance (m)",   
 y = "OTU relative abundance") +  
 geom\_line(alpha = 0.25, stat = "smooth", method = "lm", se = F, show.legend = F) +  
 scale\_y\_log10() +  
 facet\_wrap(~ found, ncol = 1,   
 labeller = as\_labeller(c(  
 `lake` = "Undetected in soils",  
 `soils` = "Present in soils")))

## Warning: Transformation introduced infinite values in continuous y-axis

## Warning: Removed 10 rows containing non-finite values (stat\_smooth).



From this figure, we note a few important points. First, we observe that core reservoir taxa that are not detected in the soil samples tend to increase in relative abundance along the reservoir transect. We also note that for the taxa that are present in the soil samples, some tend to increase drastically, while others tend to increase, along the transect. This suggests that there may be two classes of soil-derived OTUs that contribute to reservoir bacterial diversity:  
- taxa where the reservoir is a sink (i.e., maintained via mass effects from the soils) - aquatic taxa seeded by populations stored in the soils

# model distance effect on rel abundance to get slope and pval  
soil.core.mods <- apply(in.lake.core.from.soils.REL, MARGIN = 2,   
 FUN = function(x) summary(lm(x ~ design.dna$distance))$coefficients[2,c(1,4)])  
rownames(soil.core.mods) <- c("slope", "pval")  
  
# classify otus as significantly increasing or decreasing along reservoir  
soil.core.decreasing <- as.data.frame(t(soil.core.mods)) %>%   
 rownames\_to\_column("OTU") %>%   
 filter(slope < 0) %>% # rel abund decreases toward dam  
 left\_join(OTU.tax)

## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector

soil.core.increasing <- as.data.frame(t(soil.core.mods)) %>%   
 rownames\_to\_column("OTU") %>%   
 filter(slope > 0) %>% # rel abund increases toward dam  
 left\_join(OTU.tax)

## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector

nonsoil.core.mods <- apply(in.lake.core.not.soils.REL, MARGIN = 2,   
 FUN = function(x) summary(lm(x ~ design.dna$distance))$coefficients[2,c(1,4)])  
rownames(nonsoil.core.mods) <- c("slope", "pval")  
nonsoil.core.decreasing <- as.data.frame(t(nonsoil.core.mods)) %>%   
 rownames\_to\_column("OTU") %>%   
 filter(slope < 0) %>% # rel abund decreases toward dam  
 left\_join(OTU.tax)

## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector

nonsoil.core.increasing <- as.data.frame(t(nonsoil.core.mods)) %>%   
 rownames\_to\_column("OTU") %>%   
 filter(slope > 0) %>% # rel abund increases toward dam  
 left\_join(OTU.tax)

## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector

Now we will visualize the significant taxa

pander(nonsoil.core.decreasing, caption = "Core taxa not found in soils that get rarer along the transect.")

Core taxa not found in soils that get rarer along the transect. (continued below)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| OTU | slope | pval | Domain | Phylum |
| Otu00007 | -7.887e-06 | 0.2473 | Bacteria | Proteobacteria |
| Otu00020 | -1.691e-05 | 0.4613 | Bacteria | Proteobacteria |
| Otu00024 | -2.904e-06 | 0.36 | Bacteria | Bacteroidetes |
| Otu00057 | -2.992e-05 | 0.009397 | Bacteria | Proteobacteria |
| Otu00138 | -3.378e-05 | 0.01573 | Bacteria | Firmicutes |
| Otu00169 | -1.041e-05 | 0.3399 | Bacteria | Bacteria\_unclassified |
| Otu01010 | -3.383e-08 | 0.6043 | Bacteria | Actinobacteria |

Table continues below

|  |  |
| --- | --- |
| Class | Order |
| Betaproteobacteria | Burkholderiales |
| Betaproteobacteria | Burkholderiales |
| Bacteroidetes\_unclassified | Bacteroidetes\_unclassified |
| Gammaproteobacteria | Methylococcales |
| Bacilli | Bacillales |
| Bacteria\_unclassified | Bacteria\_unclassified |
| Actinobacteria | Actinomycetales |

|  |  |
| --- | --- |
| Family | Genus |
| Burkholderiaceae | Polynucleobacter |
| Alcaligenaceae | Alcaligenaceae\_unclassified |
| Bacteroidetes\_unclassified | Bacteroidetes\_unclassified |
| Methylococcaceae | Methylococcaceae\_unclassified |
| Bacillaceae\_1 | Bacillus |
| Bacteria\_unclassified | Bacteria\_unclassified |
| Dermabacteraceae | Brachybacterium |

pander(nonsoil.core.increasing, caption = "Core taxa not found in soils that get more common along the transect.")

Core taxa not found in soils that get more common along the transect. (continued below)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| OTU | slope | pval | Domain | Phylum |
| Otu00004 | 0.0001338 | 1.729e-05 | Bacteria | Actinobacteria |
| Otu00008 | 3.29e-05 | 0.02679 | Bacteria | Actinobacteria |
| Otu00015 | 0.0001364 | 0.0003654 | Bacteria | Actinobacteria |
| Otu00016 | 5.12e-05 | 0.002128 | Bacteria | Actinobacteria |
| Otu00025 | 4.605e-05 | 0.006738 | Bacteria | Actinobacteria |
| Otu00038 | 4.534e-05 | 0.0001784 | Bacteria | Actinobacteria |
| Otu00040 | 3.723e-05 | 2.428e-05 | Bacteria | Proteobacteria |
| Otu00071 | 4.773e-05 | 0.000442 | Bacteria | Planctomycetes |
| Otu00079 | 8.096e-06 | 0.001572 | Bacteria | Bacteroidetes |
| Otu00080 | 1.597e-05 | 0.1571 | Bacteria | Bacteroidetes |
| Otu00118 | 6.543e-06 | 0.03723 | Bacteria | Actinobacteria |
| Otu00156 | 8.8e-06 | 0.002755 | Bacteria | Bacteria\_unclassified |

Table continues below

|  |  |
| --- | --- |
| Class | Order |
| Actinobacteria | Actinomycetales |
| Actinobacteria | Actinomycetales |
| Actinobacteria | Actinobacteria\_unclassified |
| Actinobacteria | Actinomycetales |
| Actinobacteria | Actinomycetales |
| Actinobacteria | Actinomycetales |
| Alphaproteobacteria | Rhodospirillales |
| Planctomycetia | Planctomycetales |
| Bacteroidetes\_unclassified | Bacteroidetes\_unclassified |
| Flavobacteriia | Flavobacteriales |
| Actinobacteria | Actinobacteria\_unclassified |
| Bacteria\_unclassified | Bacteria\_unclassified |

|  |  |
| --- | --- |
| Family | Genus |
| Actinomycetales\_unclassified | Actinomycetales\_unclassified |
| Actinomycetales\_unclassified | Actinomycetales\_unclassified |
| Actinobacteria\_unclassified | Actinobacteria\_unclassified |
| Microbacteriaceae | Microbacteriaceae\_unclassified |
| Microbacteriaceae | Microbacteriaceae\_unclassified |
| Actinomycetales\_unclassified | Actinomycetales\_unclassified |
| Acetobacteraceae | Roseomonas |
| Planctomycetaceae | Planctomycetaceae\_unclassified |
| Bacteroidetes\_unclassified | Bacteroidetes\_unclassified |
| Flavobacteriaceae | Flavobacterium |
| Actinobacteria\_unclassified | Actinobacteria\_unclassified |
| Bacteria\_unclassified | Bacteria\_unclassified |

pander(soil.core.decreasing, caption = "Core taxa found in soils that get rarer along the transect.")

Core taxa found in soils that get rarer along the transect. (continued below)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| OTU | slope | pval | Domain | Phylum |
| Otu00009 | -5.115e-05 | 0.02741 | Bacteria | Proteobacteria |
| Otu00010 | -4.281e-05 | 0.5552 | Bacteria | Proteobacteria |
| Otu00011 | -1.928e-05 | 0.6028 | Bacteria | Proteobacteria |
| Otu00018 | -4.637e-05 | 0.02104 | Bacteria | Proteobacteria |
| Otu00022 | -2.499e-05 | 0.1178 | Bacteria | Verrucomicrobia |
| Otu00028 | -3.043e-05 | 0.02348 | Bacteria | Proteobacteria |
| Otu00030 | -2.222e-06 | 0.2752 | Bacteria | Actinobacteria |
| Otu00039 | -8.511e-06 | 0.1793 | Bacteria | Proteobacteria |
| Otu00045 | -7.99e-06 | 0.5274 | Bacteria | Proteobacteria |
| Otu00059 | -6.488e-05 | 0.02525 | Bacteria | Actinobacteria |
| Otu00065 | -5.535e-05 | 0.02097 | Bacteria | Bacteroidetes |
| Otu00072 | -1.884e-05 | 0.09145 | Bacteria | Proteobacteria |
| Otu00077 | -5.843e-05 | 0.0117 | Bacteria | Bacteroidetes |
| Otu00086 | -1.26e-05 | 0.0353 | Bacteria | Proteobacteria |
| Otu00094 | -2.214e-05 | 0.03137 | Bacteria | Proteobacteria |
| Otu00095 | -3.555e-05 | 0.03573 | Bacteria | Proteobacteria |
| Otu00170 | -2.475e-05 | 0.02842 | Bacteria | Bacteroidetes |
| Otu00545 | -1.25e-06 | 0.0273 | Bacteria | Actinobacteria |

Table continues below

|  |  |
| --- | --- |
| Class | Order |
| Gammaproteobacteria | Pseudomonadales |
| Proteobacteria\_unclassified | Proteobacteria\_unclassified |
| Betaproteobacteria | Betaproteobacteria\_unclassified |
| Gammaproteobacteria | Pseudomonadales |
| Opitutae | Opitutae\_unclassified |
| Gammaproteobacteria | Pseudomonadales |
| Actinobacteria | Actinomycetales |
| Betaproteobacteria | Burkholderiales |
| Betaproteobacteria | Burkholderiales |
| Actinobacteria | Actinomycetales |
| Sphingobacteriia | Sphingobacteriales |
| Alphaproteobacteria | Sphingomonadales |
| Flavobacteriia | Flavobacteriales |
| Alphaproteobacteria | Rhizobiales |
| Betaproteobacteria | Burkholderiales |
| Betaproteobacteria | Burkholderiales |
| Sphingobacteriia | Sphingobacteriales |
| Actinobacteria | Solirubrobacterales |

|  |  |
| --- | --- |
| Family | Genus |
| Pseudomonadaceae | Pseudomonas |
| Proteobacteria\_unclassified | Proteobacteria\_unclassified |
| Betaproteobacteria\_unclassified | Betaproteobacteria\_unclassified |
| Pseudomonadaceae | Pseudomonas |
| Opitutae\_unclassified | Opitutae\_unclassified |
| Pseudomonadaceae | Pseudomonas |
| Micrococcaceae | Micrococcus |
| Comamonadaceae | Comamonas |
| Oxalobacteraceae | Oxalobacteraceae\_unclassified |
| Micrococcaceae | Arthrobacter |
| Sphingobacteriaceae | Pedobacter |
| Sphingomonadaceae | Sphingomonas |
| Flavobacteriaceae | Flavobacterium |
| Bradyrhizobiaceae | Bradyrhizobium |
| Oxalobacteraceae | Duganella |
| Comamonadaceae | Comamonadaceae\_unclassified |
| Sphingobacteriaceae | Sphingobacteriaceae\_unclassified |
| Solirubrobacteraceae | Solirubrobacter |

pander(soil.core.increasing, caption = "Core taxa found in soils that get more common along the transect.")

Core taxa found in soils that get more common along the transect. (continued below)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| OTU | slope | pval | Domain | Phylum |
| Otu00001 | 1.437e-05 | 0.07357 | Bacteria | Proteobacteria |
| Otu00002 | 0.0002104 | 0.002241 | Bacteria | Actinobacteria |
| Otu00003 | 9.845e-05 | 0.006345 | Bacteria | Verrucomicrobia |
| Otu00005 | 3.593e-05 | 0.01749 | Bacteria | Bacteroidetes |
| Otu00006 | 6.515e-06 | 0.1629 | Bacteria | Bacteroidetes |
| Otu00012 | 7.565e-06 | 0.09337 | Bacteria | Proteobacteria |
| Otu00014 | 8.415e-05 | 0.0007944 | Bacteria | Actinobacteria |
| Otu00023 | 3.479e-07 | 0.7837 | Bacteria | Proteobacteria |
| Otu00029 | 3.301e-05 | 0.004547 | Bacteria | Actinobacteria |
| Otu00032 | 3.59e-06 | 0.8316 | Bacteria | Bacteroidetes |
| Otu00033 | 9.093e-06 | 0.7077 | Bacteria | Proteobacteria |

Table continues below

|  |  |
| --- | --- |
| Class | Order |
| Betaproteobacteria | Burkholderiales |
| Actinobacteria | Actinomycetales |
| Spartobacteria | Spartobacteria\_unclassified |
| Sphingobacteriia | Sphingobacteriales |
| Sphingobacteriia | Sphingobacteriales |
| Betaproteobacteria | Burkholderiales |
| Actinobacteria | Actinomycetales |
| Gammaproteobacteria | Pseudomonadales |
| Actinobacteria | Actinomycetales |
| Bacteroidetes\_unclassified | Bacteroidetes\_unclassified |
| Alphaproteobacteria | Rhizobiales |

|  |  |
| --- | --- |
| Family | Genus |
| Comamonadaceae | Comamonadaceae\_unclassified |
| Actinomycetales\_unclassified | Actinomycetales\_unclassified |
| Spartobacteria\_unclassified | Spartobacteria\_unclassified |
| Chitinophagaceae | Sediminibacterium |
| Saprospiraceae | Saprospiraceae\_unclassified |
| Comamonadaceae | Comamonadaceae\_unclassified |
| Actinomycetales\_unclassified | Actinomycetales\_unclassified |
| Moraxellaceae | Acinetobacter |
| Actinomycetales\_unclassified | Actinomycetales\_unclassified |
| Bacteroidetes\_unclassified | Bacteroidetes\_unclassified |
| Rhizobiales\_unclassified | Rhizobiales\_unclassified |

# p1 <- as.data.frame(OTUsREL[,nonsoil.core.increasing$OTU]) %>%   
# rownames\_to\_column("sampleID") %>%   
# left\_join(rownames\_to\_column(design, "sampleID")) %>%   
# gather(OTU, rel\_abund, -station, -molecule, -type, -distance, -sampleID) %>%   
# filter(molecule == "DNA") %>% left\_join(OTU.tax) %>%   
# mutate(taxon = paste(Phylum, Class, Order, Family, Genus)) %>%   
# ggplot(aes(x = distance, y = rel\_abund, group = OTU)) +   
# #geom\_point(alpha = 0.5) +   
# geom\_line(stat = "smooth", alpha = 0.5, size = 1,  
# color = "black", method = "loess", span = 1, se = FALSE) +   
# scale\_x\_reverse() +  
# scale\_y\_log10(labels = scales::scientific) +  
# theme(legend.position = "none") +  
# guides(color = guide\_legend(ncol = 1)) +  
# labs(x = "",  
# y = "Relative Abundance",  
# title = "Absent from soil and significantly increasing")  
#   
# p2 <- as.data.frame(OTUsREL[,soil.core.increasing$OTU]) %>%   
# rownames\_to\_column("sampleID") %>%   
# left\_join(rownames\_to\_column(design, "sampleID")) %>%   
# gather(OTU, rel\_abund, -station, -molecule, -type, -distance, -sampleID) %>%   
# filter(molecule == "DNA") %>% left\_join(OTU.tax) %>%   
# mutate(taxon = paste(Class, Order)) %>%   
# ggplot(aes(x = distance, y = rel\_abund, group = OTU)) +   
# #geom\_point(alpha = 0.5) +   
# geom\_line(stat = "smooth", alpha = 0.5, size = 1,  
# color = "black", method = "loess", span = 1, se = FALSE) +   
# scale\_x\_reverse() +   
# scale\_y\_log10(labels = scales::scientific) +  
# theme(legend.position = "none") +  
# guides(color = guide\_legend(ncol = 1)) +  
# labs(x = "",  
# y = "Relative Abundance",  
# title = "Present in soil and significantly increasing")  
#   
# p3 <- as.data.frame(OTUsREL[,soil.core.decreasing$OTU]) %>%   
# rownames\_to\_column("sampleID") %>%   
# left\_join(rownames\_to\_column(design, "sampleID")) %>%   
# gather(OTU, rel\_abund, -station, -molecule, -type, -distance, -sampleID) %>%   
# filter(molecule == "DNA") %>% left\_join(OTU.tax) %>%   
# mutate(taxon = paste(Class, Order)) %>%   
# ggplot(aes(x = distance, y = rel\_abund, group = OTU)) +   
# #geom\_point(alpha = 0.5) +   
# geom\_line(stat = "smooth", alpha = 0.5, size = 1,  
# color = "black", method = "loess", span = 1, se = FALSE) +   
# scale\_x\_reverse() +  
# scale\_y\_log10(labels = scales::scientific) +  
# theme(legend.position = "none") +  
# guides(color = guide\_legend(ncol = 1)) +  
# labs(x = "Reservoir Transect (m)",  
# y = "Relative Abundance",  
# title = "Present in soil and significantly decreasing")  
#   
# cowplot::plot\_grid(p1, p2, p3, align = "hv", labels = "AUTO", ncol = 1)  
  
df1 <- as.data.frame(OTUsREL[,nonsoil.core.increasing$OTU]) %>%   
 rownames\_to\_column("sampleID") %>%   
 left\_join(rownames\_to\_column(design, "sampleID")) %>%   
 gather(OTU, rel\_abund, -station, -molecule, -type, -distance, -sampleID) %>%   
 filter(molecule == "DNA") %>% left\_join(OTU.tax) %>%   
 mutate(soils = "Absent from soils", change = "Increasing")

## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector

n1 <- length(unique(df1$OTU))  
  
df2 <- as.data.frame(OTUsREL[,soil.core.increasing$OTU]) %>%   
 rownames\_to\_column("sampleID") %>%   
 left\_join(rownames\_to\_column(design, "sampleID")) %>%   
 gather(OTU, rel\_abund, -station, -molecule, -type, -distance, -sampleID) %>%   
 filter(molecule == "DNA") %>% left\_join(OTU.tax) %>%   
 mutate(soils = "Present in soils", change = "Increasing")

## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector

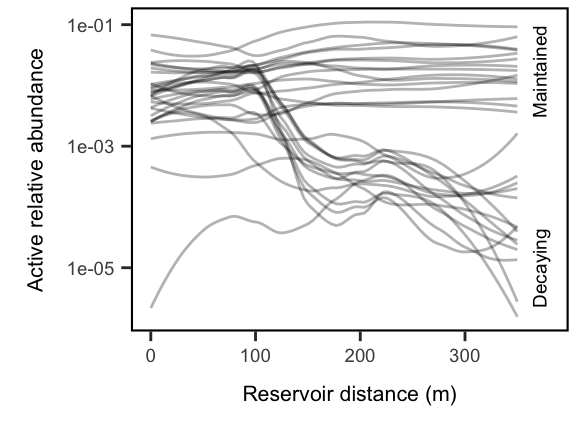
n2 <- length(unique(df2$OTU))  
  
df3 <- as.data.frame(OTUsREL[,soil.core.decreasing$OTU]) %>%   
 rownames\_to\_column("sampleID") %>%   
 left\_join(rownames\_to\_column(design, "sampleID")) %>%   
 gather(OTU, rel\_abund, -station, -molecule, -type, -distance, -sampleID) %>%   
 filter(molecule == "DNA") %>% left\_join(OTU.tax) %>%   
 mutate(soils = "Present in soils", change = "Decreasing")

## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector

n3 <- length(unique(df3$OTU))  
  
df4 <- as.data.frame(OTUsREL[,nonsoil.core.decreasing$OTU]) %>%   
 rownames\_to\_column("sampleID") %>%   
 left\_join(rownames\_to\_column(design, "sampleID")) %>%   
 gather(OTU, rel\_abund, -station, -molecule, -type, -distance, -sampleID) %>%   
 filter(molecule == "DNA") %>% left\_join(OTU.tax) %>%   
 mutate(soils = "Absent from soils", change = "Decreasing")

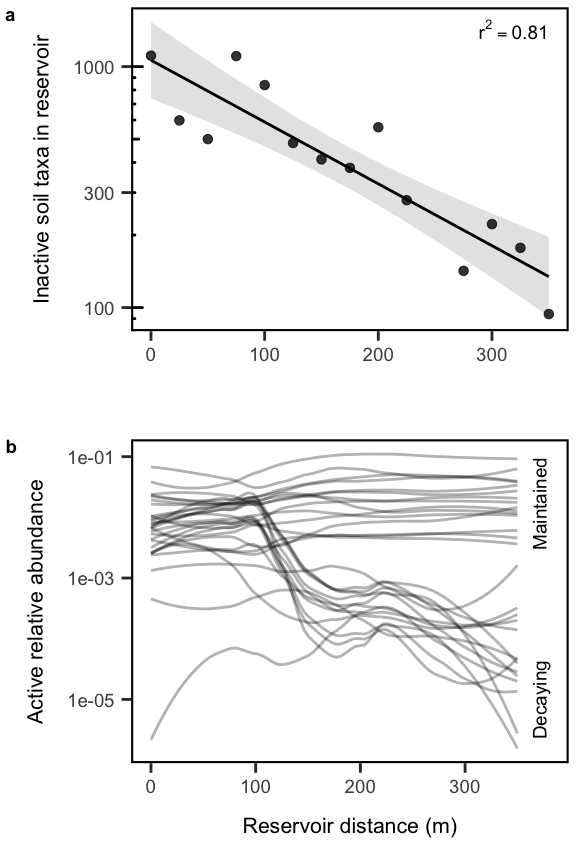
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector

n4 <- length(unique(df4$OTU))  
  
  
df.plot <- as\_tibble(rbind.data.frame(df1, df2, df3, df4)) %>% filter(type == "water")  
  
taxon\_fate.plot <- df.plot %>% mutate(rel\_abund = ifelse(rel\_abund == 0, 1e-6, rel\_abund)) %>%   
 filter(soils == "Present in soils") %>%   
 #mutate(change = ifelse(change == "Increasing",   
 # paste0("Increasing (n = ", n2,")"),  
 # paste0("Decreasing (n = ", n3,")"))) %>%   
 ggplot(aes(x = distance, y = rel\_abund, group = OTU)) +   
 #geom\_jitter(alpha = 0.15) +   
 geom\_line(stat = "smooth", alpha = 0.3, size = 1,  
 method = "loess", span = .7, se = FALSE) +  
 scale\_y\_log10(labels = scales::scientific) +  
 scale\_x\_continuous(limits = c(0,380)) +  
 #theme(legend.position = "none") +  
 #guides(color = guide\_legend(ncol = 1)) +  
 labs(x = "Reservoir distance (m)",  
 y = "Active relative abundance") +  
 annotate("text", x = 365, y = 1e-1, size = 5, hjust = 1, vjust = 1, angle = 90,  
 label = "Maintained") +  
 annotate("text", x = 365, y = 1e-5, size = 5, hjust = 0.5, vjust = 1, angle = 90,  
 label = "Decaying") +  
 ggsave("figures/taxa\_origins.pdf")  
 taxon\_fate.plot



# how much do the different core components contribute to total abundances  
in.lake.core.soil.REL <- rowSums(in.lake.core.from.soils) / rowSums(w.dna)  
in.lake.core.water.REL <- rowSums(in.lake.core.not.soils) / rowSums(w.dna)

plot\_grid(transient.plot + labs(x = ""),  
 taxon\_fate.plot,  
 align = "hv", axis = "rltb",  
 labels = "auto",  
 ncol = 1) +  
 ggsave("figures/fate\_panel.pdf")



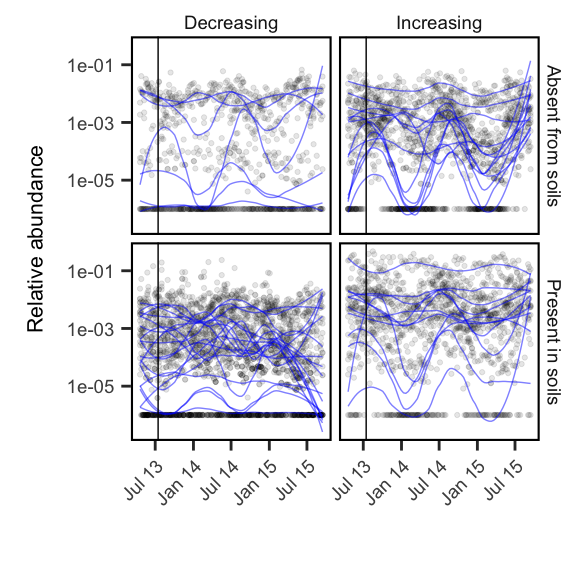
# soil.mods <- t(soil.core.mods) %>% as.data.frame()  
# soil.mods$habitat <- "Present in soils"  
# soil.mods <- soil.mods %>% rownames\_to\_column(var = "OTU")  
# nonsoil.mods <- t(nonsoil.core.mods) %>% as.data.frame()  
# nonsoil.mods$habitat <- "Absent from soils"  
# nonsoil.mods <- nonsoil.mods %>% rownames\_to\_column(var = "OTU")  
# rbind.data.frame(soil.mods, nonsoil.mods) %>%   
# filter(pval < 0.05) %>%   
# ggplot(aes(x = -slope, fill = habitat, color = habitat)) +  
# geom\_line(stat = "density", alpha = 0.5, adjust = .8) +  
# geom\_density(color = NA, adjust = .8, alpha = 0.2)

# Are the “persistent” reservoir taxa really representative? Look over time…

total.OTUs <- read.otu(shared = shared, cutoff = "0.03") # 97% Similarity  
  
# Import Taxonomy  
total.OTU.tax <- read.tax(taxonomy = taxon, format = "rdp")  
  
# Subset to just the time series sites  
UL.ts.OTUs <- total.OTUs[str\_which(rownames(total.OTUs), "UL"),]  
  
# make sure OTU table matches up with design order  
UL.ts.design <- read\_csv("data/UL\_timeseries\_design.csv")  
UL.ts.OTUs <- UL.ts.OTUs[match(UL.ts.design$sample.name, rownames(UL.ts.OTUs)),]  
UL.ts.OTUs.RNA <- decostand(UL.ts.OTUs[which(UL.ts.design$sample.type == "RNA"),], method = "total")  
UL.ts.OTUs.DNA <- decostand(UL.ts.OTUs[which(UL.ts.design$sample.type == "DNA"),], method = "total")  
  
  
env.ts.data <- read.table("data/ul-seedbank.env.txt", sep="\t", header=TRUE)  
env.ts.data$date <- as.Date(parse\_date\_time(env.ts.data$date, "m d y"))  
env.ts.data$doc[which(env.ts.data$doc == "\*\*")] <- NA  
env.ts.data$doc <- as.numeric(env.ts.data$doc)  
summary(env.ts.data)

## sample.id date temp spc   
## Min. : 1.00 Min. :2013-04-19 Min. : 2.21 Min. :0.3300   
## 1st Qu.: 31.75 1st Qu.:2013-11-20 1st Qu.: 5.50 1st Qu.:0.4600   
## Median : 62.50 Median :2014-06-23 Median :17.73 Median :0.5320   
## Mean : 62.50 Mean :2014-06-24 Mean :16.18 Mean :0.5172   
## 3rd Qu.: 93.25 3rd Qu.:2015-01-25 3rd Qu.:25.05 3rd Qu.:0.5660   
## Max. :124.00 Max. :2015-09-14 Max. :29.77 Max. :0.6700   
## NA's :2 NA's :2   
## oxygen salinity secchi ph   
## Min. : 1.870 Min. :0.1500 Min. :0.200 Min. : 6.890   
## 1st Qu.: 5.237 1st Qu.:0.2200 1st Qu.:1.200 1st Qu.: 7.920   
## Median : 8.355 Median :0.2550 Median :1.600 Median : 8.415   
## Mean : 8.961 Mean :0.2487 Mean :1.668 Mean : 8.567   
## 3rd Qu.:10.178 3rd Qu.:0.2700 3rd Qu.:2.200 3rd Qu.: 9.123   
## Max. :22.240 Max. :0.3200 Max. :3.600 Max. :10.860   
## NA's :2 NA's :2 NA's :1 NA's :2   
## chla tp tn doc   
## Min. : 0.92 Min. : 8.26 Min. : 0.407 Min. : 2.00   
## 1st Qu.: 12.63 1st Qu.: 26.30 1st Qu.: 0.882 1st Qu.: 32.25   
## Median : 37.67 Median : 34.85 Median : 1.210 Median : 61.50   
## Mean : 79.25 Mean : 84.25 Mean : 1.889 Mean : 61.57   
## 3rd Qu.:121.31 3rd Qu.: 47.95 3rd Qu.: 1.490 3rd Qu.: 90.75   
## Max. :523.56 Max. :3200.00 Max. :42.600 Max. :121.00   
## NA's :2 NA's :2 NA's :3 NA's :2   
## orp air.temp   
## Min. :-41.800 Min. :-11.60   
## 1st Qu.: 9.325 1st Qu.: 7.00   
## Median : 21.700 Median : 18.50   
## Mean : 50.507 Mean : 15.57   
## 3rd Qu.:104.975 3rd Qu.: 24.00   
## Max. :225.200 Max. : 32.00   
## NA's :68 NA's :2

UL.ts.design <- left\_join(UL.ts.design, env.ts.data[,c("sample.id", "date")])  
env.ts.data <- env.ts.data[-which(!(env.ts.data$date %in% UL.ts.design$date)),]  
  
OTUs.in.core <- UL.ts.OTUs.RNA[, which(colnames(UL.ts.OTUs) %in% df.plot$OTU)]  
cbind.data.frame(UL.ts.design[which(UL.ts.design$sample.type == "RNA"),], OTUs.in.core) %>% as\_tibble() %>%   
 gather(-sample.name, -sample.type, -sample.id, -date, key = OTU, value = rel\_abund) %>%   
 mutate(soils = ifelse(OTU %in% unique(c(df2$OTU, df3$OTU)),   
 "Present in soils", "Absent from soils")) %>%   
 mutate(change = ifelse(OTU %in% unique(c(df3$OTU, df4$OTU)),   
 "Decreasing", "Increasing")) %>%   
 mutate(rel\_abund = ifelse(rel\_abund == 0, 1e-6, rel\_abund)) %>%   
 ggplot(aes(x = date, y = rel\_abund, group = OTU)) +  
 geom\_point(alpha = .1) +   
 geom\_line(stat = "smooth", method = "loess", color = "blue",  
 alpha = 0.5, span = .5, se = F) +  
 geom\_vline(aes(xintercept = as\_date("2013-07-15"))) +  
 scale\_y\_log10() +   
 scale\_x\_date(labels = scales::date\_format(format = "%b %y")) +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) +  
 facet\_grid(soils ~ change) +  
 labs(x = "",  
 y = "Relative abundance")

 Many of them do appear to track the seasons quite well, suggesting there could be a seasonality component to the role of terrestrial inputs into the reservoir.

## Ecosystem functions

metab <- read.table("data/res.grad.metab.txt", sep="\t", header=TRUE)  
colnames(metab) <- c("dist", "BP", "BR")  
BGE <- round((metab$BP/(metab$BP + metab$BR)),3)  
metab <- cbind(metab, BGE)  
metab <- metab[-c(16:18),]  
metab$dist <- 350 - metab$dist  
  
  
# Quadratic regression for BP  
dist <- metab$dist  
dist2 <- metab$dist^2  
BP.fit <- lm(metab$BP ~ dist + dist2)  
BP.R2 <- round(summary(BP.fit)$r.squared, 2)  
  
# Simple linear regression for BR  
BR.fit <- lm(metab$BR ~ metab$dist)  
BR.R2 <- round(summary(BR.fit)$r.squared, 2)  
BR.int <- BR.fit$coefficients[1]  
BR.slp <- BR.fit$coefficients[2]  
  
# Simple linear regression for BGE  
BGE.fit <- lm(metab$BGE ~ metab$dist)  
BGE.R2 <- round(summary(BGE.fit)$r.squared, 2)  
BGE.int <- BGE.fit$coefficients[1]  
BGE.slp <- BGE.fit$coefficients[2]  
  
BP.R2

## [1] 0.36

BR.R2

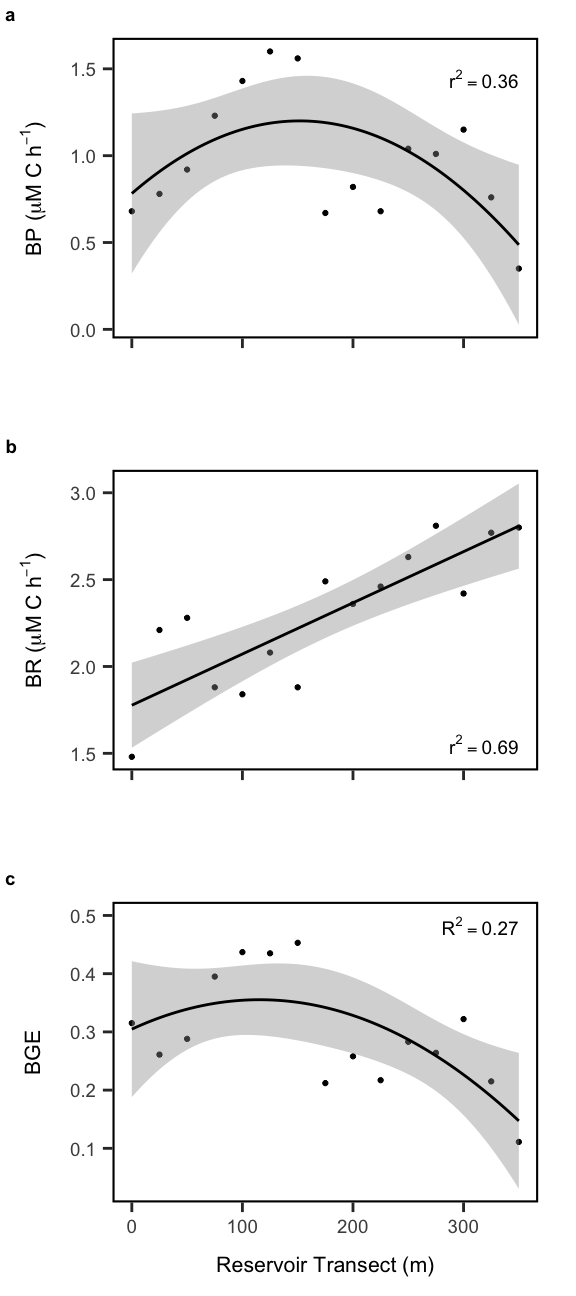
## [1] 0.69

BGE.R2

## [1] 0.27

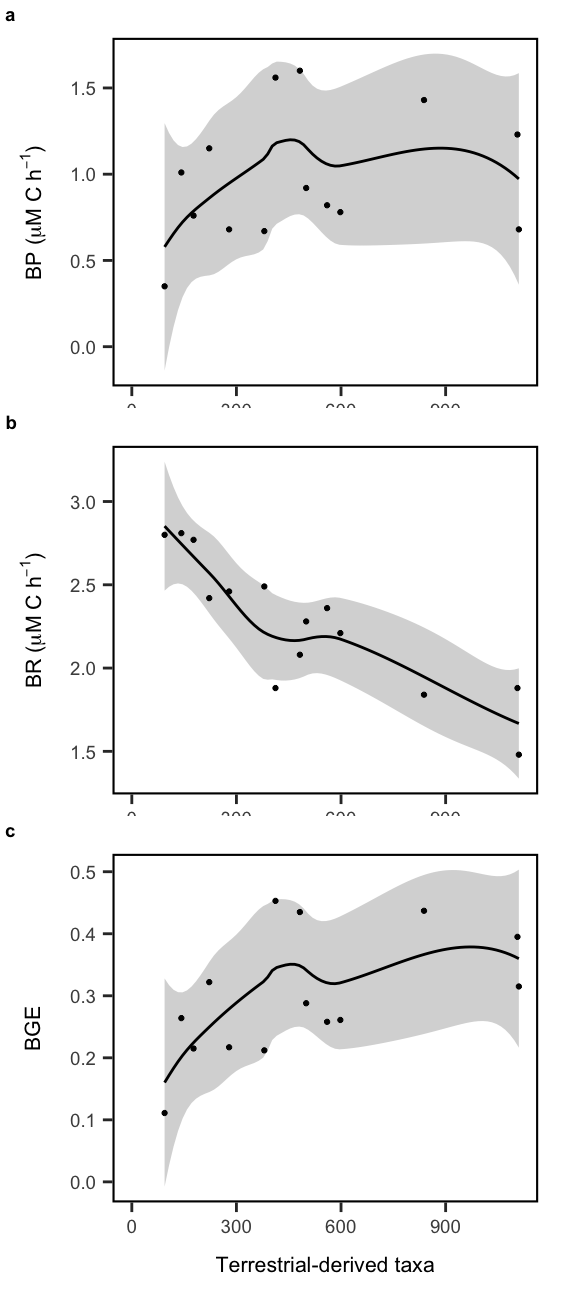
BP.plot <- ggplot(metab, aes(x = dist, y = BP)) +   
 geom\_point() +   
 geom\_smooth(method = "lm", formula = y ~ x + I(x^2), color = "black") +  
 annotate(geom = "text", x = 350, y = 1.5, size = 5, hjust = 1, vjust = 1,  
 label = paste0("r^2== ",BP.R2), parse = T) +  
 labs(y = expression(paste('BP (', mu ,'M C h'^-1\* ')')),   
 x = "Reservoir Transect (m)")  
BR.plot <- ggplot(metab, aes(x = dist, y = BR)) +   
 geom\_point() +   
 geom\_smooth(method = "lm", formula = y ~ x, color = "black") +   
 annotate("text", x = 350, y = 1.5, size = 5, hjust = 1, vjust = 0,  
 label = paste0("r^2== ",BR.R2), parse = T ) +  
 labs(y = expression(paste('BR (', mu ,'M C h'^-1\* ')')),   
 x = "Reservoir Transect (m)")  
BGE.plot <- ggplot(metab, aes(x = dist, y = BGE)) +   
 geom\_point() +   
 geom\_smooth(method = "lm", formula = y ~ x + I(x^2), color = "black") +  
 annotate("text", x = 350, y = .5, size = 5, hjust = 1, vjust = 1,  
 label = paste0("R^2== ",BGE.R2), parse = T ) +  
 labs(y = "BGE",   
 x = "Reservoir Transect (m)")

plot\_grid(BP.plot + theme(axis.title.x = element\_blank(), axis.text.x = element\_blank(),   
 plot.margin = unit(c(1, 1, -1, 0), "cm")),   
 BR.plot + theme(axis.title.x = element\_blank(), axis.text.x = element\_blank(),  
 plot.margin = unit(c(-1, 1, -1, 0), "cm")),   
 BGE.plot + theme(plot.margin = unit(c(-1, 1, 0, 0), "cm")),   
 align = "hv", ncol = 1, labels = "auto")



## Relation of ecosystem functions and community structure

metab.joined <- cbind.data.frame(design.dna, metab[-5,])  
  
transient.metabolism <- cbind.data.frame(transients = terr.rich, metab.joined)   
  
p1 <- transient.metabolism %>%   
 ggplot(aes(x=transients, y = BP)) +  
 geom\_smooth(color = "black") +  
 geom\_point() +   
 scale\_x\_continuous(limits = c(0, NA)) +  
 labs(x = "Terrestrial-derived taxa",  
 y = expression(paste('BP (', mu ,'M C h'^-1\* ')'))) +  
 theme(axis.title.x = element\_blank(),   
 plot.margin = unit(c(1, 1, 0, 0), "cm"))  
p2 <- transient.metabolism %>%   
 ggplot(aes(x=transients, y = BR)) +  
 geom\_smooth(color = "black") +  
 geom\_point() +   
 scale\_x\_continuous(limits = c(0, NA)) +  
 labs(x = "Terrestrial-derived taxa",  
 y = expression(paste('BR (', mu ,'M C h'^-1\* ')'))) +  
 theme(axis.title.x = element\_blank(),  
 plot.margin = unit(c(0, 1, 0, 0), "cm"))  
p3 <- transient.metabolism %>%   
 ggplot(aes(x=transients, y = BGE)) +  
 geom\_smooth(color = "black") +  
 geom\_point() +   
 scale\_x\_continuous(limits = c(0, NA)) +  
 labs(x = "Terrestrial-derived taxa") +  
 theme(plot.margin = unit(c(0, 1, 0, 0), "cm"))  
  
plot\_grid(p1, NULL, p2, NULL, p3,   
 rel\_heights = c(1, -.15, 1, -.15, 1), align = "hv",   
 ncol = 1, labels = c("a", "NULL", "b", "NULL", "c")) +  
 ggsave("figures/functions.pdf")



# identify otus in soil samples and lake samples  
in.soil <- OTUs[, which(colSums(OTUs[c(1:3),]) > 0 )]  
  
# isolate just the rna water samples and convert to presence-absence  
in.lake.rna <- OTUs[which(design$molecule == "RNA" & design$type == "water"), ]  
in.lake.rna.pa <- (in.lake.rna > 0) \* 1  
  
threshlist <- c(.3, .4, .5, .6, .7, .8, .9)  
df.plot <- data.frame()  
for(thresh in threshlist){  
 # define the 'core' taxa as otus present in 50% of samples  
in.lake.core <- w.dna[, which((colSums(in.lake.rna.pa) / nrow(in.lake.rna.pa)) >= thresh)]  
  
# of the core, how many are also in the soil samples?  
in.lake.core.from.soils <- in.lake.core[, intersect(colnames(in.lake.core), colnames(in.soil))]  
  
# of the core which are not in the soil samples  
in.lake.core.not.soils <- in.lake.core[, setdiff(colnames(in.lake.core), colnames(in.soil))]  
  
# Find the relative abundance of the core taxa and prepare data frame to plot  
in.lake.core.from.soils.REL <- in.lake.core.from.soils / rowSums(w.dna)  
  
in.soil.to.plot <- as.data.frame(in.lake.core.from.soils.REL) %>%   
 rownames\_to\_column("sample\_ID") %>%   
 gather(otu\_id, rel\_abundance, -sample\_ID) %>%   
 left\_join(rownames\_to\_column(design.dna, "sample\_ID")) %>%   
 add\_column(found = "soils")  
  
in.lake.core.not.soils.REL <- in.lake.core.not.soils / rowSums(w.dna)  
  
in.lake.to.plot <- as.data.frame(in.lake.core.not.soils.REL) %>%   
 rownames\_to\_column("sample\_ID") %>%   
 gather(otu\_id, rel\_abundance, -sample\_ID) %>%   
 left\_join(rownames\_to\_column(design.dna, "sample\_ID")) %>%   
 add\_column(found = "lake")  
  
# model distance effect on rel abundance to get slope and pval  
soil.core.mods <- apply(in.lake.core.from.soils.REL, MARGIN = 2,   
 FUN = function(x) summary(lm(x ~ design.dna$distance))$coefficients[2,c(1,4)])  
rownames(soil.core.mods) <- c("slope", "pval")  
  
# classify otus as significantly increasing or decreasing along reservoir  
soil.core.decreasing <- as.data.frame(t(soil.core.mods)) %>%   
 rownames\_to\_column("OTU") %>%   
 filter(slope < 0) %>% # rel abund decreases toward dam  
 left\_join(OTU.tax)  
soil.core.increasing <- as.data.frame(t(soil.core.mods)) %>%   
 rownames\_to\_column("OTU") %>%   
 filter(slope > 0) %>% # rel abund increases toward dam  
 left\_join(OTU.tax)  
  
nonsoil.core.mods <- apply(in.lake.core.not.soils.REL, MARGIN = 2,   
 FUN = function(x) summary(lm(x ~ design.dna$distance))$coefficients[2,c(1,4)])  
rownames(nonsoil.core.mods) <- c("slope", "pval")  
nonsoil.core.decreasing <- as.data.frame(t(nonsoil.core.mods)) %>%   
 rownames\_to\_column("OTU") %>%   
 filter(slope < 0) %>% # rel abund decreases toward dam  
 left\_join(OTU.tax)  
nonsoil.core.increasing <- as.data.frame(t(nonsoil.core.mods)) %>%   
 rownames\_to\_column("OTU") %>%   
 filter(slope > 0) %>% # rel abund increases toward dam  
 left\_join(OTU.tax)  
  
df1 <- as.data.frame(OTUsREL[,nonsoil.core.increasing$OTU]) %>%   
 rownames\_to\_column("sampleID") %>%   
 left\_join(rownames\_to\_column(design, "sampleID")) %>%   
 gather(OTU, rel\_abund, -station, -molecule, -type, -distance, -sampleID) %>%   
 filter(molecule == "DNA") %>% left\_join(OTU.tax) %>%   
 mutate(soils = "Absent from soils", change = "Increasing")  
n1 <- length(unique(df1$OTU))  
  
df2 <- as.data.frame(OTUsREL[,soil.core.increasing$OTU]) %>%   
 rownames\_to\_column("sampleID") %>%   
 left\_join(rownames\_to\_column(design, "sampleID")) %>%   
 gather(OTU, rel\_abund, -station, -molecule, -type, -distance, -sampleID) %>%   
 filter(molecule == "DNA") %>% left\_join(OTU.tax) %>%   
 mutate(soils = "Present in soils", change = "Increasing")  
n2 <- length(unique(df2$OTU))  
  
df3 <- as.data.frame(OTUsREL[,soil.core.decreasing$OTU]) %>%   
 rownames\_to\_column("sampleID") %>%   
 left\_join(rownames\_to\_column(design, "sampleID")) %>%   
 gather(OTU, rel\_abund, -station, -molecule, -type, -distance, -sampleID) %>%   
 filter(molecule == "DNA") %>% left\_join(OTU.tax) %>%   
 mutate(soils = "Present in soils", change = "Decreasing")  
n3 <- length(unique(df3$OTU))  
  
df4 <- as.data.frame(OTUsREL[,nonsoil.core.decreasing$OTU]) %>%   
 rownames\_to\_column("sampleID") %>%   
 left\_join(rownames\_to\_column(design, "sampleID")) %>%   
 gather(OTU, rel\_abund, -station, -molecule, -type, -distance, -sampleID) %>%   
 filter(molecule == "DNA") %>% left\_join(OTU.tax) %>%   
 mutate(soils = "Absent from soils", change = "Decreasing")  
n4 <- length(unique(df4$OTU))  
  
  
df.plot <- as\_tibble(rbind.data.frame(df1, df2, df3, df4)) %>%   
 mutate(thresh = thresh) %>% filter(type == "water") %>%   
 bind\_rows(df.plot)  
  
}

## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector  
  
## Warning: Column `OTU` joining character vector and factor, coercing into  
## character vector

taxon\_fate.plot <- df.plot %>% mutate(rel\_abund = ifelse(rel\_abund == 0, 1e-6, rel\_abund)) %>%   
 filter(soils == "Present in soils") %>%   
 #mutate(change = ifelse(change == "Increasing",   
 # paste0("Increasing (n = ", n2,")"),  
 # paste0("Decreasing (n = ", n3,")"))) %>%   
 ggplot(aes(x = distance, y = rel\_abund, group = OTU)) +   
 #geom\_jitter(alpha = 0.15) +   
 geom\_line(stat = "smooth", alpha = 0.3, size = .5,  
 method = "loess", span = .7, se = FALSE) +  
 scale\_y\_log10(labels = scales::scientific) +  
 scale\_x\_continuous(limits = c(0,380)) +  
 facet\_wrap(~thresh) +  
 #theme(legend.position = "none") +  
 #guides(color = guide\_legend(ncol = 1)) +  
 labs(x = "Reservoir distance (m)",  
 y = "Active relative abundance") +  
 # annotate("text", x = 365, y = 1e-1, size = 5, hjust = 1, vjust = 1, angle = 90,  
 # label = "Maintained") +  
 # annotate("text", x = 365, y = 1e-5, size = 5, hjust = 0.5, vjust = 1, angle = 90,  
 # label = "Decaying") +  
 ggsave("figures/taxa\_origins\_threshold.pdf", width = 8, height = 6, units = "in")  
taxon\_fate.plot

