

# eDNAexpeditions\_trialSamples\_analysis

Saara Suominen

4/18/2023

## eDNA expeditions trial samples

This document shows the overall results of the eDNA expeditions trial sample analysis.

The analysis consisted of a mixture of samples of

- eDNA filters
- control filters
- pcr controls

Which were first amplified in 3 different types of reactions:

- Mifish multiplex (all mifish and mimammal primers combined in one reaction)
- All multiplex (all primers combined in one reaction)
- All singleplex (all primers amplified separately, then combined for sequencing)

The samples are from the eDNA expeditions project, and are from three locations:

- Wadden Sea
- Brazil
- Australia

Random positive samples (PCR bands detected) from each of the countries were combined together for the analysis.

The primers that were used for the analysis were:

- Mifish-UE, Mimammal-UEB
- COI (Leray fragment)
- 16S rRNA for vertebrates
- Teleo 12S rRNA

The purpose of this analysis was to test the sample processing pipeline and to understand which reaction type works best with the samples.

All sequences were analysed with the PacMAN pipeline, with each primer set analysed separately. Cutadapt was used to search for the primer sequences and discard any reads where the primer sequences were not found. For the MifishUE and the Mimammal, the different primer sets in each category were combined as a consensus sequence. Still there is a large overlap within the sequences found by Mimammal and Mifish primers, as the primers are very similar. The overlap in ASVs is removed in the analysis stage.

The sequences were obtained with a paired run of a 150 bp Novaseq kit. Only paired sequences from the mifish and mimammal analyses were merged in the analysis. The COI and 16S rRNA markers were too long to be merged (313 bp and 250 bp respectively) as the length of good quality sequences for the 16S rRNA was also about 125 bp. For Teleo, it is a short read, but there were large differences in the amount of primers recognized in forward reads or reverse reads by cutadapt (possibly something to still check why). Cutadapt removes reads if the other pair is not found in paired-end mode, therefore the reads for teleo were run separately. In this case, each read (both forward and reverse), cover the full range of the marker.

A reference database built with the reference database creator either from ncbi or from the mitofish database (for the mifish primers), and cut to the specific primer region that was used. Remaining unknown sequences were searched against the full ncbi-nt database.

The sequence annotation in this first trial was not strict, with an 85% identity cutoff for both the targeted database file and blast. Therefore the annotation results need to be considered carefully.

## Upload phyloseq data

The PacMAN pipeline outputs the data in phyloseq format to make starting the analysis easy. However, these are not names that have been checked with the WoRMS database and therefore the DwC files should be used in the final analysis.

```
#Preparation of data

#First all data is uploaded from the phyloseq objects from the pacMAN pipeline:

MifishUE=readRDS("../eDNAexpeditions/Analysis_Trials/phyloseq_object_MiFishUE.rds"); MifishUE

## phyloseq-class experiment-level object
## otu_table()    OTU Table:          [ 10251 taxa and 45 samples ]
## sample_data()  Sample Data:       [ 45 samples by 22 sample variables ]
## tax_table()    Taxonomy Table:   [ 10251 taxa by 14 taxonomic ranks ]
## refseq()       DNAStringSet:     [ 10251 reference sequences ]

Mimammal=readRDS("../eDNAexpeditions/Analysis_Trials/phyloseq_object_MiMammal.rds"); Mimammal

## phyloseq-class experiment-level object
## otu_table()    OTU Table:          [ 10097 taxa and 45 samples ]
## sample_data()  Sample Data:       [ 45 samples by 22 sample variables ]
## tax_table()    Taxonomy Table:   [ 10097 taxa by 14 taxonomic ranks ]
## refseq()       DNAStringSet:     [ 10097 reference sequences ]

CO1=readRDS("../eDNAexpeditions/Analysis_Trials/phyloseq_object_COI.rds"); CO1

## phyloseq-class experiment-level object
## otu_table()    OTU Table:          [ 3443 taxa and 30 samples ]
## sample_data()  Sample Data:       [ 30 samples by 23 sample variables ]
## tax_table()    Taxonomy Table:   [ 3443 taxa by 14 taxonomic ranks ]
## refseq()       DNAStringSet:     [ 3443 reference sequences ]

Teleo=readRDS("../eDNAexpeditions/Analysis_Trials/phyloseq_object_teleoSsingle.rds"); Teleo
```

```

## phyloseq-class experiment-level object
## otu_table() OTU Table: [ 2321 taxa and 30 samples ]
## sample_data() Sample Data: [ 30 samples by 21 sample variables ]
## tax_table() Taxonomy Table: [ 2321 taxa by 14 taxonomic ranks ]
## refseq() DNAStringSet: [ 2321 reference sequences ]

rna16S=readRDS("../../.eDNAexpeditions/Analysis_Trials/phyloseq_object_16Svert.rds");rna16S

## phyloseq-class experiment-level object
## otu_table() OTU Table: [ 1004 taxa and 30 samples ]
## sample_data() Sample Data: [ 30 samples by 22 sample variables ]
## tax_table() Taxonomy Table: [ 1004 taxa by 14 taxonomic ranks ]
## refseq() DNAStringSet: [ 1004 reference sequences ]

pseqs=list(MifishUE, Mimammal, C01, Teleo, rna16S)
names(pseqs)=c("MifishUE", "Mimammal", "C01", "Teleo", "rna16S")

#Subset data for reads in different categories
fish_class=c("Actinopteri", "Chondrichthyes", "Coelacanthe", "Myxini", "Petromyzonti")

#To combine all data to one object, I will change the asv names to the ASV sequence again (so that ASVs
for (i in 1:length(pseqs)) {
  taxa_names(pseqs[[i]])<-pseqs[[i]]@tax_table[, "DNA_sequence"]
}

To avoid the issue of duplication on Mifish/Mimammal reads, keep only unique mimammal reads for merging
(and use the same sample names)

#Mimammal and mifish share a lot of their asvs. So I should remove the number of shared asvs from the t
length(intersect(pseqs$MifishUE@tax_table[, "DNA_sequence"], pseqs$Mimammal@tax_table[, "DNA_sequence"]))

## [1] 8727

shared_names=intersect(taxa_names(pseqs$MifishUE), taxa_names(pseqs$Mimammal))
unique_mimammal_names=taxa_names(pseqs$Mimammal) [!(taxa_names(pseqs$Mimammal)%in%shared_names)]

mimammal_unique = prune_taxa(unique_mimammal_names, pseqs$Mimammal)

pseqs_merged=merge_phyloseq(pseqs$MifishUE, mimammal_unique, pseqs$C01, pseqs$Teleo, pseqs$rna16S)
pseqs_merged

## phyloseq-class experiment-level object
## otu_table() OTU Table: [ 18389 taxa and 45 samples ]
## sample_data() Sample Data: [ 45 samples by 23 sample variables ]
## tax_table() Taxonomy Table: [ 18389 taxa by 14 taxonomic ranks ]
## refseq() DNAStringSet: [ 18389 reference sequences ]

#Fix some mistakes in the sample data mistakes:
pseqs_merged@sam_data$pseqs_merged@sam_data$Country%in%c("Denmark", "The Netherlands", "Country")="Wadden"
pseqs_merged@sam_data[rownames(pseqs_merged@sam_data)=="S-22-multi-Mi", "Sample_type"]="Purification bla
```

```

#Add sample_name field to sample_data to help showing the info
samnames=sub('^(.-+-[^-]+).*', '\\1', sample_names(pseqs_merged))
pseqs_merged@sam_data$Sample_name=samnames

#Make a relative abundance to compare to:
pseqs_merged_rel=transform_sample_counts(pseqs_merged, function(x) x / sum(x))

```

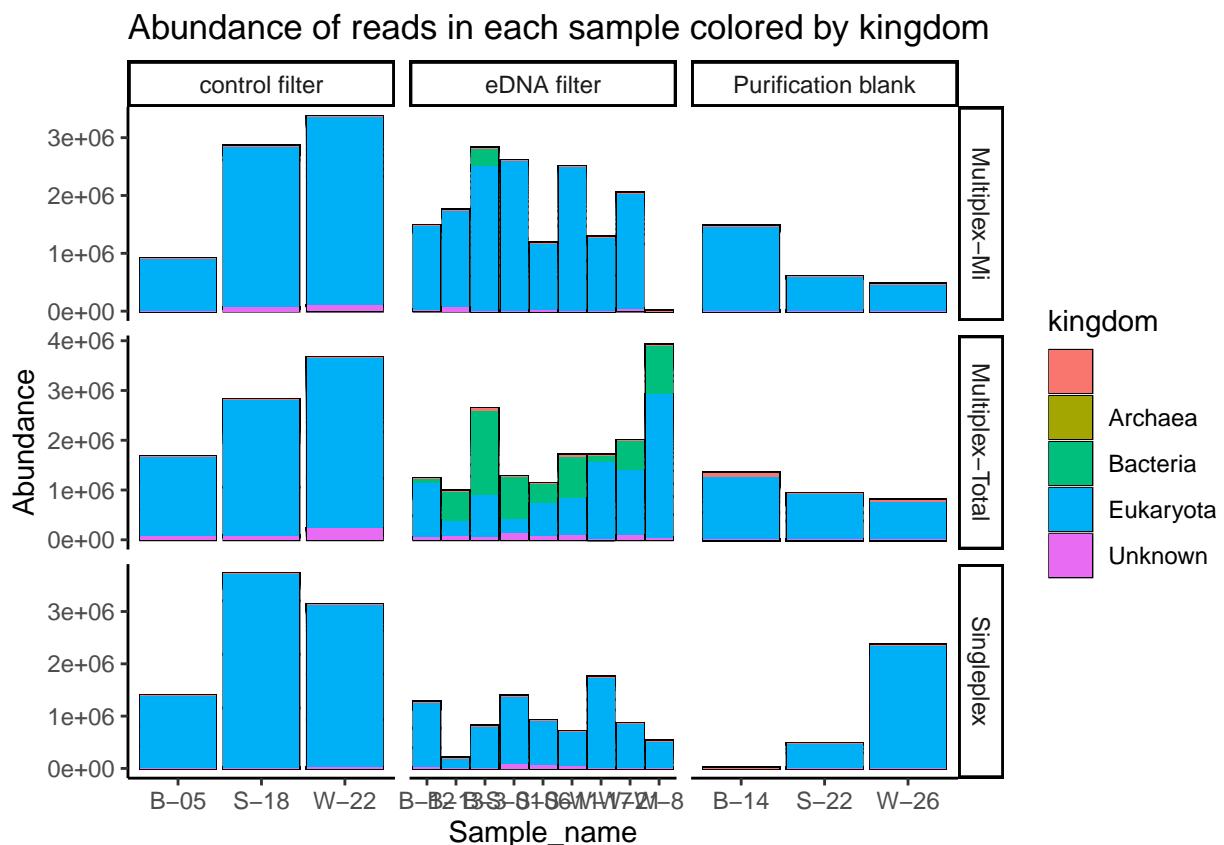
## Reads

First look at the of the amount of reads in different categories:

```

plot_bar(pseqs_merged, x = "Sample_name", fill="kingdom")+
  geom_bar(stat="identity")+
  facet_grid(ReactionType~Sample_type, scales="free")+
  theme_classic()+
  ggtitle("Abundance of reads in each sample colored by kingdom")

```

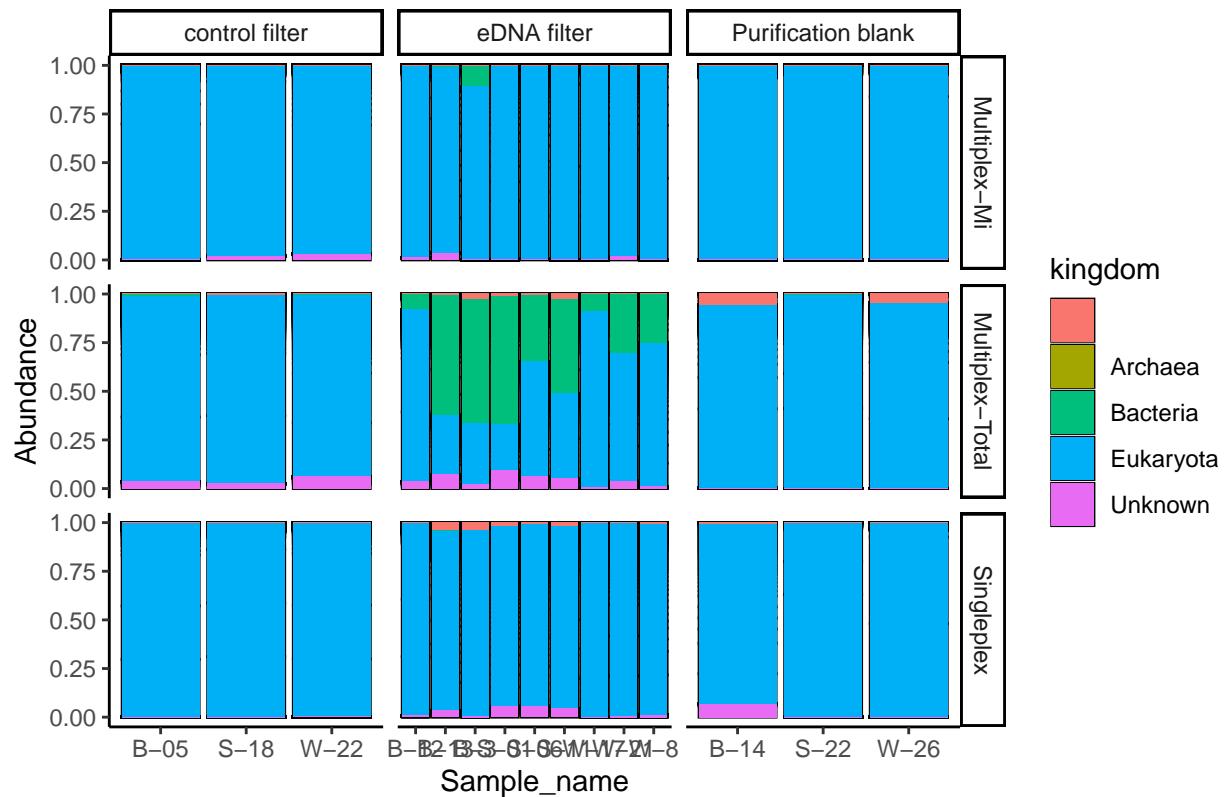


```

plot_bar(pseqs_merged_rel, x = "Sample_name", fill="kingdom")+
  geom_bar(stat="identity")+
  facet_grid(ReactionType~Sample_type, scales="free")+
  theme_classic()+
  ggtitle("Relative abundance of reads in each sample colored by kingdom")

```

### Relative abundance of reads in each sample colored by kingdom



The first thing we can see is that there is bacterial reads in significant amounts only in the eDNA filters with the Multiplex-total reaction type.

This is likely because there are more bacteria in the environmental samples and these are picked up by a combination of factors in the multiplexing (also seen in multiple bands in the PCR).

We are also interested to understand which part of the reads can be attributed to the different primer analyses? Look at amount of total reads that were sequences for each sample:

```
# I want to keep the information on total read counts for each base sample before the primers were separated
#setwd("/Users/saara/Dropbox (IPOfI)/Saara/github/Mock communities/eDNAexpeditions/Analysis_Trials/")
read_stats=read.csv("multiqc_general_stats.txt", sep="\t")
sample_names=str_split_fixed(read_stats$Sample, " ", 2)
read_stats$Sample=sample_names[,1]
read_stats$Orientation=sample_names[,2]
read_stats_f=read_stats[read_stats$Orientation=="| qc | fw",]
read_stats_r=read_stats[read_stats$Orientation=="| qc | rv",]
row.names(read_stats_f)=read_stats_f$Sample
row.names(read_stats_r)=read_stats_r$Sample

#Add total reads to the sam_data table of MifishUE
sample_data(pseqs$MifishUE)$Total_reads_forward=read_stats_f[sample_names(pseqs$MifishUE), "FastQC_mqc"]
sample_data(pseqs$MifishUE)$Total_reads_reverse=read_stats_r[sample_names(pseqs$MifishUE), "FastQC_mqc"]

#And also add total reads of each pseq (sample type) to the sam_data table
for (i in 1:length(pseqs)) {
  sample_data(pseqs[[i]])$sample_reads=sample_sums(pseqs[[i]])}
```

```

}

#And also add total number of ASVs
for (i in 1:length(pseqs)) {
sample_data(pseqs[[i]])$sample_asvs=colSums(pseqs[[i]]@otu_table>0)
}

reads_analysis=pseqs$MifishUE@sam_data[,c("Total_reads_forward", "Total_reads_reverse","sample_reads")]

#Define total reads as reads forward + reads reverse as most reads were not merged.
reads_analysis$Total_reads=reads_analysis$Total_reads_forward+reads_analysis$Total_reads_reverse

#Double the amount of reads analysed in MifishUE as these were the merged ones
reads_analysis$MifishUE=reads_analysis$sample_reads*2

#Add amounts of reads from other pseqs to the table
reads=sample_sums(mimammal_unique)
reads_analysis$Mimammal_single=reads[rownames(reads_analysis)]
#Double the reads also from the mimammal
reads_analysis$Mimammal=reads_analysis$Mimammal_single*2

reads=pseqs$C01@sam_data[rownames(reads_analysis),"sample_reads"]
reads_analysis$C01=reads$sample_reads

reads=pseqs$Teleo@sam_data[rownames(reads_analysis),"sample_reads"]
reads_analysis$Teleo=reads$sample_reads

reads=pseqs$rna16S@sam_data[rownames(reads_analysis),"sample_reads"]
reads_analysis$rna16S=reads$sample_reads

#Calculate the proportion of reads by dividing with the total reads
reads_fraction=reads_analysis[,c(5,6, 8, 9, 10)]/reads_analysis$Total_reads

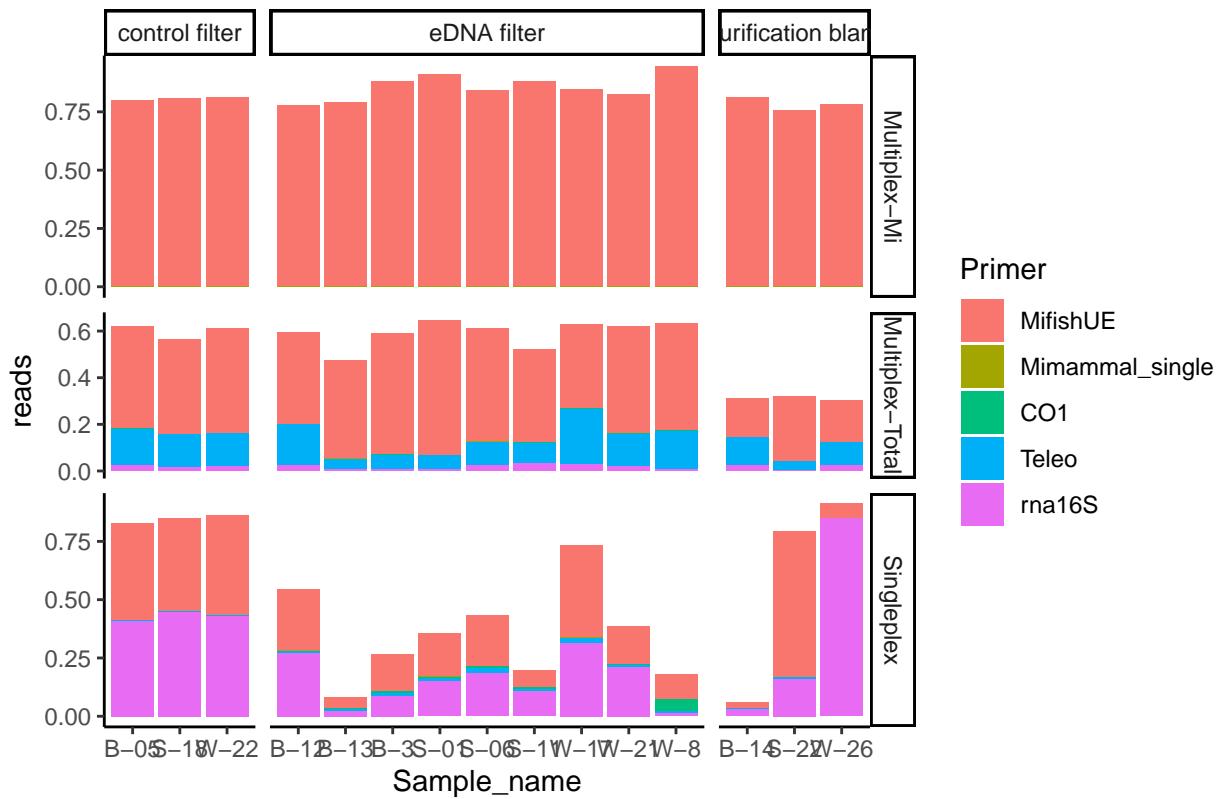
#Add other metadata to the table:
reads_fraction$Sample=rownames(reads_fraction)
reads_fraction <- cbind(reads_fraction, pseqs_merged@sam_data[rownames(reads_fraction), c("ReactionType"]

#Prepare data for plotting
df <- melt(reads_fraction, id.vars = c("Sample_name", "ReactionType", "Sample_type", "Country"), variable_name="Primer")
df$reads=as.numeric(df$reads)

ggplot(df, aes(x = Sample_name , y= reads, fill = Primer)) +
  geom_bar( stat = "identity") + facet_grid(ReactionType~Sample_type, scales="free", space="free")+
  theme_classic()
  ggtitle("Propotion of total reads used in the analysis, colored by the primer sequences")

```

Propportion of total reads used in the analysis, colored by the primer sequence



In many samples a large part of the reads survived until the analysis. Only for the singleplex, there is a clear lack of sequences, and it will be important to check what the remaining reads are, is there a reason that primers are not recognized by cutadapt in these sequences, or is something else going on?

The difference in which primers contributed to the sequences in which reaction type is likewise difficult to understand. The expectation would be that in the singleplex run, there would be an equal amount of reads from each primer type, as these were mixed after the PCR (in equimolar amounts? how were the samples cleaned?).

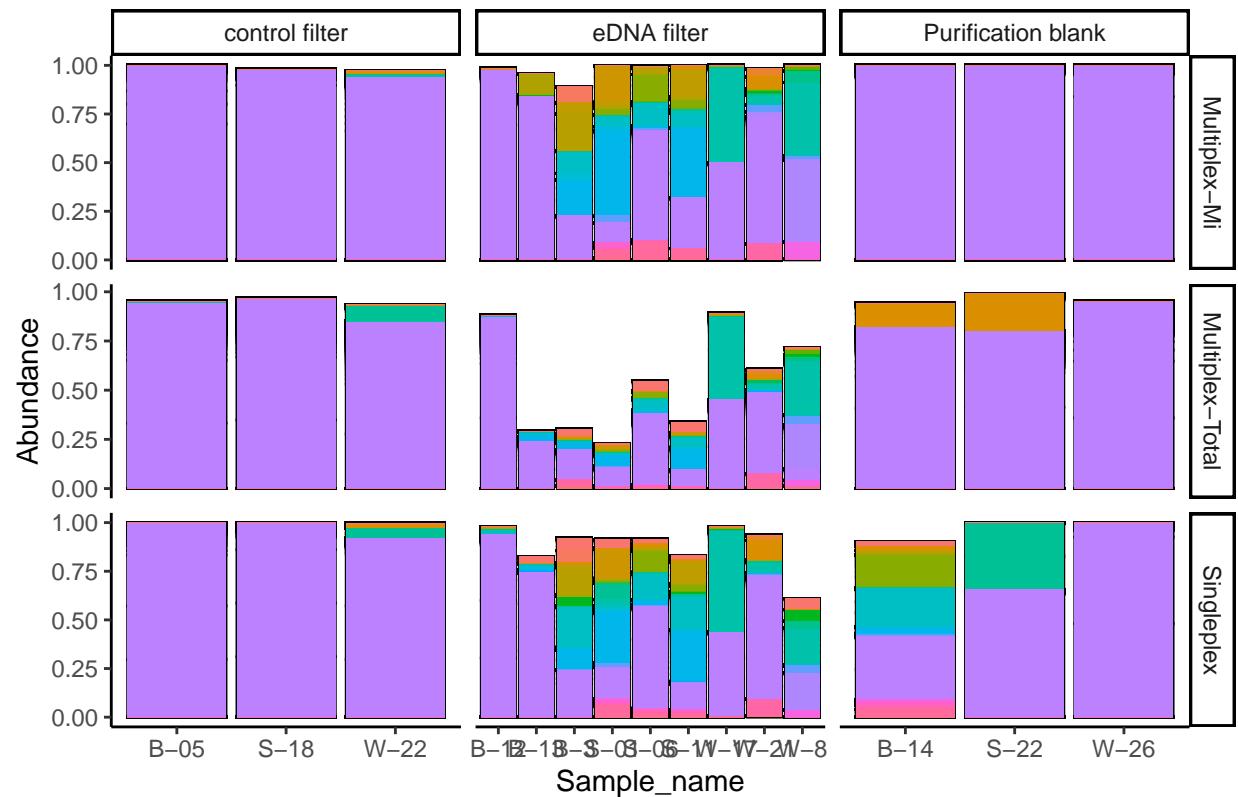
## Taxonomic composition of reads

To look closer into the taxonomic composition of the samples, we will subset.

```
pseqs_merged_rel_euk=subset_taxa(pseqs_merged_rel, kingdom=="Eukaryota")
pseqs_merged_euk=subset_taxa(pseqs_merged, kingdom=="Eukaryota")
pseqs_merged_rel_chor=subset_taxa(pseqs_merged_rel_euk, phylum=="Chordata")
pseqs_merged_chor=subset_taxa(pseqs_merged_euk, phylum=="Chordata")

plot_bar(pseqs_merged_rel_chor, x = "Sample_name", fill="order")+
  geom_bar(stat="identity")+
  facet_grid(ReactionType~Sample_type, scales="free")+
  theme_classic()+
  #theme(axis.text.x=element_blank())+
  theme(legend.position = "none")+
  ggtitle("Chordata reads in samples colored by order")
```

## Chordata reads in samples colored by order



It is clear that most of the reads and relative abundance is human DNA (not shown, but the violet color is Primates). However good news is that for the eDNA samples there is less human reads.

Looking at the other primates there is “*Macaca Fasciatus*” multiple times in the Brazilian samples. This is from the blast analysis, and when I run blast on some of the reads, it is clear that the blast-lca has worked fine; there are multiple hits of this macaca with 87% identity and 100% coverage (from the COI analysis). However, this is clearly not the correct assignation, as this is not a local species (Asian species). Only the first hit of blast is actually a marine sponge; 89% similarity, 100% coverage, which is much more likely. Therefore the annotation or last common ancestor strategy should be changed to a more accurate version.

Since the reference databases are known to be lacking, and we can expect that there are not several instances of the rare species that we are looking for, we will need to change how the blast-lca is working: less targets evaluated? More stringent criteria? (i.e. 97% similarity). Or simply take only the first blast hit? This should be tested with the help of mock communities, or a mock dataset.

## Remove control sequences

Remove ASVs that are found in the control samples from eDNA samples. This will allow to evaluate the ‘real’ sequences that are found. This can be done also at the level of the PacMAN pipeline also, but I have left them in so that we can look at them in detail.

```
#Remove all the ASVs found in the control filters
control_samples=subset_samples(pseqs_merged_rel, Sample_type=="Purification blank" | Sample_type=="Control")
control_asvs = filter_taxa(control_samples, function(x) sum(x) > 0, TRUE)
control_asvs_names = taxa_names(control_asvs)
asv_names=taxa_names(pseqs_merged_rel) [!(taxa_names(pseqs_merged_rel)%in%control_asvs_names)]
```

```

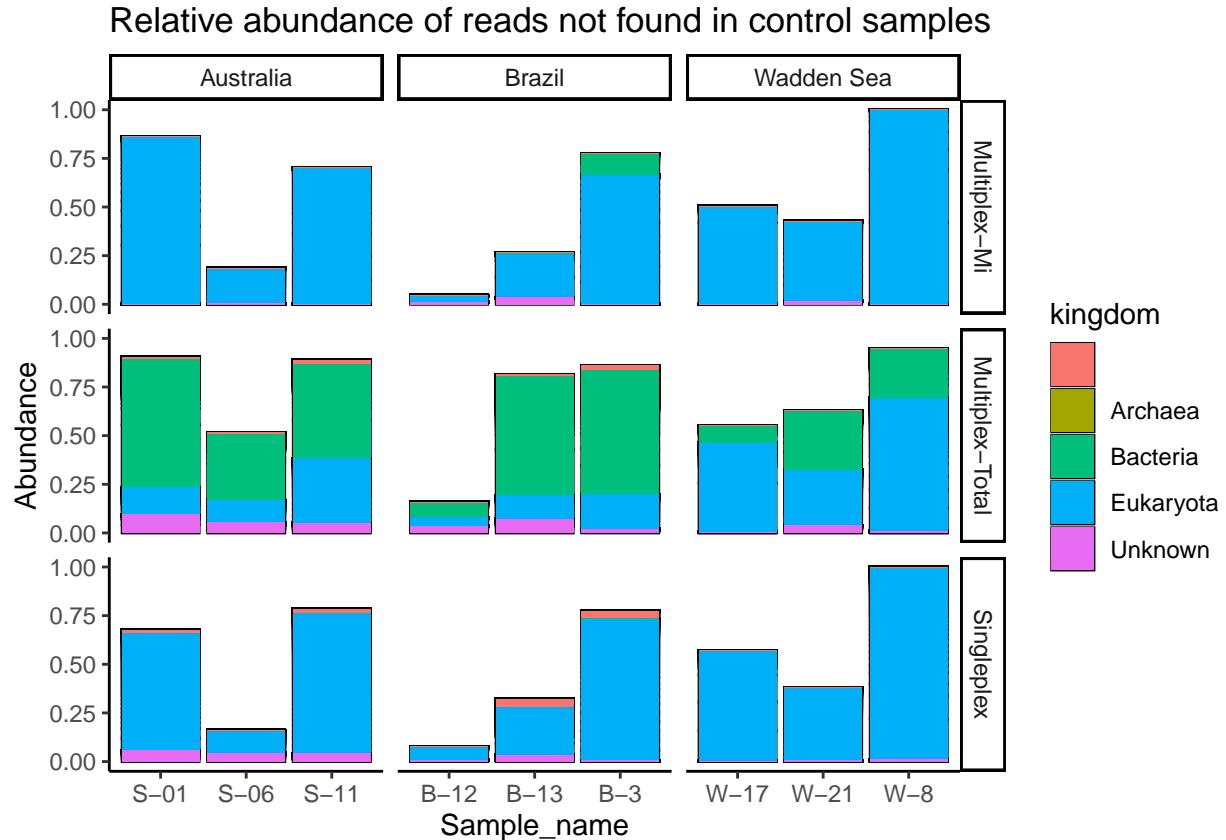
pseqs_merged_rel_eDNA<-subset_samples(pseqs_merged_rel, Sample_type=="eDNA filter")
pseqs_merged_eDNA<-subset_samples(pseqs_merged, Sample_type=="eDNA filter")

pseqs_merged_rel_eDNA_nocontrol = prune_taxa(asv_names, pseqs_merged_rel_eDNA)
pseqs_merged_eDNA_nocontrol = prune_taxa(asv_names, pseqs_merged_eDNA)
summary(sample_sums(pseqs_merged_rel_eDNA_nocontrol))

##      Min. 1st Qu. Median    Mean 3rd Qu.   Max.
## 0.04638 0.35094 0.62822 0.58425 0.83763 1.00000

#26-100% of reads from the eDNA samples are not in the control samples
plot_bar(pseqs_merged_rel_eDNA_nocontrol, x = "Sample_name", fill="kingdom")+
  geom_bar(stat="identity")+
  facet_grid(ReactionType~Country, scales="free_x")+
  theme_classic()+
  ggtitle("Relative abundance of reads not found in control samples")

```

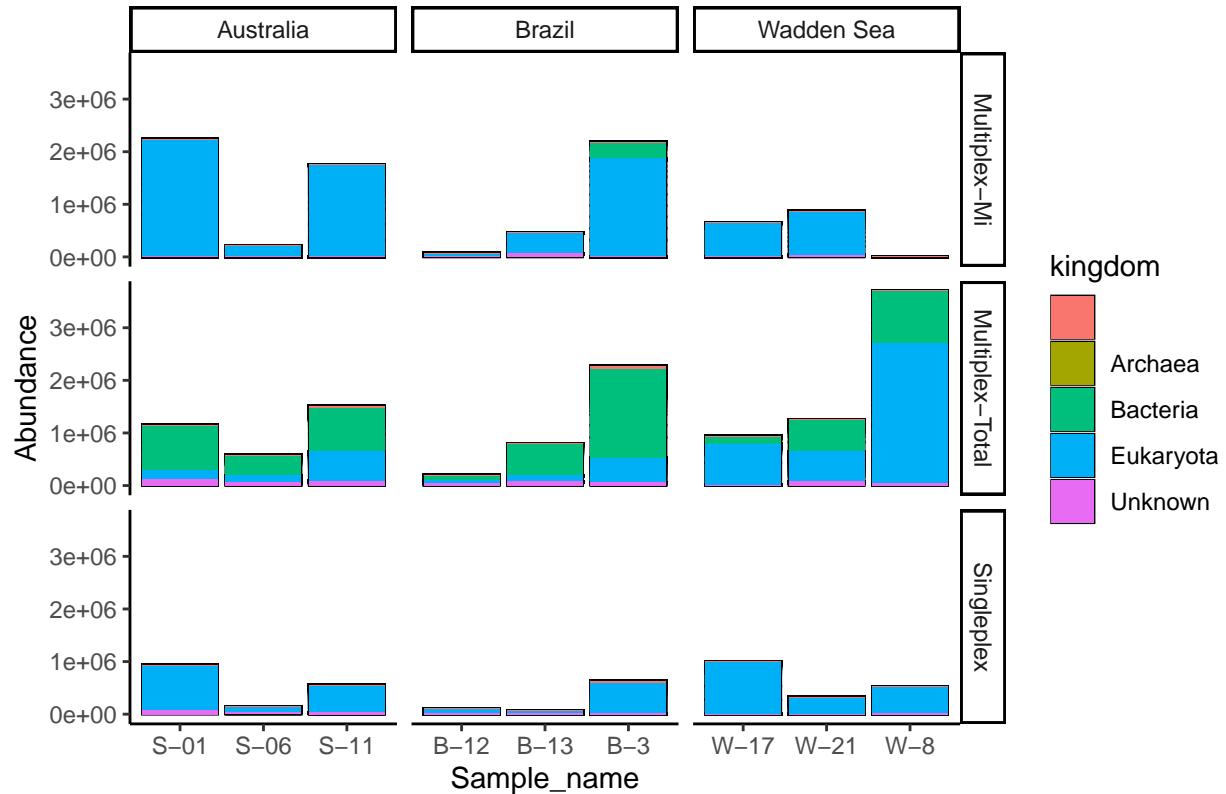


```

plot_bar(pseqs_merged_eDNA_nocontrol, x = "Sample_name", fill="kingdom")+
  geom_bar(stat="identity")+
  facet_grid(ReactionType~Country, scales="free_x")+
  theme_classic()+
  ggtitle("Abundance of reads not found in control samples")

```

## Abundance of reads not found in control samples

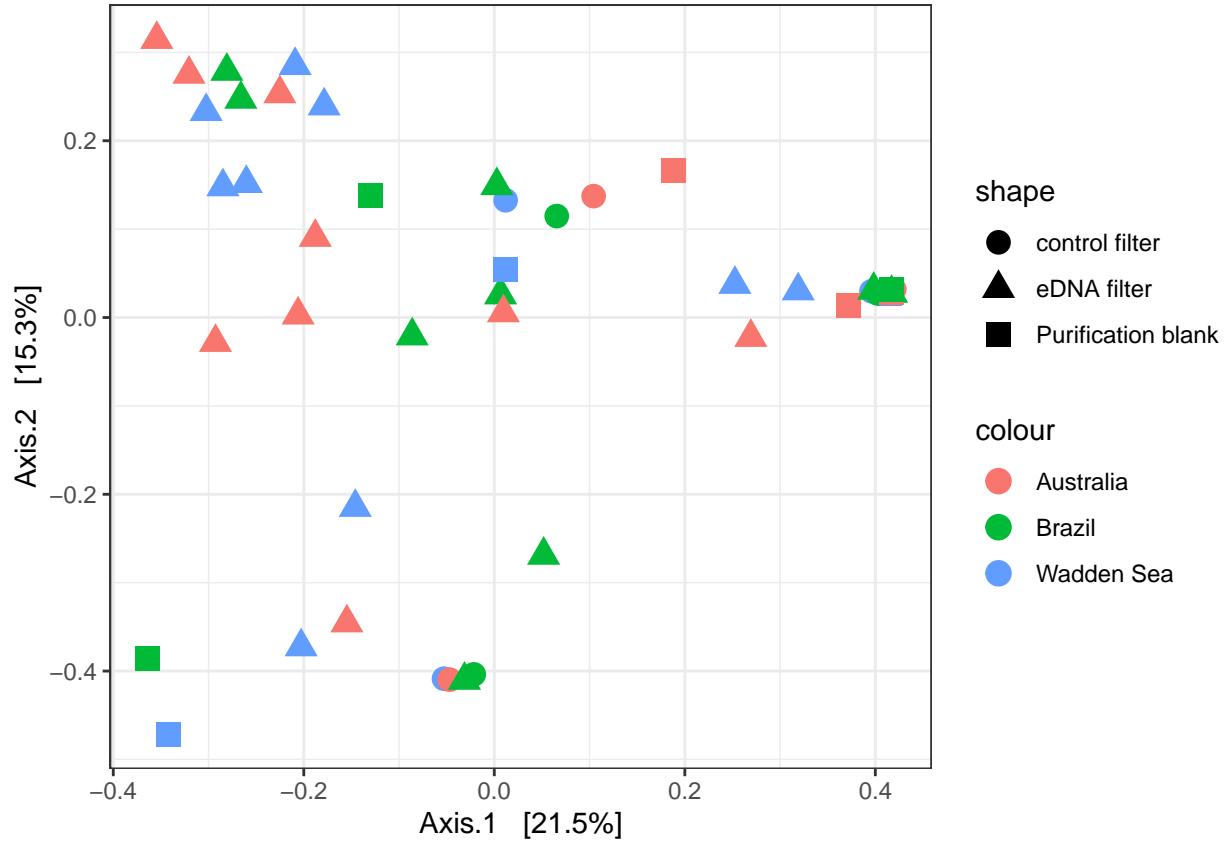


The amount of reads that are not from the controls has a lot to do with the samples, but also the reaction type (multiplex total). There is a large variation between the samples. It is important to check if this can be traced back to DNA-yield, or PCR efficiency possibly, so that critical steps in the lab can be optimized.

Try ordination to see how different/similar the samples from different locations are:

```
pseq.ord <- ordinate(pseqs_merged_rel, "PCoA", "bray")

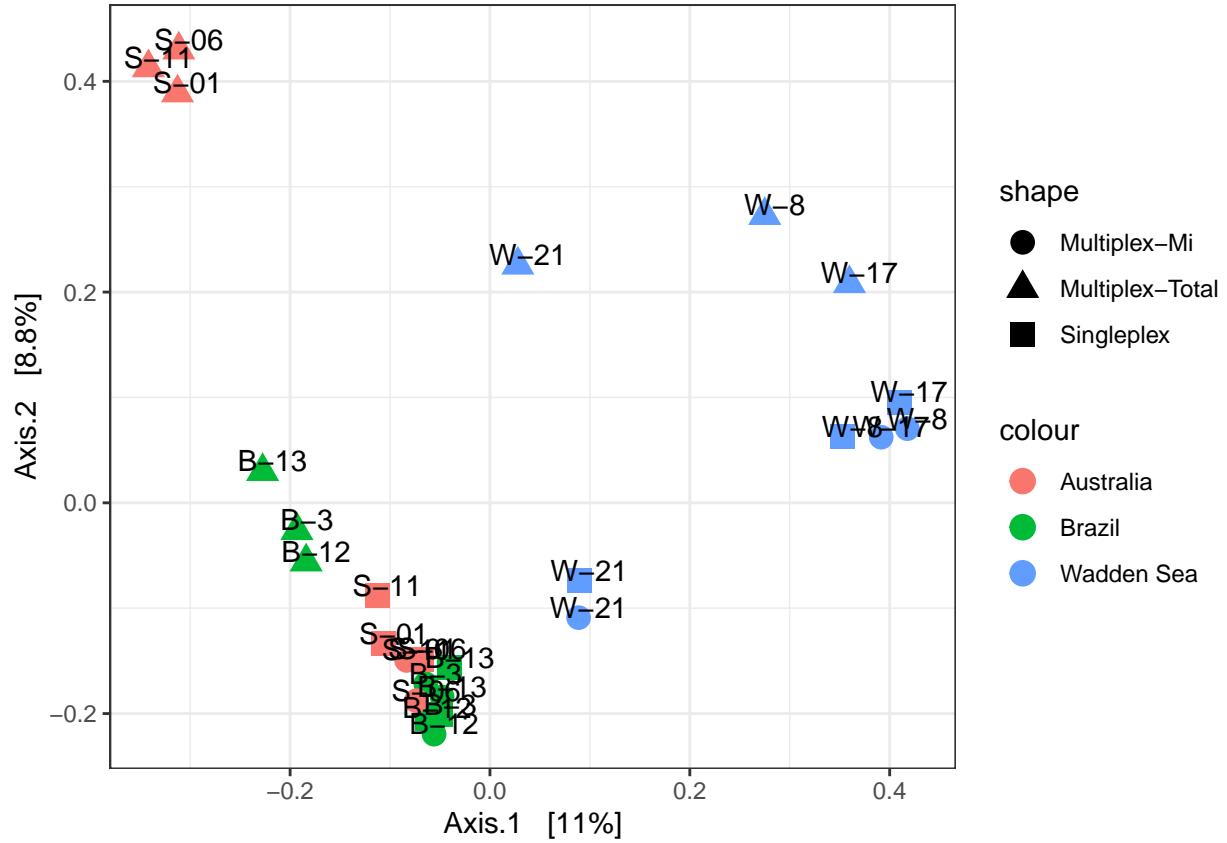
plot_ordination(pseqs_merged_rel, pseq.ord) +
  geom_point(aes(color=Country, shape= Sample_type), size=4) +
  theme_bw()
```



The first ordination does not show separation, because the human reads are so prevalent. Instead the ordination will be done only with eDNA samples.

```
#The data with only the sample info (no control asvs.)
#pseqs_merged_rel_eDNA_nocontrol_euk<-subset_taxa(pseqs_merged_rel_eDNA_nocontrol, kingdom=="Eukaryota")
pseq.ord <- ordinate(pseqs_merged_rel_eDNA_nocontrol, "PCoA", "bray")

plot_ordination(pseqs_merged_rel_eDNA_nocontrol, pseq.ord) +
  geom_point(aes(color=Country, shape= ReactionType), size=4) +
  geom_text(label=pseqs_merged_rel_eDNA_nocontrol@sam_data$Sample_name, nudge_x = 0.01, nudge_y = 0.01)
  theme_bw()
```

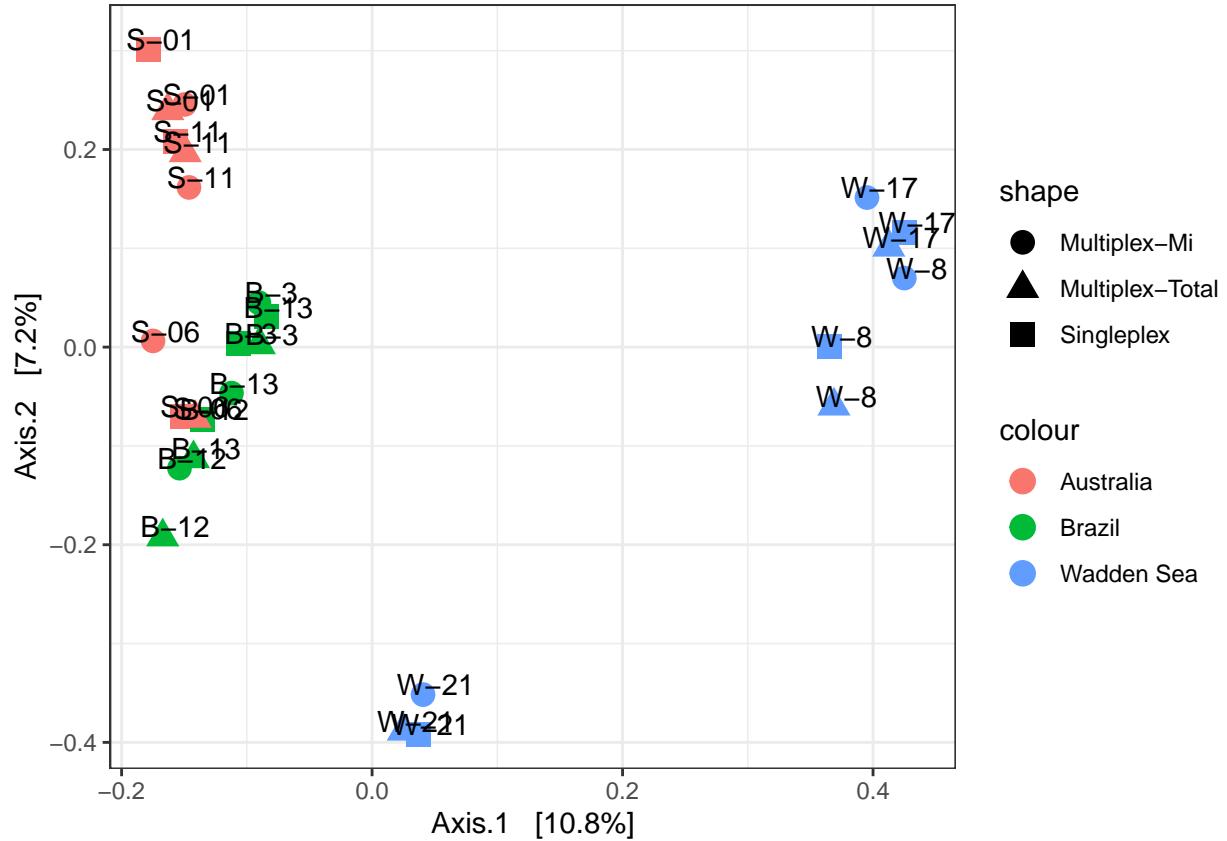


Here we see that the multiplex total reads separate from the remaining reads. This is likely mostly affected by the bacterial reads found in the samples.

Without bacterial reads:

```
#The data with only the sample info (no control asvs.), and only Eukaryotes (so it is not affected by the bacterial reads)
pseqs_merged_rel_eDNA_nocontrol_euk<-subset_taxa(pseqs_merged_rel_eDNA_nocontrol, kingdom=="Eukaryota")
pseq.ord <- ordinate(pseqs_merged_rel_eDNA_nocontrol_euk, "PCoA", "bray")

plot_ordination(pseqs_merged_rel_eDNA_nocontrol_euk, pseq.ord) +
  geom_point(aes(color=Country, shape= ReactionType), size=4) +
  geom_text(label=pseqs_merged_rel_eDNA_nocontrol_euk@sam_data$Sample_name, nudge_x = 0.01, nudge_y = 0)
  theme_bw()
```

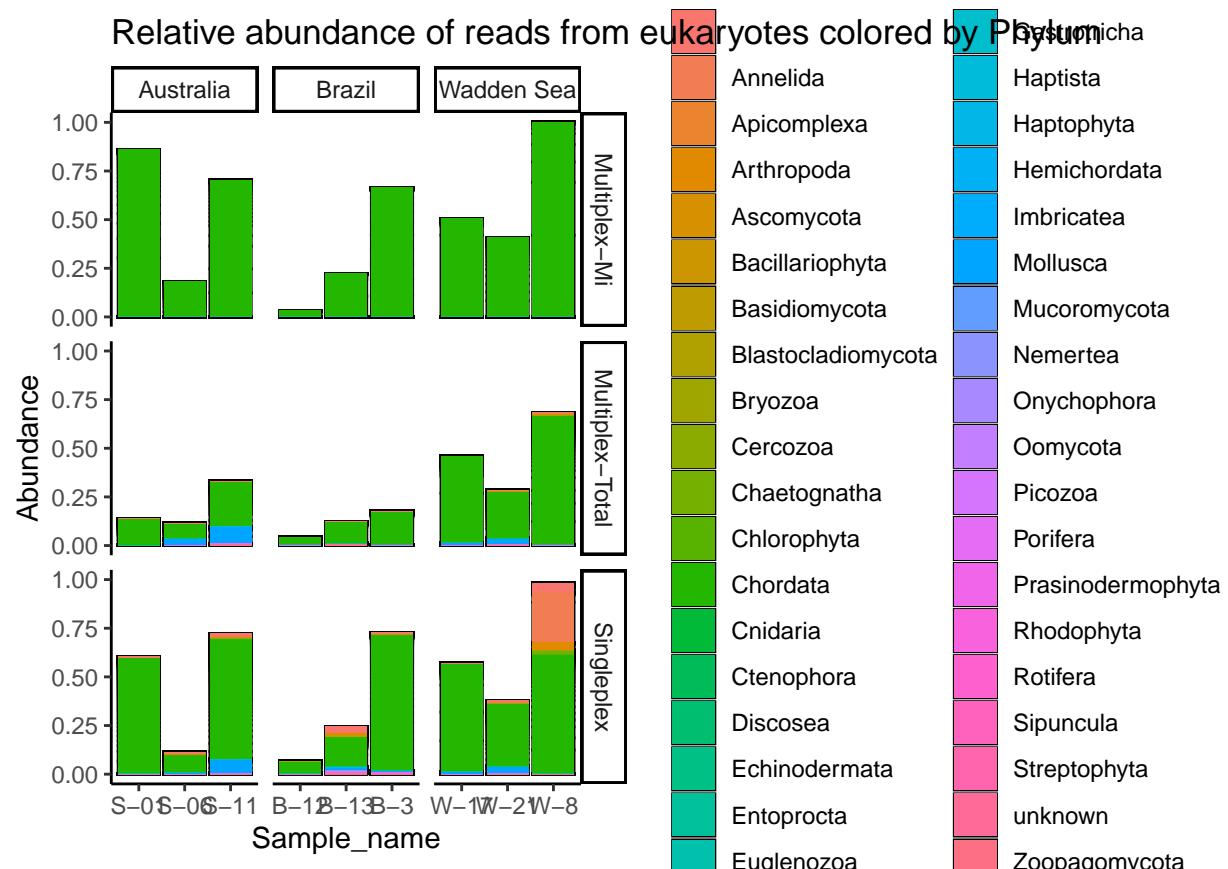


In this way we see more clearly the separation by site. Some of the low-abundance samples are clustering together, as there is not enough data.

A closer look at the taxonomic composition of the eDNA samples:

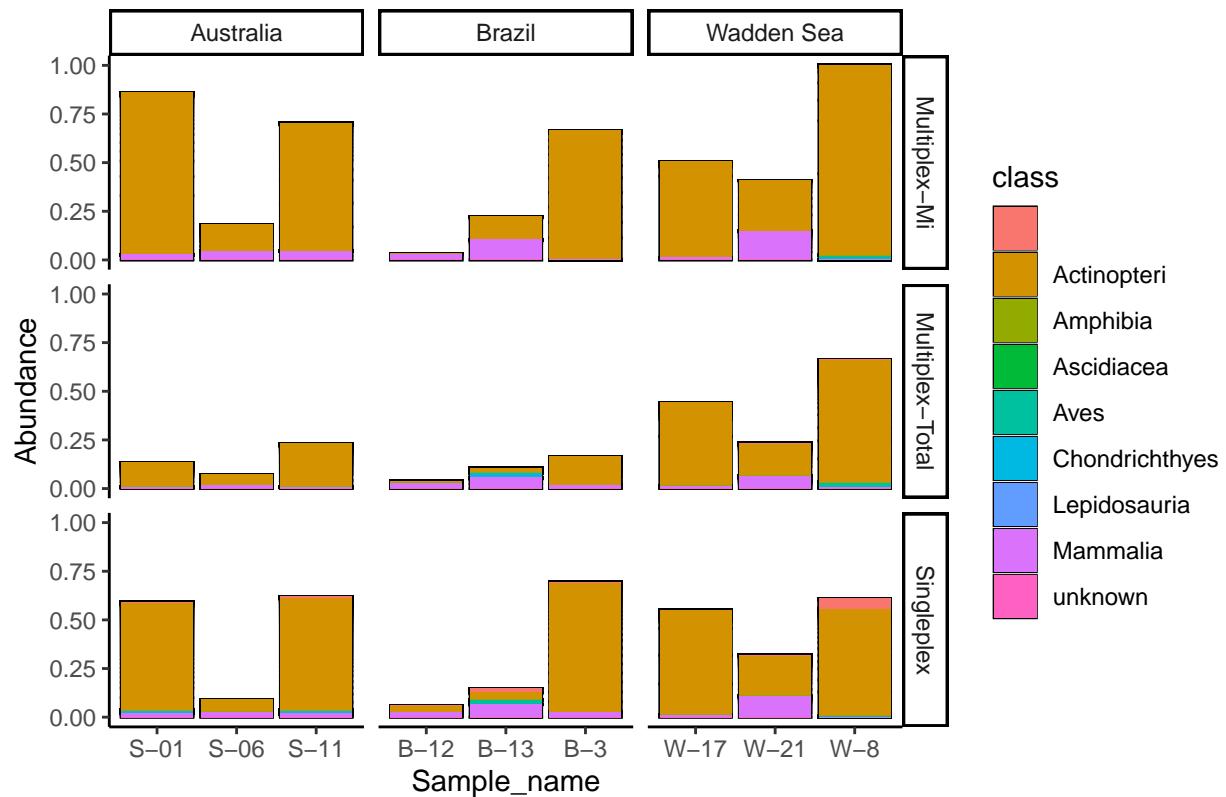
```
pseqs_merged_rel_eDNA_nocontrol_euk<-subset_taxa(pseqs_merged_rel_eDNA_nocontrol, kingdom=="Eukaryota")
pseqs_merged_eDNA_nocontrol_euk<-subset_taxa(pseqs_merged_eDNA_nocontrol, kingdom=="Eukaryota")
pseqs_merged_rel_eDNA_nocontrol_chor<-subset_taxa(pseqs_merged_rel_eDNA_nocontrol_euk, phylum=="Chordata")
pseqs_merged_eDNA_nocontrol_chor<-subset_taxa(pseqs_merged_eDNA_nocontrol_euk, phylum=="Chordata")

plot_bar(pseqs_merged_rel_eDNA_nocontrol_euk, x = "Sample_name", fill="phylum")+
  geom_bar(stat="identity")+
  facet_grid(ReactionType~Country, scales="free_x")+
  theme_classic()+
  ggtitle("Relative abundance of reads from eukaryotes colored by Phylum")
```



```
plot_bar(pseqs_merged_rel_eDNA_nocontrol_chor, x = "Sample_name", fill="class")+
  geom_bar(stat="identity")+
  facet_grid(ReactionType~Country, scales="free_x")+
  theme_classic()+
  ggtitle("Relative abundance of reads from Chordata colored by Class")
```

### Relative abundance of reads from Chordata colored by Class



We can see that the majority of reads are from phylum Chordata and class Actinopteri (i.e. fish), as targeted.

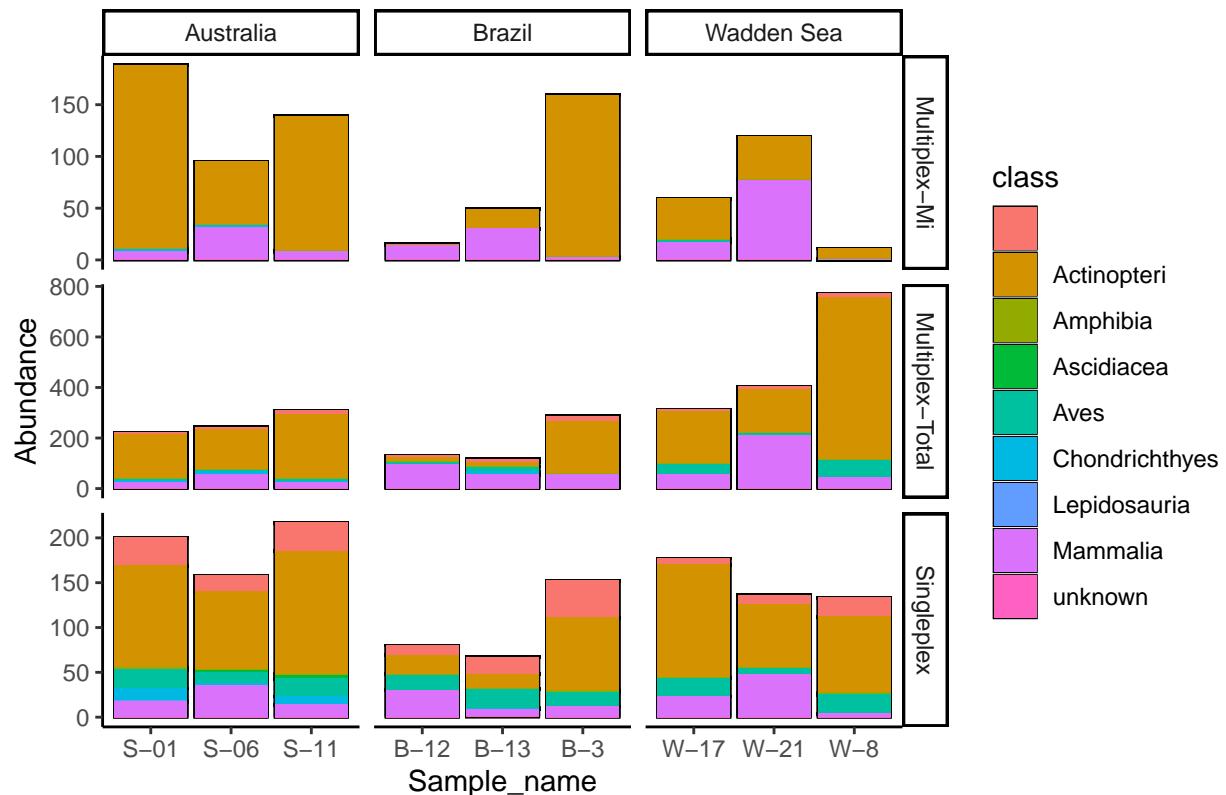
### Number of ASVs

We also want to check what is the number of different ASVs (annotated species) in the different samples.

```
ps_pa<-transform_sample_counts(pseqs_merged_eDNA_nocontrol_chor, function(abund) 1*(abund>0))

plot_bar(ps_pa, x = "Sample_name", fill="class")+
  geom_bar(stat="identity")+
  facet_grid(ReactionType~Country, scales="free")+
  theme_classic()+
  ggtitle("Number of Chordata ASVs found in each sample in different classes")
```

## Number of Chordata ASVs found in each sample in different classes



## Number of different taxonomic ranks assigned at each level:

Here we check the number of different names found in each sample.

```
#Calculate number of ASVs, by making a presence-absence ASV table.
#I will keep only eukaryota as the naming of bacteria is not cohesive (i.e. species name: uncultured pr
ps_pa<-transform_sample_counts(pseqs_merged_eDNA_nocontrol_euk, function(abund) 1*(abund>0))

#Clean up the species names:
#Replace those species names with numerical as NA, and with "uncultured" and with "environmental" "sp."
ps_pa@tax_table[grepl('0-9|uncultured|environmental|sp.',ps_pa@tax_table[, "species"]),"species"] =NA

#replace underscores with space
ps_pa@tax_table[, "species"]<-gsub('_', " ",ps_pa@tax_table[, "species"])

ps_taxa=list()
ps_unique_taxa=list()

#Count the number of unique names at each taxonomic rank:
for (i in 2:7){
  tax_rank=colnames(ps_pa@tax_table)[i]
  ps_taxa[[i-1]]<-tax_glm(ps_pa, taxrank=tax_rank, NArm=T)
  names(ps_taxa)[i-1] <- tax_rank

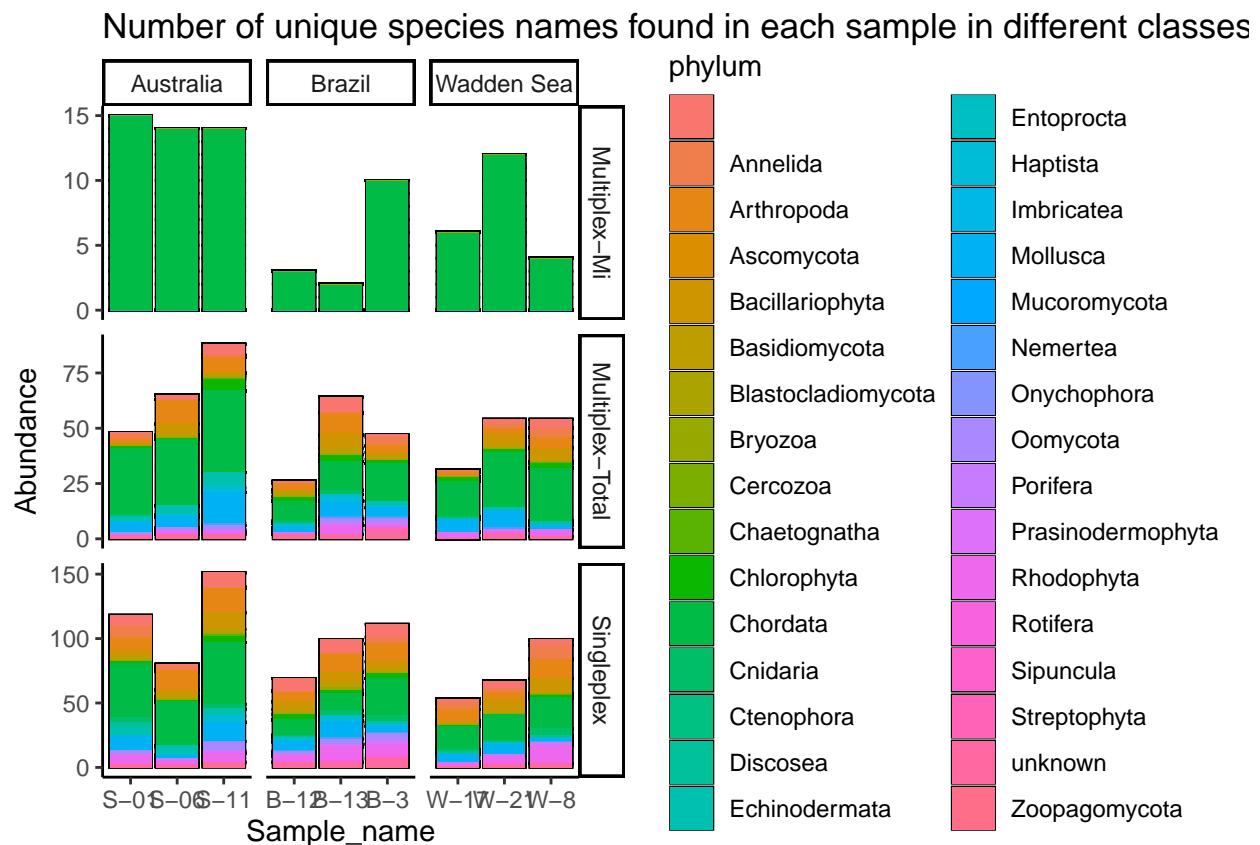
  ps_unique_taxa[[i-1]] <- transform_sample_counts(ps_taxa[[i-1]], function(abund) 1*(abund>0))
```

```

    names(ps_unique_taxa)[i-1] <- tax_rank
}

plot_bar(ps_unique_taxa$species, x = "Sample_name", fill="phylum")+
  geom_bar(stat="identity")+
  facet_grid(ReactionType~Country, scales="free")+
  theme_classic()+
  ggtitle("Number of unique species names found in each sample in different classes")

```



```

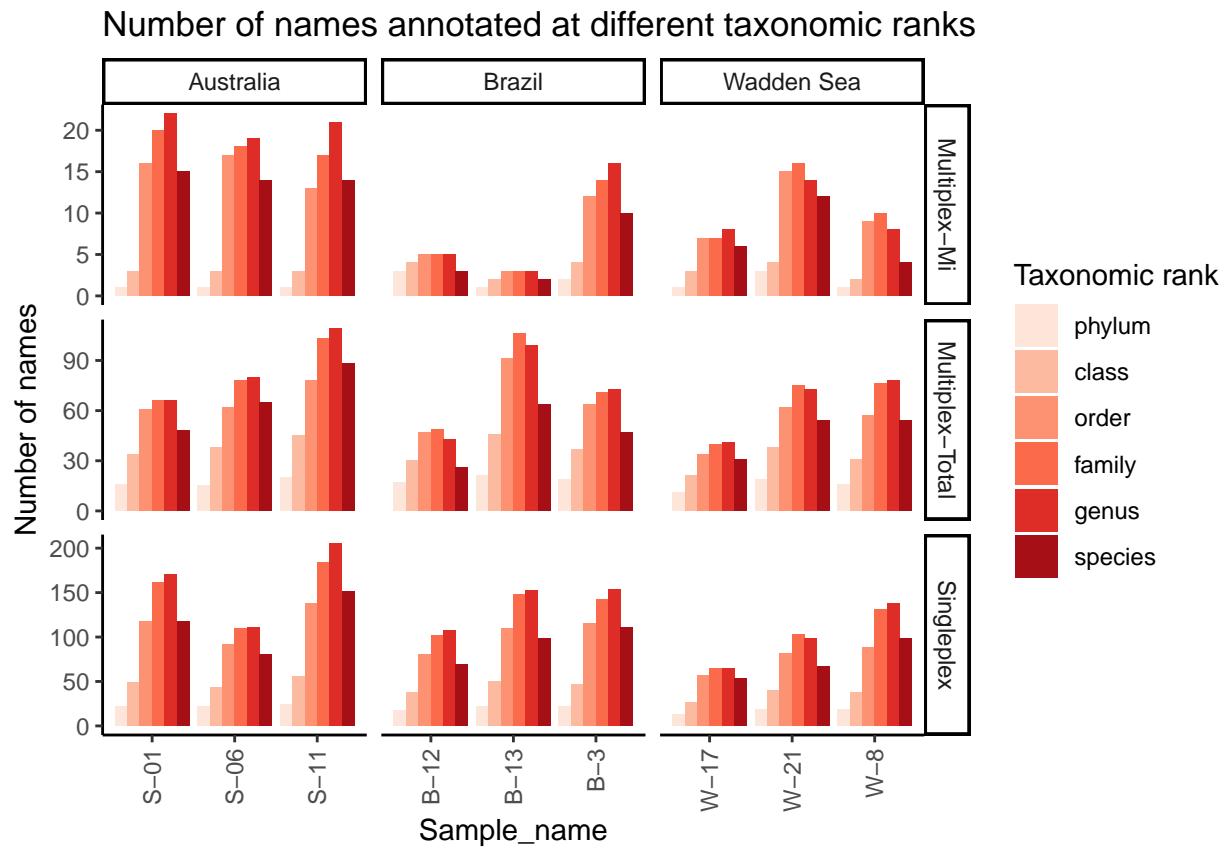
#Combine the amount of names at each taxonomic level in each sample
df_taxa=lapply(ps_unique_taxa, sample_sums)
df_taxa=as.data.frame(df_taxa)

#Add other sample information to the table
df_taxa$Sample=rownames(df_taxa)
melt <- gather(data = df_taxa, key = "Taxonomic rank", "Number of names", phylum:species)
melt <- cbind(melt, pseqs_merged@sam_data[melt$Sample, c("ReactionType", "Sample_type", "Sample_name", "")]
melt$`Taxonomic rank` <- factor(melt$`Taxonomic rank`, levels=c(unique(melt$`Taxonomic rank`)))

ggplot(melt, aes(x=Sample_name, y=`Number of names`, fill=`Taxonomic rank`))+
  geom_bar(stat="identity", position = "dodge")+
  facet_grid(ReactionType~Country, scales="free")+
  scale_fill_brewer(palette = 14)+
  theme_classic()

```

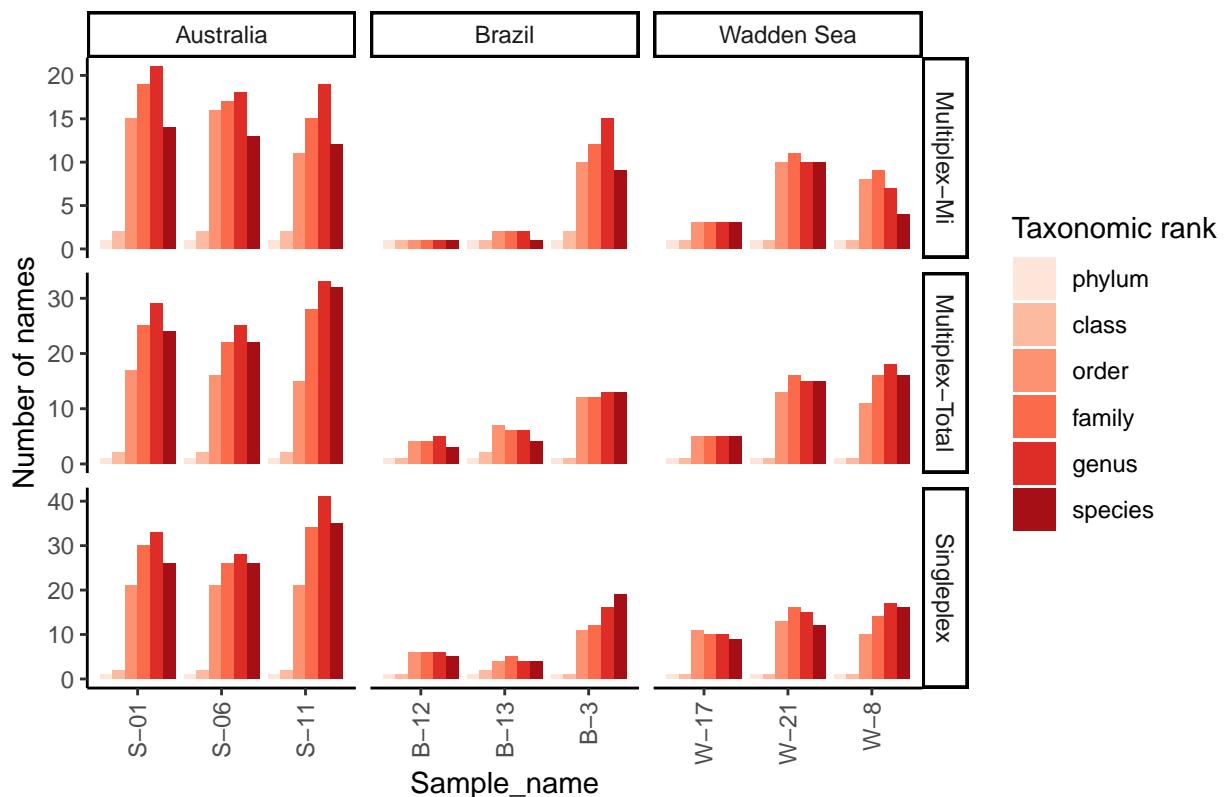
```
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+  
ggtitle("Number of names annotated at different taxonomic ranks")
```



We can see that most annotations are at the genus and family level. The low number of names in the Brazilian samples is clear also in the other steps of the analysis, low amounts of reads compared to other samples. Likely this is therefore due to issues in sampling. However it seems that there are still a good amount of reads in the other categories than mifish, could the issue therefore be more about a lack of fish dna in the samples?

Run the same analysis for only fish:

## Number of names for fish species annotated at different taxonomic ranks



## Conclusions

Conclusions from the different steps of the analysis:

- Bacterial reads in significant amounts only in the eDNA filters with the Multiplex-total reaction type.
  - Multiplexing all primers likely results in the amplification of reads not in the target range
  - Either due to the combination of the different primers, or the unspecificity of the used annealing temperature
- Most reads from the original sequencing files were analysed in the pipeline
  - Still unclear:
    - Why was a smaller portion of the singleplex reads picked up by the pipeline (likely cutadapt?)
    - The reason for why the singleplex reads were not better distributed across the different primers is not clear
    - Why were teleo reads poorly recognized (i.e. f and r reads) in the singleplex reactions? A cutadapt issue?
- Large proportion of reads were human, but on average 58% of reads in the eDNA samples were from environmental sources
- Most analysed environmental reads were fish as targeted
- The amount of species names differed largely between samples

- For Brazil in fact very few fish species (but the amount of other species was as high as the other samples)
- This indicates that there likely was not enough eDNA of fish in the samples to enable analysis.