

Phototrophy analysis

Julia Golebiowska

2024-03-14

```
#or just load prepared initial data
load(file="phototrophy.RData")
```

1. Rhodopsins analysis

1.1. Find rhodopsins genes

```
## [1] "K04643"
```

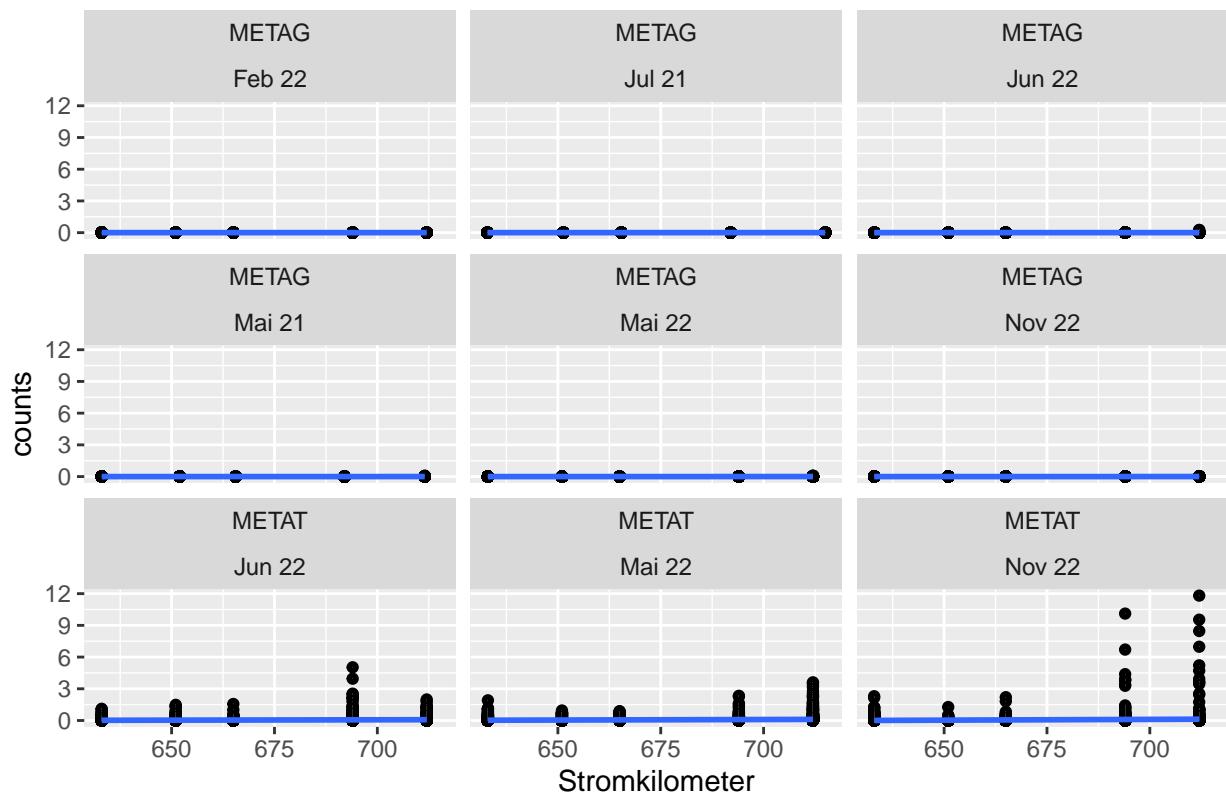
1.2. Explore rhodopsins spatial and temporal distribution

```
#prepare the table with metadata for plot
meta_rhd_clust <- rhd_clust %>%
  pivot_longer(!c("QUERY", "KO", "BRITE", "CAZY", "COG", "DISEASE", "DRUG", "ENZYME", "GO", "MODULE", "P
  merge(metadata)
meta_rhd_clust$Stromkilometer <- sub(", ", ".", meta_rhd_clust$Stromkilometer)
meta_rhd_clust$Stromkilometer <- as.double(meta_rhd_clust$Stromkilometer)

#plot it!
meta_rhd_clust %>%
  filter(Station != 'BunthausSpitze') %>%
  filter(Sample_date != "Nov 21") %>%
  mutate(counts=counts/1000) %>%
  ggplot(aes(x=Stromkilometer, y=counts)) + geom_point() +
  geom_tile() + facet_wrap(~data_type *Sample_date) + ggtitle("Rhd counts across different seasons, riv

## `geom_smooth()` using formula = 'y ~ x'
```

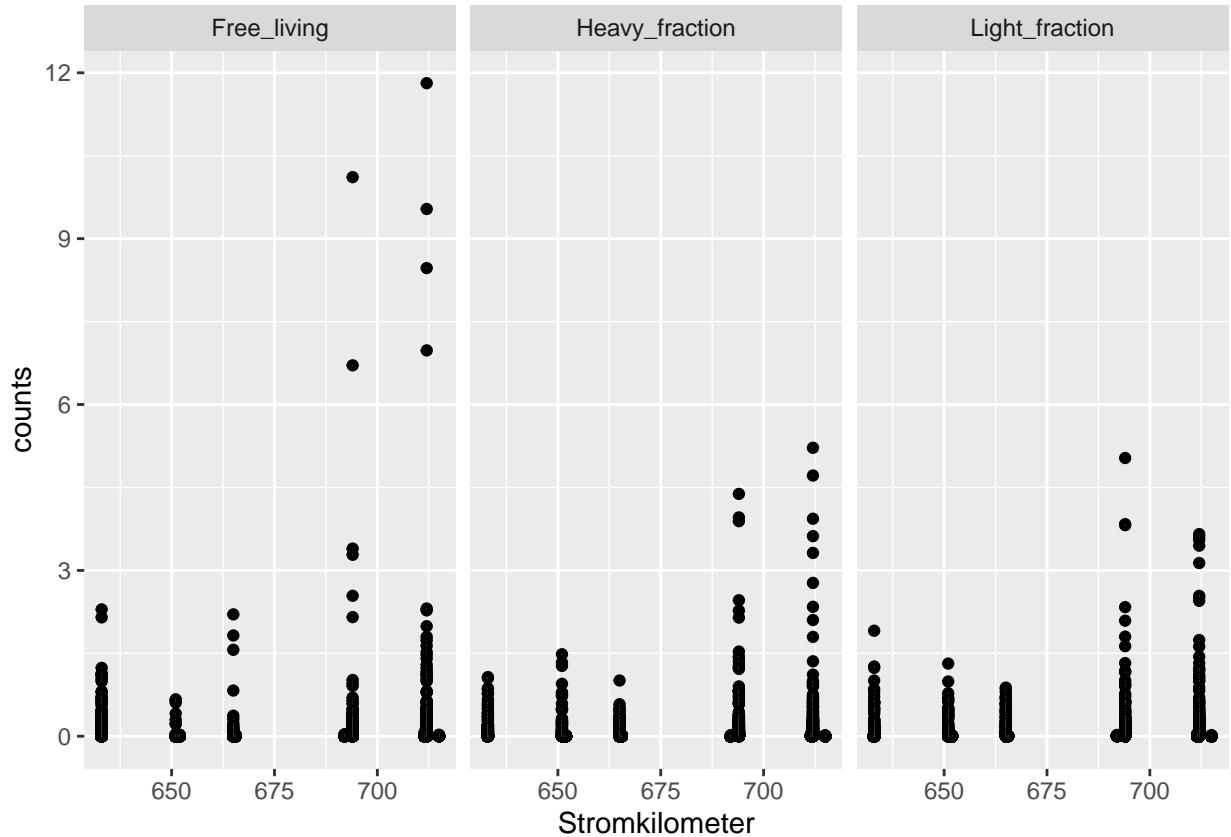
Rhd counts across different seasons, river sites and data types



```
#not enough to do meaningful analysis
#interestingly all rhodopsins were found in viral scaffolds ?
#usually the highest variance in the last station
```

1.3. Check different fractions

```
#check different fractions
#plot it!
meta_rhd_clust %>%
  filter(Station != 'BunthausSpitze') %>%
  filter(Sample_date != "Nov 21") %>%
  mutate(counts=counts/1000) %>%
  ggplot(aes(x=Stromkilometer, y=counts)) + geom_point() +
  geom_tile() + facet_wrap(~ Sample_type)
```



#higher variance in free living

1.4. Check correlations

```
#explore correlations for METAT counts
numeric_col <- c("counts", "Stromkilometer", "Sat_02_TBDHereon", "Turbidity_TBDHereon", "SPM_mgperL", "DO")
meta_rhd_clust_df <- meta_rhd_clust %>% filter(data_type == "METAT") %>%
  filter(Station != 'BunthausSpitze') %>%
  filter(Sample_date != "Nov 21") %>%
  filter(data_type == 'METAT') %>%
  select(numeric_col)

## Warning: Using an external vector in selections was deprecated in tidyselect 1.1.0.
## i Please use `all_of()` or `any_of()` instead.
##   # Was:
##   data %>% select(numeric_col)
##
##   # Now:
##   data %>% select(all_of(numeric_col))
##
## See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```

meta_rhd_clust_df[ meta_rhd_clust_df == ""] <- NA

for (i in c(3:length(meta_rhd_clust_df[1]))){
  meta_rhd_clust_df[,i] <- as.numeric(sub(",",".", meta_rhd_clust_df[,i])) #replace , to . to get a s
}
meta_rhd_clust_df[is.na(meta_rhd_clust_df)] <- 0
head(meta_rhd_clust_df)

##   counts Stromkilometer Sat_02_TBDHereon Turbidity_TBDHereon SPM_mgperL
## 1    0.00          633      55.04           0       68
## 2    0.00          633      55.04           0       68
## 3 133.27          633      55.04           0       68
## 4    0.00          633      55.04           0       68
## 5    0.00          633      55.04           0       68
## 6   27.79          633      55.04           0       68
##   DOC_mg.L DIC_mg.L Temperature_TBDHereon pH_TBDHereon Silicate_mg.L
## 1    7.714     11.08          18.89       7.5    0.327
## 2    7.714     11.08          18.89       7.5    0.327
## 3    7.714     11.08          18.89       7.5    0.327
## 4    7.714     11.08          18.89       7.5    0.327
## 5    7.714     11.08          18.89       7.5    0.327
## 6    7.714     11.08          18.89       7.5    0.327
##   Nitrite_mg.L TotalDissolvedPhosphate_mg.L RespirationRate_02ug.L.h
## 1      0.008            0.036386        6.716419
## 2      0.008            0.036386        6.716419
## 3      0.008            0.036386        6.716419
## 4      0.008            0.036386        6.716419
## 5      0.008            0.036386        6.716419
## 6      0.008            0.036386        6.716419
##   Phosphate_pM O2_TBDHereon TN_mg.L DIC_uM.L DOC_uM.L POC_mgperL PTH_mgperL
## 1      0     159.61     1.784 922.5108 642.2607  2.838645  0.8837012
## 2      0     159.61     1.784 922.5108 642.2607  2.838645  0.8837012
## 3      0     159.61     1.784 922.5108 642.2607  2.838645  0.8837012
## 4      0     159.61     1.784 922.5108 642.2607  2.838645  0.8837012
## 5      0     159.61     1.784 922.5108 642.2607  2.838645  0.8837012
## 6      0     159.61     1.784 922.5108 642.2607  2.838645  0.8837012
##   Ammonium_mg.L SRP_pM Salinity_TBDHereon PTC_mgperL PTN_mgperL
## 1      0.04      0         0.44  3.266749  2.705108
## 2      0.04      0         0.44  3.266749  2.705108
## 3      0.04      0         0.44  3.266749  2.705108
## 4      0.04      0         0.44  3.266749  2.705108
## 5      0.04      0         0.44  3.266749  2.705108
## 6      0.04      0         0.44  3.266749  2.705108

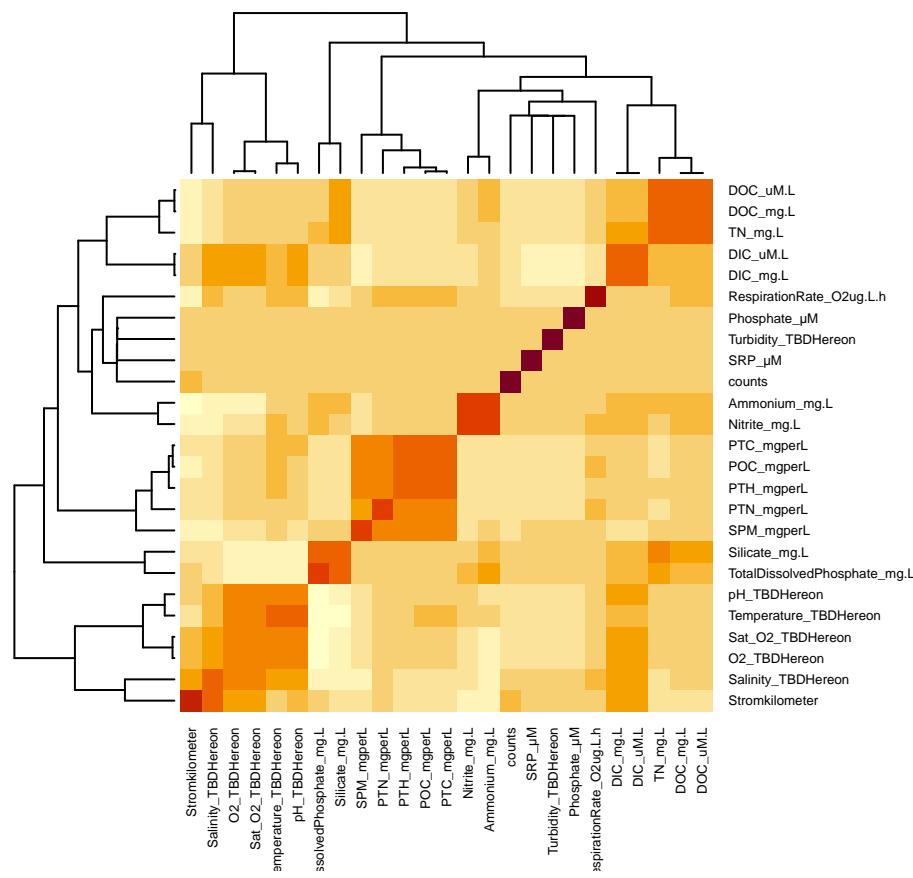
#to numeric
meta_rhd_clust_matrix <- matrix(apply(meta_rhd_clust_df,2, as.numeric), ncol=ncol(meta_rhd_clust_df), d

meta_rhd_clust_matrix[is.na(meta_rhd_clust_matrix)] <- 0
res <- cor(meta_rhd_clust_matrix)

## Warning in cor(meta_rhd_clust_matrix): the standard deviation is zero

```

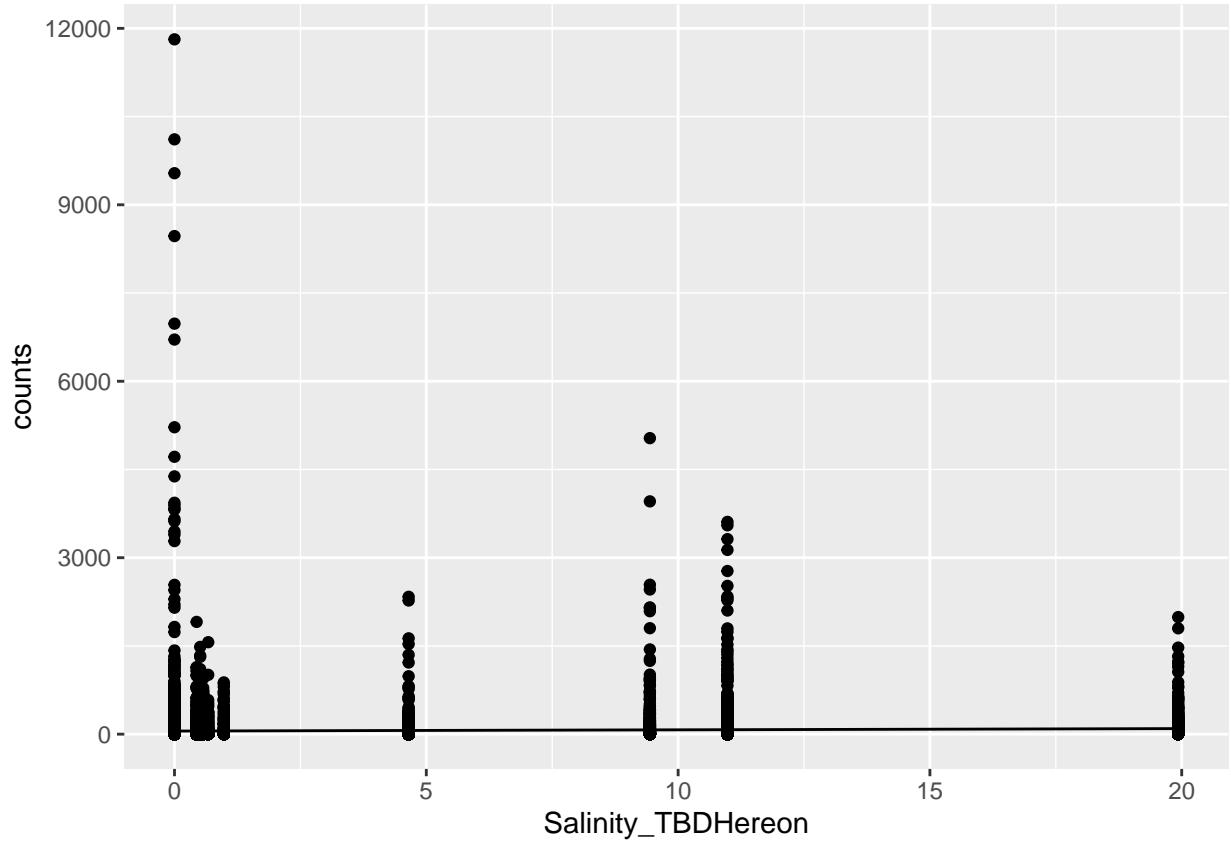
```
res[ is.na(res) ] <- 0
heatmap(round(res, 2), cexRow = 0.5, cexCol = 0.5)
```



Very strong correlation between “DOC_mgL”, “TN_mgL”. Quite strong between “TN_mgL”, “DIC_mgL” ## 1.5. Try to make a model

```
#lets plot two parameters agains rhodopsins counts
#salinity
salinitymodel <- lm(counts ~ Salinity_TBDHeron, data = meta_rhd_clust_df) # didn't work

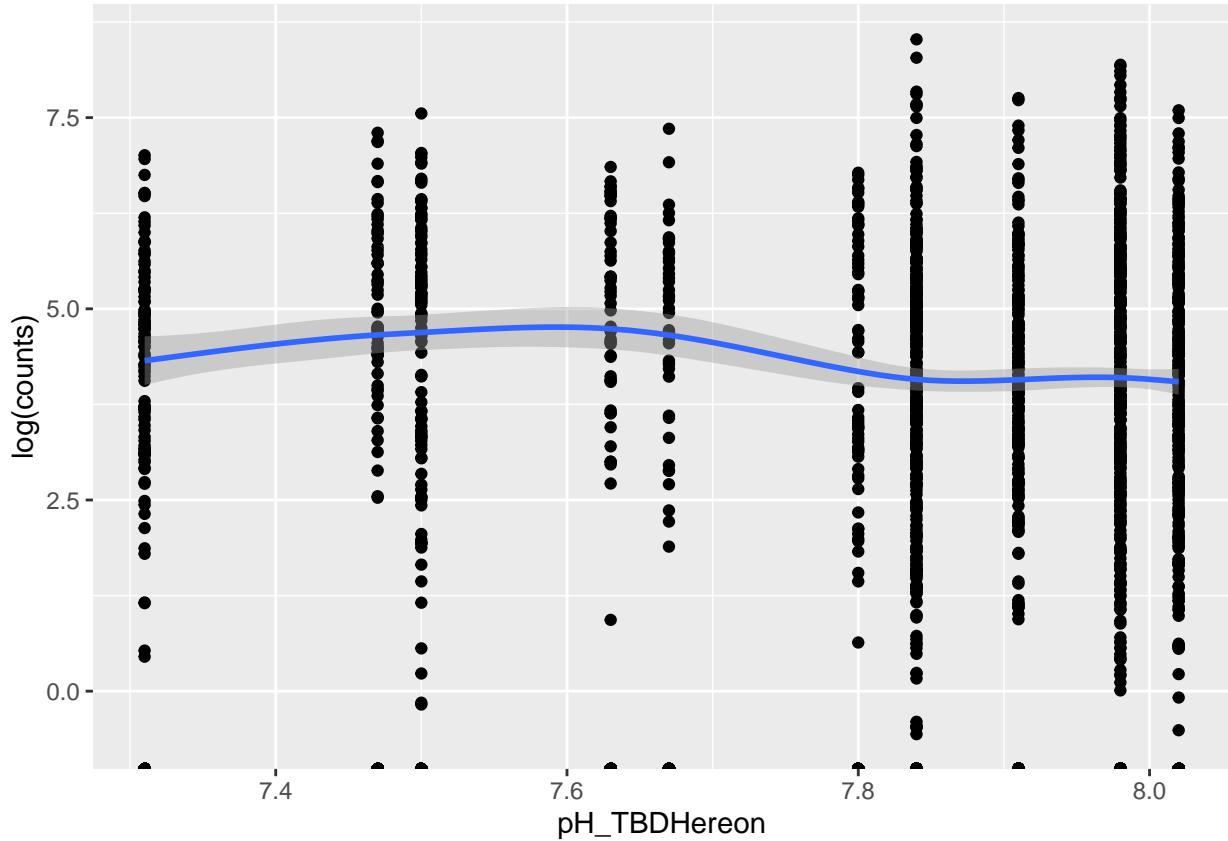
meta_rhd_clust_df %>%
ggplot(aes(x=Salinity_TBDHeron, y=counts)) + geom_point() + geom_line(aes(x= Salinity_TBDHeron , y =
```



```
#pH
ggplot(meta_rhd_clust_df[meta_rhd_clust_df["pH_TBDHereon"]>0,], aes(x=pH_TBDHereon, y=log(counts))) + g

## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'

## Warning: Removed 4310 rows containing non-finite outside the scale range
## ('stat_smooth()').
```



Very subtle tendency ## 1.6. Explore variance

```
#explore the variance as it looks like is different across stations
var_meta_rhd_clust_station <- meta_rhd_clust %>%
  filter(Station != 'BunthausSpitze') %>%
  filter(Sample_date != "Nov 21") %>%
  group_by(Station) %>%
  summarise(var = var(counts))

#variance across stations (similarly the variance is lower n the brackish water)
rhd_var_station <- var_meta_rhd_clust_station %>% arrange(factor(Station, levels =c("Muehlenberger Loch", "Schwarztonnensand", "Meedem Grund", "Twielensfleth", "Brunsbuttel")))

rhd_var_station #change the order
```

```
## # A tibble: 5 x 2
##   Station      var
##   <chr>     <dbl>
## 1 Twielensfleth 3344.
## 2 Schwarztonnensand 4041.
## 3 Meedem Grund  91201.
## 4 Brunsbüttel    52040.
## 5 Mühlenberger Loch 7308.
```

2. PsbA analysis

2.1 Find PsbA gene and check its spatial and temporal distribution

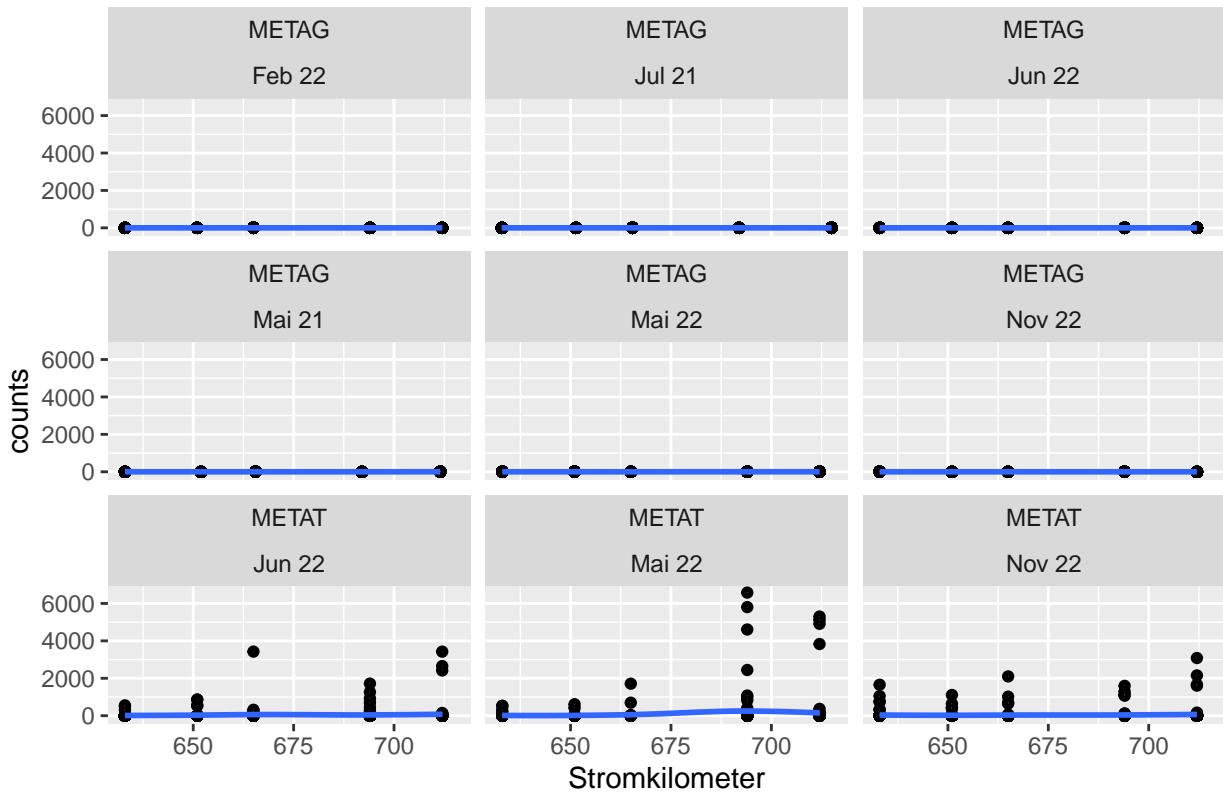
```
#find psbA genes
psbA_clust <- annot %>%
  filter(KO %in% c("K02703")) %>%
  left_join(geneabund, join_by(QUERY == gene_cluster))

#prepare the table with metadata for plot
meta_psbA_clust <- psbA_clust %>%
  pivot_longer(!c("QUERY", "KO", "BRITE", "CAZY", "COG", "DISEASE", "DRUG", "ENZYME", "GO", "MODULE", "J"))
  merge(metadata)
meta_psbA_clust$Stromkilometer <- sub(", ", ".", meta_psbA_clust$Stromkilometer)
meta_psbA_clust$Stromkilometer <- as.double(meta_psbA_clust$Stromkilometer)

#plot it!
meta_psbA_clust %>%
  filter(Station != 'BunthausSpitze') %>%
  filter(Sample_date != "Nov 21") %>%
  mutate(counts=counts/1000) %>%
  ggplot(aes(x=Stromkilometer, y=counts)) + ggtitle("psbA counts across different seasons and river sites")
  geom_tile() + facet_wrap(~data_type *Sample_date) + geom_smooth()

## 'geom_smooth()' using method = 'loess' and formula = 'y ~ x'
```

psbA counts across different seasons and river sites

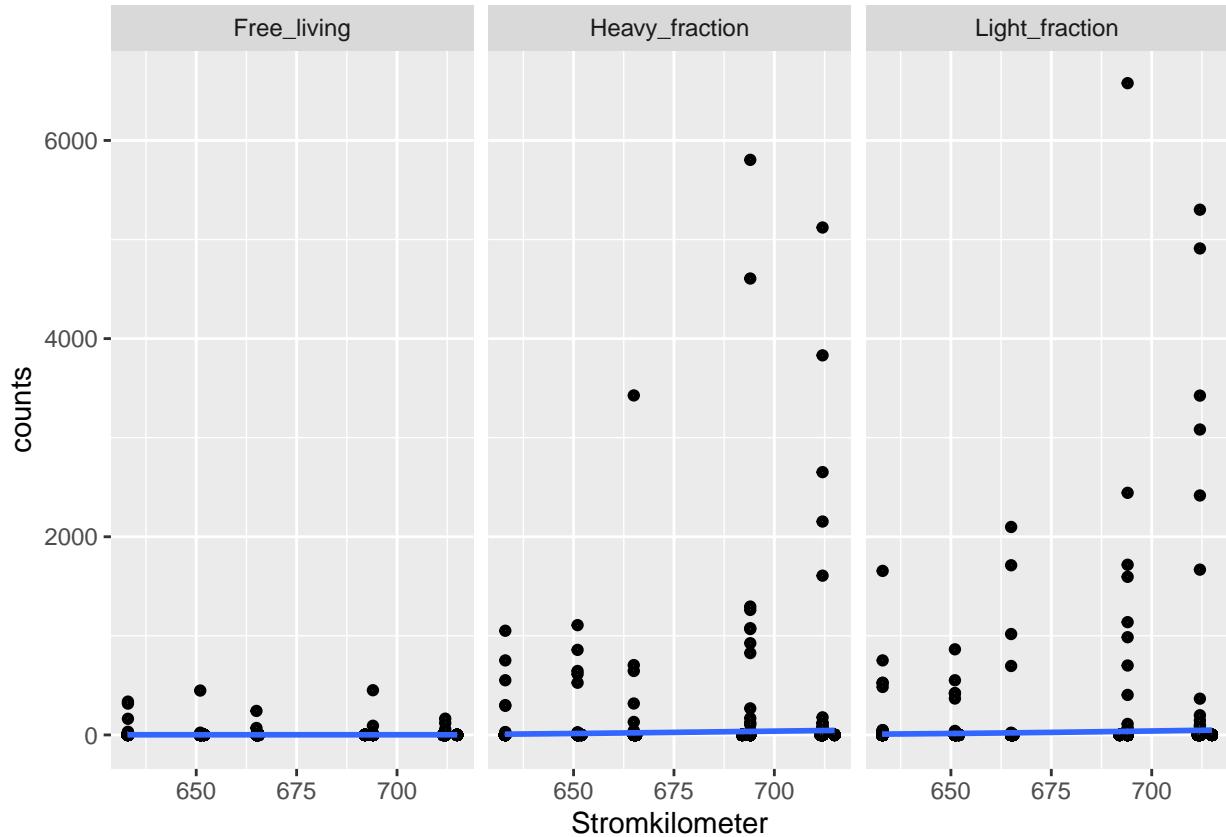


```
#usually the highest variance in the beginning and at the end of the estuary
#but very similar across the samples
```

2.2 Check different fractions

```
#check different fractions
#plot it!
meta_psbA_clust %>%
  filter(Station != 'BunthausSpitze') %>%
  filter(Sample_date != "Nov 21") %>%
  mutate(counts=counts/1000) %>%
  ggplot(aes(x=Stromkilometer, y=counts)) + geom_point() +
  geom_tile() + facet_wrap(~ Sample_type) + geom_smooth()

## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



```
#higher variance in particle associated - interesting opposite to the rhodopsins
```

```
#explore correlations
meta_psbA_clust_df <- meta_psbA_clust %>%
  filter(data_type == "METAT") %>%
  filter(Station != 'BunthausSpitze') %>%
  filter(Sample_date != "Nov 21") %>%
  select(numeric_col)

meta_psbA_clust_df[ meta_psbA_clust_df == ""] <- NA

for (i in c(3:length(meta_psbA_clust_df[1]))){
  meta_psbA_clust_df[,i] <- sub(", ", ".", meta_psbA_clust_df[,i]) #replace , to . to get a string which
```

}

```
head(meta_psbA_clust_df)
```

```
##   counts Stromkilometer Sat_02_TBDHereon Turbidity_TBDHereon SPM_mgperL
## 1    0.00          633      55.04            <NA>       68
## 2    0.00          633      55.04            <NA>       68
## 3    0.00          633      55.04            <NA>       68
## 4    0.00          633      55.04            <NA>       68
## 5   4.82          633      55.04            <NA>       68
## 6    0.00          633      55.04            <NA>       68
##   DOC_mg.L DIC_mg.L Temperature_TBDHereon pH_TBDHereon Silicate_mg.L
```

```

## 1 7.714 11.08 18.89 7.5 0.327
## 2 7.714 11.08 18.89 7.5 0.327
## 3 7.714 11.08 18.89 7.5 0.327
## 4 7.714 11.08 18.89 7.5 0.327
## 5 7.714 11.08 18.89 7.5 0.327
## 6 7.714 11.08 18.89 7.5 0.327
## Nitrite_mg.L TotalDissolvedPhosphate_mg.L RespirationRate_02ug.L.h
## 1 0.008 0.036386 6.716418768
## 2 0.008 0.036386 6.716418768
## 3 0.008 0.036386 6.716418768
## 4 0.008 0.036386 6.716418768
## 5 0.008 0.036386 6.716418768
## 6 0.008 0.036386 6.716418768
## Phosphate_muM 02_TBDHereon TN_mg.L DIC_uM.L DOC_uM.L POC_mgperL
## 1 <NA> 159.61 1.784 922.5107612 642.2606509 2.838644568
## 2 <NA> 159.61 1.784 922.5107612 642.2606509 2.838644568
## 3 <NA> 159.61 1.784 922.5107612 642.2606509 2.838644568
## 4 <NA> 159.61 1.784 922.5107612 642.2606509 2.838644568
## 5 <NA> 159.61 1.784 922.5107612 642.2606509 2.838644568
## 6 <NA> 159.61 1.784 922.5107612 642.2606509 2.838644568
## PTH_mgperL Ammonium_mg.L SRP_muM Salinity_TBDHereon PTC_mgperL PTN_mgperL
## 1 0.88370118 0.04 <NA> 0.44 3.266748695 2.705107501
## 2 0.88370118 0.04 <NA> 0.44 3.266748695 2.705107501
## 3 0.88370118 0.04 <NA> 0.44 3.266748695 2.705107501
## 4 0.88370118 0.04 <NA> 0.44 3.266748695 2.705107501
## 5 0.88370118 0.04 <NA> 0.44 3.266748695 2.705107501
## 6 0.88370118 0.04 <NA> 0.44 3.266748695 2.705107501

```

```

#to numeric
meta_psbA_clust_matrix <- matrix(apply(meta_psbA_clust_df, 2, as.numeric), ncol=ncol(meta_psbA_clust_df))
meta_psbA_clust_matrix[ is.na(meta_psbA_clust_matrix)] <- 0
res <- cor(meta_psbA_clust_matrix)

```

```

## Warning in cor(meta_psbA_clust_matrix): the standard deviation is zero

```

```

res[ is.na(res)] <- 0
round(res, 2)

```

	counts	Stromkilometer	Sat_02_TBDHereon
## counts	1.00	0.08	0.07
## Stromkilometer	0.08	1.00	0.33
## Sat_02_TBDHereon	0.07	0.33	1.00
## Turbidity_TBDHereon	0.00	0.00	0.00
## SPM_mgperL	0.00	-0.25	-0.08
## DOC_mg.L	0.00	-0.24	0.16
## DIC_mg.L	0.04	0.24	0.59
## Temperature_TBDHereon	0.02	-0.02	0.86
## pH_TBDHereon	0.04	0.09	0.93
## Silicate_mg.L	-0.01	-0.18	-0.36
## Nitrite_mg.L	-0.03	-0.38	-0.20
## TotalDissolvedPhosphate_mg.L	-0.03	-0.07	-0.55
## RespirationRate_02ug.L.h	0.00	-0.10	0.18
## Phosphate_muM	0.00	0.00	0.00

## O2_TBDHereon	0.07	0.34	1.00
## TN_mg.L	0.00	-0.17	0.09
## DIC_uM.L	0.04	0.24	0.59
## DOC_uM.L	0.00	-0.24	0.16
## POC_mgperL	0.02	-0.15	0.23
## PTH_mgperL	0.02	-0.15	0.19
## Ammonium_mg.L	-0.04	-0.42	-0.24
## SRP_pM	0.00	0.00	0.00
## Salinity_TBDHereon	0.06	0.61	0.71
## PTC_mgperL	0.02	-0.13	0.24
## PTN_mgperL	0.06	-0.04	0.18
##	Turbidity_TBDHereon	SPM_mgperL	DOC_mg.L
## counts	0	0.00	0.00
## Stromkilometer	0	-0.25	-0.24
## Sat_O2_TBDHereon	0	-0.08	0.16
## Turbidity_TBDHereon	1	0.00	0.00
## SPM_mgperL	0	1.00	0.03
## DOC_mg.L	0	0.03	1.00
## DIC_mg.L	0	-0.09	0.46
## Temperature_TBDHereon	0	-0.01	0.28
## pH_TBDHereon	0	-0.04	0.26
## Silicate_mg.L	0	0.12	0.51
## Nitrite_mg.L	0	-0.07	0.19
## TotalDissolvedPhosphate_mg.L	0	0.04	0.24
## RespirationRate_O2ug.L.h	0	0.10	0.29
## Phosphate_pM	0	0.00	0.00
## O2_TBDHereon	0	-0.08	0.16
## TN_mg.L	0	0.05	0.95
## DIC_uM.L	0	-0.09	0.46
## DOC_uM.L	0	0.03	1.00
## POC_mgperL	0	0.77	0.08
## PTH_mgperL	0	0.78	0.08
## Ammonium_mg.L	0	-0.03	0.31
## SRP_pM	0	0.00	0.00
## Salinity_TBDHereon	0	-0.24	-0.02
## PTC_mgperL	0	0.77	0.09
## PTN_mgperL	0	0.58	0.08
##	Temperature_TBDHereon	pH_TBDHereon	Silicate_mg.L
## counts	0.02	0.04	-0.01
## Stromkilometer	-0.02	0.09	-0.18
## Sat_O2_TBDHereon	0.86	0.93	-0.36
## Turbidity_TBDHereon	0.00	0.00	0.00
## SPM_mgperL	-0.01	-0.04	0.12
## DOC_mg.L	0.28	0.26	0.51
## DIC_mg.L	0.52	0.57	0.34
## Temperature_TBDHereon	1.00	0.99	-0.39
## pH_TBDHereon	0.99	1.00	-0.39
## Silicate_mg.L	-0.39	-0.39	1.00
## Nitrite_mg.L	0.17	0.08	0.07
## TotalDissolvedPhosphate_mg.L	-0.60	-0.60	0.85
## RespirationRate_O2ug.L.h	0.28	0.24	0.00
## Phosphate_pM	0.00	0.00	0.00
## O2_TBDHereon	0.85	0.92	-0.35
## TN_mg.L	0.14	0.14	0.65

## DIC_uM.L	0.52	0.57	0.34
## DOC_uM.L	0.28	0.26	0.51
## POC_mgperL	0.32	0.28	-0.01
## PTH_mgperL	0.28	0.24	0.00
## Ammonium_mg.L	0.11	0.02	0.25
## SRP_pM	0.00	0.00	0.00
## Salinity_TBDHereon	0.45	0.53	-0.25
## PTC_mgperL	0.32	0.29	-0.01
## PTN_mgperL	0.21	0.20	0.00
## Nitrite_mg.L	Nitrite_mg.L	TotalDissolvedPhosphate_mg.L	
## counts	-0.03	-0.03	
## Stromkilometer	-0.38	-0.07	
## Sat_O2_TBDHereon	-0.20	-0.55	
## Turbidity_TBDHereon	0.00	0.00	
## SPM_mgperL	-0.07	0.04	
## DOC_mg.L	0.19	0.24	
## DIC_mg.L	0.17	0.24	
## Temperature_TBDHereon	0.17	-0.60	
## pH_TBDHereon	0.08	-0.60	
## Silicate_mg.L	0.07	0.85	
## Nitrite_mg.L	1.00	0.13	
## TotalDissolvedPhosphate_mg.L	0.13	1.00	
## RespirationRate_02ug.L.h	0.15	-0.13	
## Phosphate_pM	0.00	0.00	
## O2_TBDHereon	-0.20	-0.55	
## TN_mg.L	0.08	0.39	
## DIC_uM.L	0.17	0.24	
## DOC_uM.L	0.19	0.24	
## POC_mgperL	0.02	-0.13	
## PTH_mgperL	0.01	-0.10	
## Ammonium_mg.L	0.96	0.33	
## SRP_pM	0.00	0.00	
## Salinity_TBDHereon	-0.23	-0.30	
## PTC_mgperL	0.01	-0.13	
## PTN_mgperL	0.02	-0.08	
## RespirationRate_02ug.L.h	RespirationRate_02ug.L.h	Phosphate_pM	O2_TBDHereon
## counts	0.00	0	0.07
## Stromkilometer	-0.10	0	0.34
## Sat_O2_TBDHereon	0.18	0	1.00
## Turbidity_TBDHereon	0.00	0	0.00
## SPM_mgperL	0.10	0	-0.08
## DOC_mg.L	0.29	0	0.16
## DIC_mg.L	0.16	0	0.59
## Temperature_TBDHereon	0.28	0	0.85
## pH_TBDHereon	0.24	0	0.92
## Silicate_mg.L	0.00	0	-0.35
## Nitrite_mg.L	0.15	0	-0.20
## TotalDissolvedPhosphate_mg.L	-0.13	0	-0.55
## RespirationRate_02ug.L.h	1.00	0	0.16
## Phosphate_pM	0.00	1	0.00
## O2_TBDHereon	0.16	0	1.00
## TN_mg.L	0.16	0	0.10
## DIC_uM.L	0.16	0	0.59
## DOC_uM.L	0.29	0	0.16

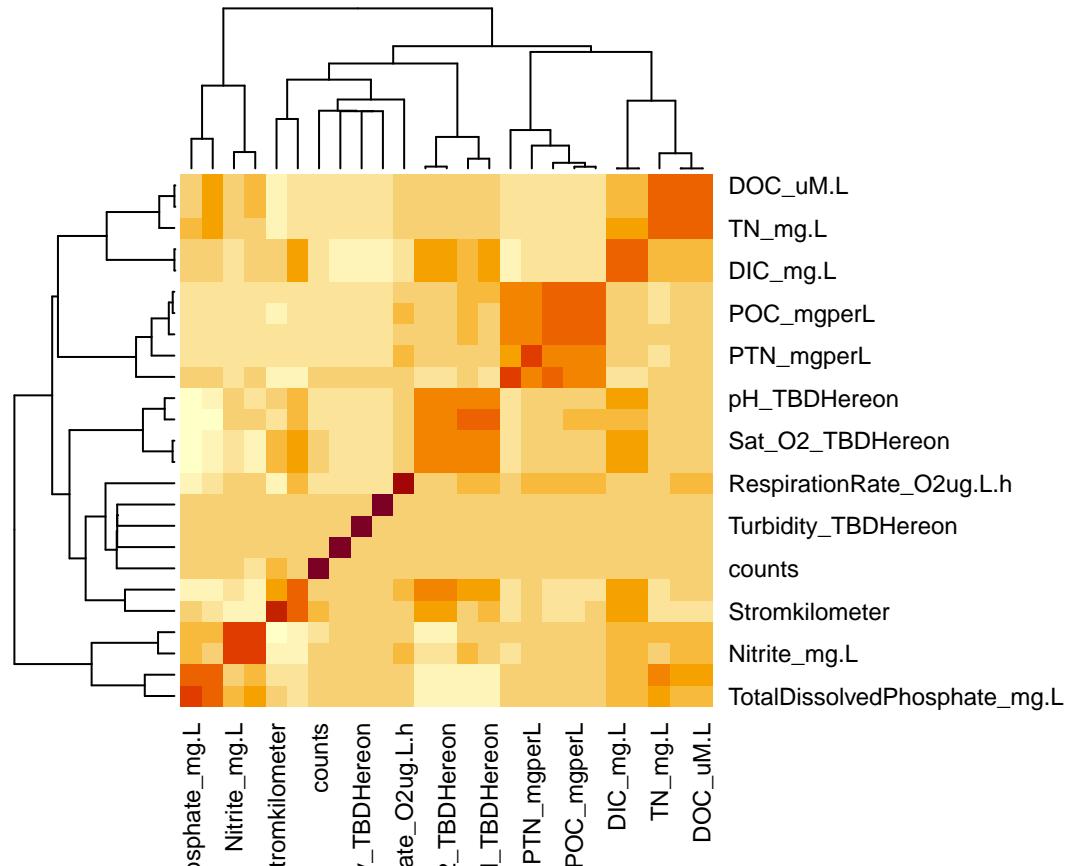
## POC_mgperL		0.29	0	0.22	
## PTH_mgperL		0.26	0	0.19	
## Ammonium_mg.L		0.13	0	-0.24	
## SRP_pM		0.00	0	0.00	
## Salinity_TBDHereon		0.24	0	0.68	
## PTC_mgperL		0.28	0	0.23	
## PTN_mgperL		0.31	0	0.18	
##	TN_mg.L	DIC_uM.L	DOC_uM.L	POC_mgperL	PTH_mgperL
## counts	0.00	0.04	0.00	0.02	0.02
## Stromkilometer	-0.17	0.24	-0.24	-0.15	-0.15
## Sat_02_TBDHereon	0.09	0.59	0.16	0.23	0.19
## Turbidity_TBDHereon	0.00	0.00	0.00	0.00	0.00
## SPM_mgperL	0.05	-0.09	0.03	0.77	0.78
## DOC_mg.L	0.95	0.46	1.00	0.08	0.08
## DIC_mg.L	0.48	1.00	0.46	0.12	0.08
## Temperature_TBDHereon	0.14	0.52	0.28	0.32	0.28
## pH_TBDHereon	0.14	0.57	0.26	0.28	0.24
## Silicate_mg.L	0.65	0.34	0.51	-0.01	0.00
## Nitrite_mg.L	0.08	0.17	0.19	0.02	0.01
## TotalDissolvedPhosphate_mg.L	0.39	0.24	0.24	-0.13	-0.10
## RespirationRate_02ug.L.h	0.16	0.16	0.29	0.29	0.26
## Phosphate_pM	0.00	0.00	0.00	0.00	0.00
## O2_TBDHereon	0.10	0.59	0.16	0.22	0.19
## TN_mg.L	1.00	0.48	0.95	0.05	0.06
## DIC_uM.L	0.48	1.00	0.46	0.12	0.08
## DOC_uM.L	0.95	0.46	1.00	0.08	0.08
## POC_mgperL	0.05	0.12	0.08	1.00	0.99
## PTH_mgperL	0.06	0.08	0.08	0.99	1.00
## Ammonium_mg.L	0.22	0.27	0.31	0.03	0.02
## SRP_pM	0.00	0.00	0.00	0.00	0.00
## Salinity_TBDHereon	-0.07	0.53	-0.02	-0.04	-0.09
## PTC_mgperL	0.06	0.13	0.09	1.00	0.99
## PTN_mgperL	0.06	0.11	0.08	0.76	0.75
##	Ammonium_mg.L	SRP_pM	Salinity_TBDHereon	PTC_mgperL	
## counts	-0.04	0	0.06	0.02	
## Stromkilometer	-0.42	0	0.61	-0.13	
## Sat_02_TBDHereon	-0.24	0	0.71	0.24	
## Turbidity_TBDHereon	0.00	0	0.00	0.00	
## SPM_mgperL	-0.03	0	-0.24	0.77	
## DOC_mg.L	0.31	0	-0.02	0.09	
## DIC_mg.L	0.27	0	0.53	0.13	
## Temperature_TBDHereon	0.11	0	0.45	0.32	
## pH_TBDHereon	0.02	0	0.53	0.29	
## Silicate_mg.L	0.25	0	-0.25	-0.01	
## Nitrite_mg.L	0.96	0	-0.23	0.01	
## TotalDissolvedPhosphate_mg.L	0.33	0	-0.30	-0.13	
## RespirationRate_02ug.L.h	0.13	0	0.24	0.28	
## Phosphate_pM	0.00	0	0.00	0.00	
## O2_TBDHereon	-0.24	0	0.68	0.23	
## TN_mg.L	0.22	0	-0.07	0.06	
## DIC_uM.L	0.27	0	0.53	0.13	
## DOC_uM.L	0.31	0	-0.02	0.09	
## POC_mgperL	0.03	0	-0.04	1.00	
## PTH_mgperL	0.02	0	-0.09	0.99	

```

## Ammonium_mg.L          1.00    0      -0.30    0.02
## SRP_pM                  0.00    1       0.00    0.00
## Salinity_TBDHeron        -0.30   0       1.00    -0.03
## PTC_mgperL                0.02   0      -0.03    1.00
## PTN_mgperL                0.02   0       0.00    0.75
##                                     PTN_mgperL
## counts                      0.06
## Stromkilometer              -0.04
## Sat_02_TBDHeron             0.18
## Turbidity_TBDHeron           0.00
## SPM_mgperL                  0.58
## DOC_mg.L                     0.08
## DIC_mg.L                     0.11
## Temperature_TBDHeron         0.21
## pH_TBDHeron                  0.20
## Silicate_mg.L                 0.00
## Nitrite_mg.L                  0.02
## TotalDissolvedPhosphate_mg.L -0.08
## RespirationRate_02ug.L.h      0.31
## Phosphate_pM                  0.00
## O2_TBDHeron                   0.18
## TN_mg.L                      0.06
## DIC_uM.L                      0.11
## DOC_uM.L                      0.08
## POC_mgperL                    0.76
## PTH_mgperL                    0.75
## Ammonium_mg.L                  0.02
## SRP_pM                         0.00
## Salinity_TBDHeron               0.00
## PTC_mgperL                     0.75
## PTN_mgperL                     1.00

```

```
heatmap(round(res, 2))
```



```
#no interresting correlations for PsbA counts
```

2.3 Variance analysis

```
#variance analysis
var_meta_psbA_clust <- meta_psbA_clust %>%
  filter(Station != 'BunthausSpitze') %>%
  filter(Sample_date != "Nov 21") %>%
  group_by(Station, Sample_date) %>%
  summarise(var = var(counts))

## `summarise()` has grouped output by 'Station'. You can override using the
## `.` argument.

var_meta_psbA_clust_station <- meta_psbA_clust %>%
  filter(Station != 'BunthausSpitze') %>%
  filter(Sample_date != "Nov 21") %>%
  group_by(Station) %>%
  summarise(var = var(counts))

#variance across stations (similarly the variance is lower n the brakish water)
var_meta_psbA_clust_station %>% arrange(factor(Station, levels = c("Muehlenberger Loch", "Twielensfleth")))
```

```
## # A tibble: 5 x 2
##   Station           var
##   <chr>            <dbl>
## 1 Twielenfleth    4478928997.
## 2 Schwarztonnensand 20865462587.
## 3 Meedem Grund    116304863625.
## 4 Brunsbüttel      99637195528.
## 5 Mühlenberger Loch  5495165096.
```

The variance is increasing with a station order