Omics Analysis Report 69 subjects

Powered by MicrobiomeStat (Ver 1.1.1)

2023-09-15

Contents

1. Data overview and summary statistics

1.1 Parameter setting

Parameter	Value
data.obj	subset_T2D.obj
feature.dat.type	count
group.var	subject_race
test.adj.vars	NULL
vis.adj.vars	NULL
strata.var	subject_gender
subject.var	subject_id
time.var	visit_number_num
t0.level	NULL
ts.levels	NULL
alpha.obj	NULL
alpha.name	shannon, observed_species
depth	NULL
dist.obj	NULL
dist.name	BC, Jaccard
pc.obj	NULL
prev.filter	0.1
abund.filter	1e-04
feature.analysis.rarafy	TRUE
feature.change.func	relative change
bar.area.feature.no	40
heatmap.feature.no	40
vis.feature.level	Family, Genus
test.feature.level	Family
feature.mt.method	none
feature.sig.level	0.3
feature.box.axis.transform	sqrt
base.size	20
theme.choice	bw
custom.theme	NULL
palette	NULL
pdf	TRUE
file.ann	NULL
pdf.wid	11
pdf.hei	8.5

1.2 Summary statistics

```
mStat_results <- mStat_summarize_data_obj(data.obj = data.obj,
                                          time.var = time.var,
                                          group.var = group.var,
                                          palette = palette)
```

Histogram of Sample Counts over Time

Boxplot of Sequencing Depth over Time
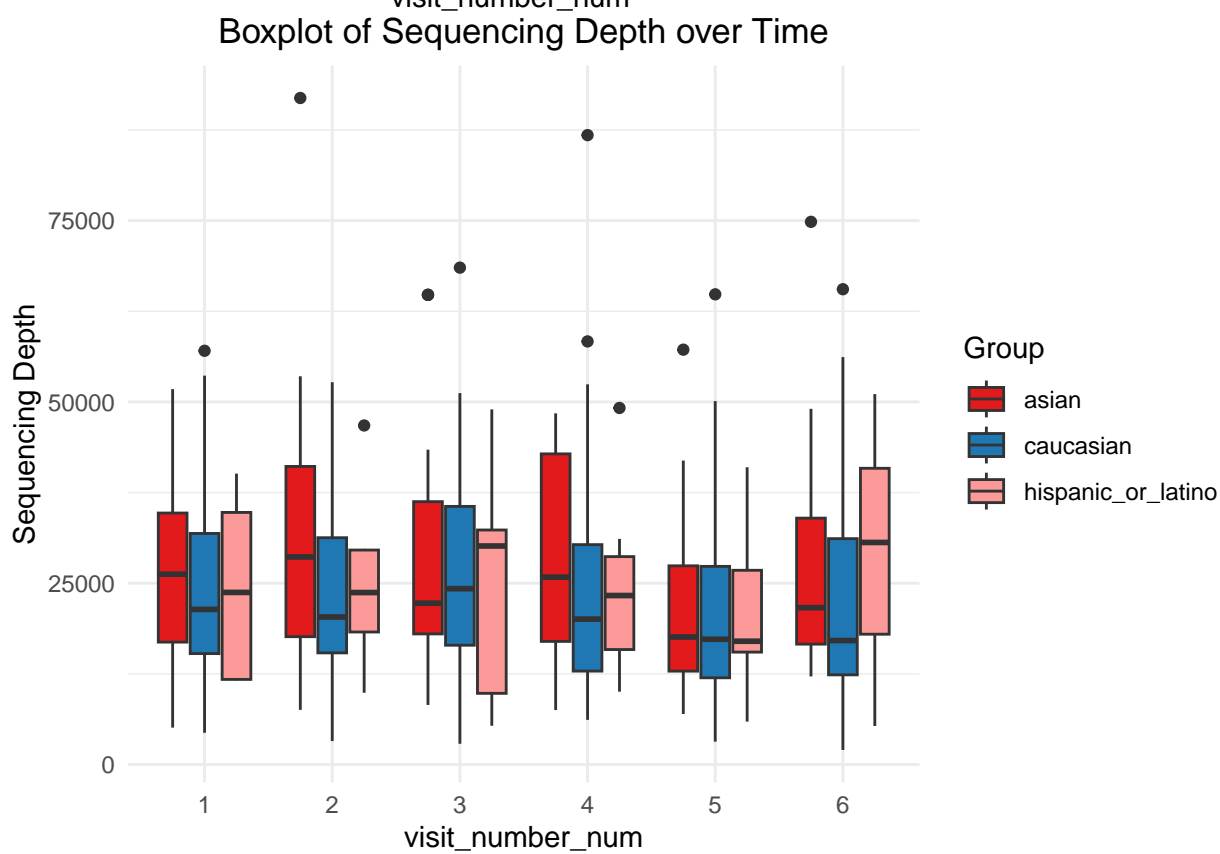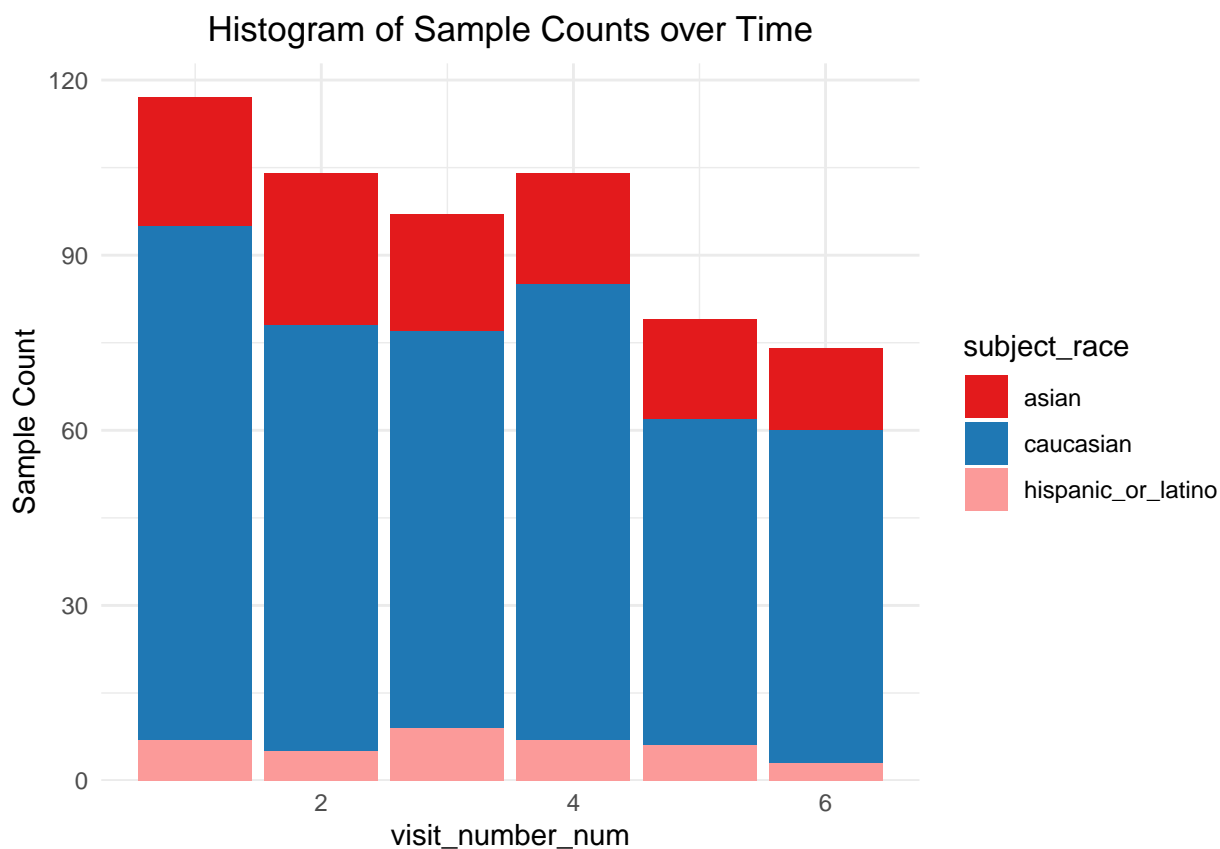
Category	Variable	Value
Basic Statistics	Number of samples	575
Basic Statistics	Number of features	9533
Basic Statistics	Min. reads per sample	2007
Basic Statistics	Max. reads per sample	91908
Basic Statistics	Total reads across all samples	14138179
Basic Statistics	Average reads per sample	1483.078
Basic Statistics	Median reads per sample	21062
Basic Statistics	Proportion of zero counts	0.963
Basic Statistics	Count of features that only appear once	1505
Metadata	Number of metadata variables	14
Feature Annotations	Proportion of missing annotations in Kingdom	0
Feature Annotations	Proportion of missing annotations in Phylum	0
Feature Annotations	Proportion of missing annotations in Class	0.002
Feature Annotations	Proportion of missing annotations in Order	0.012
Feature Annotations	Proportion of missing annotations in Family	0.128
Feature Annotations	Proportion of missing annotations in Genus	0.484
Feature Annotations	Proportion of missing annotations in Species	0.887
Phylogenetic Tree	Exists in the dataset	No
Time-Series Information	Earliest sample time-point	1
Time-Series Information	Latest sample time-point	6
Distribution of sample counts	Sample Count at Time-point: 1	117
Distribution of sample counts	Sample Count at Time-point: 2	104
Distribution of sample counts	Sample Count at Time-point: 3	97
Distribution of sample counts	Sample Count at Time-point: 4	104
Distribution of sample counts	Sample Count at Time-point: 5	79
Distribution of sample counts	Sample Count at Time-point: 6	74

No rarefaction depth is specified. The minimum depth, 2007, is used as the rarefaction depth. After rarefaction, 575 samples remain in the analysis. alpha.obj is calculated based on the rarefied data.obj. dist.obj is calculated based on the rarefied data.obj. pc.obj is calculated based on the dist.obj.

1.3 Data visualization(overall)

Rarefaction has been enabled for feature-level analysis.

Reason: The observed abundance of rare/low-abundance features can be strongly influenced by sequence depth. Rarefaction is an effective method to control the effect of sequence depth variation. By employing rarefaction, we can potentially increase the power of detecting those rare/low-abundance features, even though it introduces some variation due to under-sampling. In essence, this step helps to ensure more accurate and consistent results across samples with varying sequence depths.

If you do not wish to perform rarefaction during feature-level analysis, please turn feature.analysis.rarafy to FALSE.

1.3.1 Feature areaplot

```
taxa_areaplot_long_results <- generate_taxa_areaplot_long(
  data.obj = data.obj,
  subject.var = subject.var,
  time.var = time.var,
  t0.level = t0.level,
```

```
    ts.levels = ts.levels,
    group.var = group.var,
    strata.var = strata.var,
    feature.level = vis.feature.level,
    feature.dat.type = feature.dat.type,
    feature.number = bar.area.feature.no,
    base.size = base.size,
    theme.choice = theme.choice,
    custom.theme = custom.theme,
    palette = palette,
    pdf = pdf,
    file.ann = file.ann,
    pdf.wid = pdf.wid,
    pdf.hei = pdf.hei
)
```

The following plots display the average proportions for each time point, group, and stratum.

$Family

$Genus
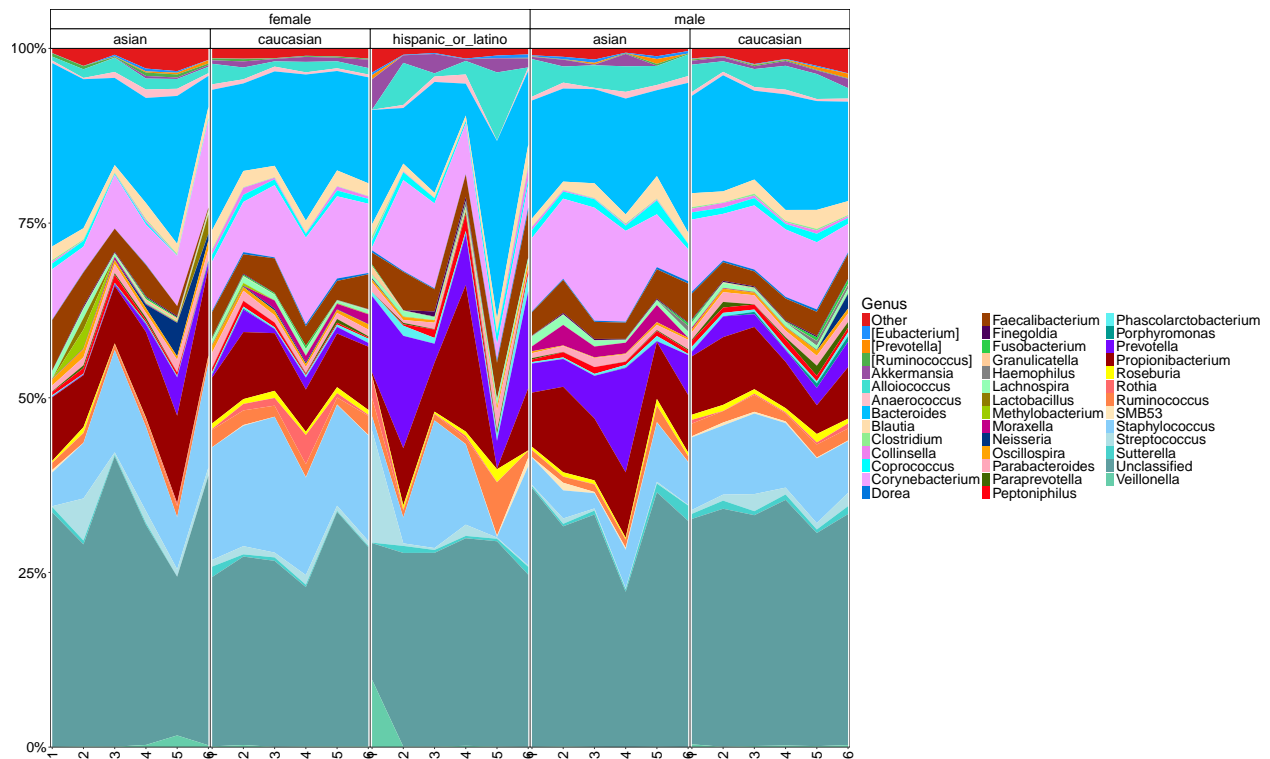
1.3.2 Feature heatmap

```
taxa_heatmap_long_results <- generate_taxa_heatmap_long(
  data.obj = data.obj,
  subject.var = subject.var,
  time.var = time.var,
  t0.level = t0.level,
  ts.levels = ts.levels,
  group.var = group.var,
  strata.var = strata.var,
  feature.level = vis.feature.level,
  feature.dat.type = feature.dat.type,
  features.plot = NULL,
  top.k.plot = heatmap.feature.no,
  top.k.func = 'mean',
  prev.filter = prev.filter,
  abund.filter = abund.filter,
  base.size = base.size,
  palette = palette,
  cluster.cols = NULL,
  cluster.rows = NULL,
  pdf = pdf,
  file.ann = file.ann,
  pdf.wid = pdf.wid,
  pdf.hei = pdf.hei
)
```

The following plots display the average proportions for each time point, group, and stratum.

$Family

$Genus

The following plots display the individual proportions for each time point, group, and stratum.

$Family
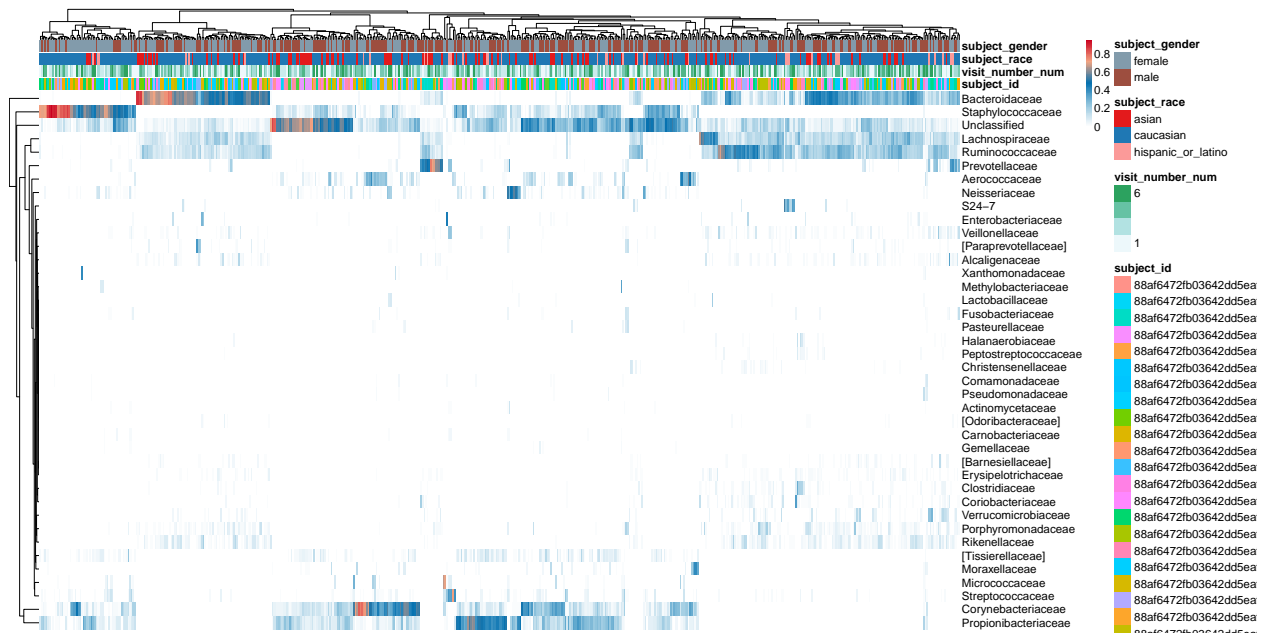
$Genus

1.3.3 Feature change heatmap

```r
taxa_change_heatmap_long_results <- generate_taxa_change_heatmap_long(
  data.obj = data.obj,
  subject.var = subject.var,
  time.var = time.var,
  t0.level = t0.level,
  ts.levels = ts.levels,
  group.var = group.var,
  strata.var = strata.var,
  feature.level = vis.feature.level,
```
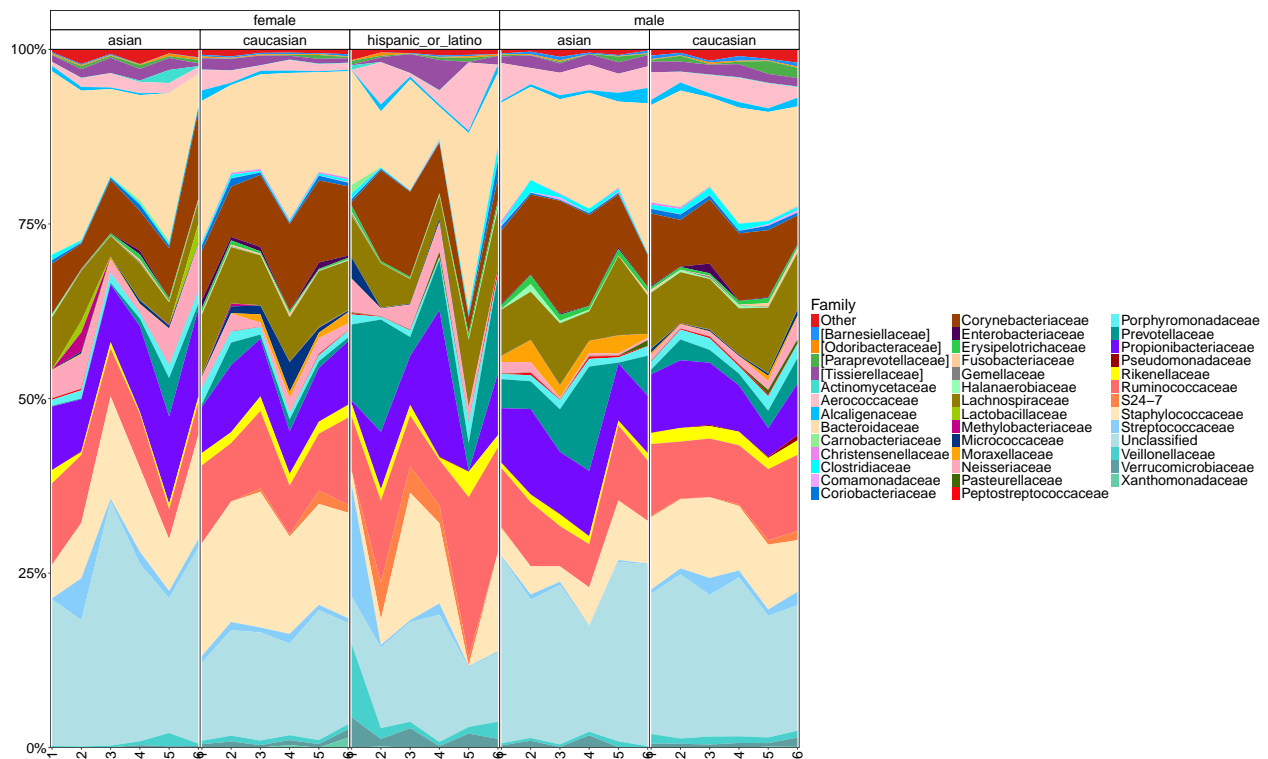
```
    feature.dat.type = feature.dat.type,
    features.plot = NULL,
    top.k.plot = heatmap.feature.no,
    top.k.func = 'mean',
    feature.change.func = feature.change.func,
    prev.filter = prev.filter,
    abund.filter = abund.filter,
    base.size = base.size,
    palette = palette,
    cluster.cols = NULL,
    cluster.rows = NULL,
    pdf = pdf,
    file.ann = file.ann,
    pdf.wid = pdf.wid,
    pdf.hei = pdf.hei
)
```

The changes from t0.level were computed as the difference between the current value and t0.level divided by the sum of the two.The following plots display the average changes for each time point, group, and stratum.

$Family

$Genus

1.3.4 Feature barplot

```
taxa_barplot_long_results <- generate_taxa_barplot_long(
  data.obj = data.obj,
  subject.var = subject.var,
  time.var = time.var,
  group.var = group.var,
  strata.var = strata.var,
  feature.level = vis.feature.level,
  feature.dat.type = feature.dat.type,
  feature.number = bar.area.feature.no,
  t0.level = t0.level,
  ts.levels = ts.levels,
  base.size = base.size,
  theme.choice = theme.choice,
  custom.theme = custom.theme,
  palette = palette,
  pdf = pdf,
  file.ann = file.ann,
  pdf.wid = pdf.wid,
  pdf.hei = pdf.hei
)
```

The following plots display the average proportions for each time point, group, and stratum.
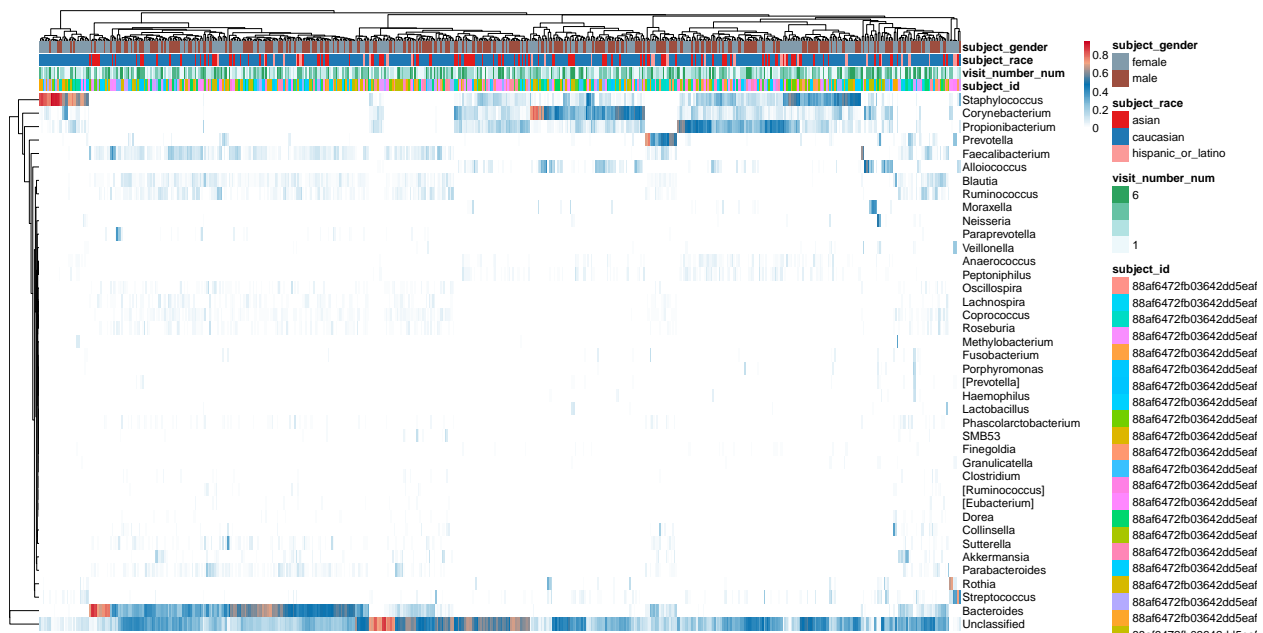
$Family

$Genus

2. Alpha diversity analysis

2.1 Data visualization

2.1.1 Alpha diversity boxplot

```
alpha_boxplot_results <- generate_alpha_boxplot_long(data.obj = data.obj,
                                                     alpha.obj = alpha.obj,
                                                     alpha.name = alpha.name,
                                                     depth = depth,
                                                     subject.var = subject.var,
                                                     time.var = time.var,
                                                     t0.level = t0.level,
                                                     ts.levels = ts.levels,
                                                     group.var = group.var,
                                                     strata.var = strata.var,
                                                     adj.vars = vis.adj.vars,
                                                     base.size = base.size,
                                                     theme.choice = theme.choice,
                                                     custom.theme = custom.theme,
                                                     palette = palette,
                                                     pdf = pdf,
                                                     file.ann = file.ann,
                                                     pdf.wid = pdf.wid,
                                                     pdf.hei = pdf.hei)
alpha_boxplot_results
```
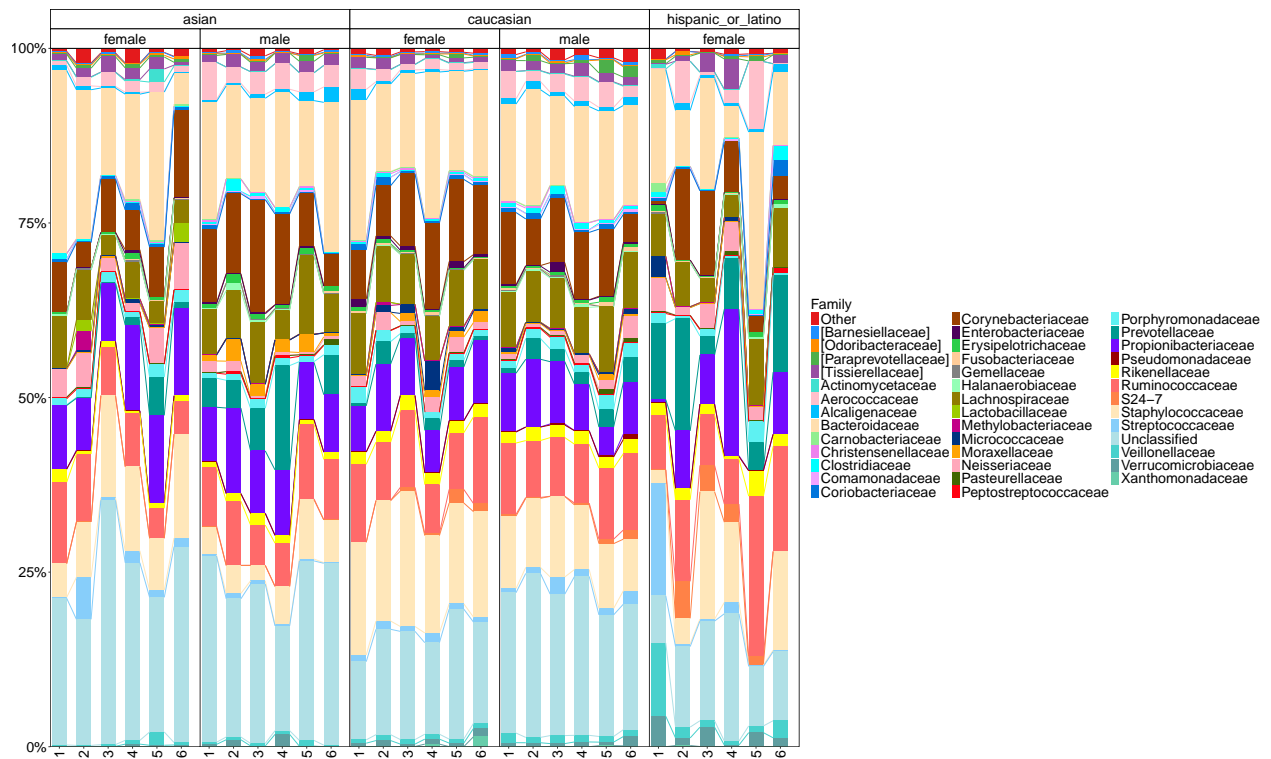

$shannon

$observed_species

2.1.2 Alpha diversity spaghettiplot

```
alpha_spaghettiplot_results <- generate_alpha_spaghettiplot_long(
                                        data.obj = data.obj,
                                        alpha.obj = alpha.obj,
                                        alpha.name = alpha.name,
                                        depth = depth,
                                        subject.var = subject.var,
                                        time.var = time.var,
                                        t0.level = t0.level,
                                        ts.levels = ts.levels,
                                        group.var = group.var,
                                        strata.var = strata.var,
                                        adj.vars = vis.adj.vars,
                                        base.size = base.size,
                                        theme.choice = theme.choice,
                                        custom.theme = custom.theme,
                                        palette = palette,
                                        pdf = pdf,
```
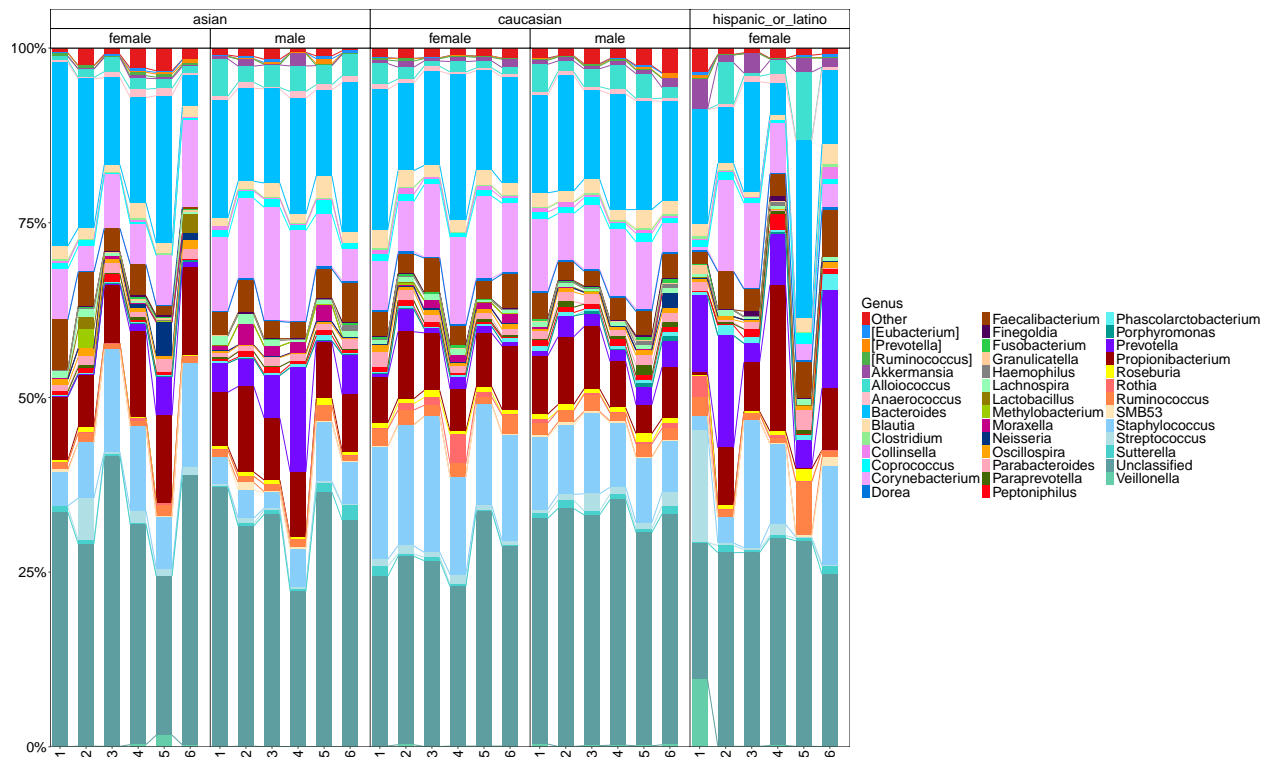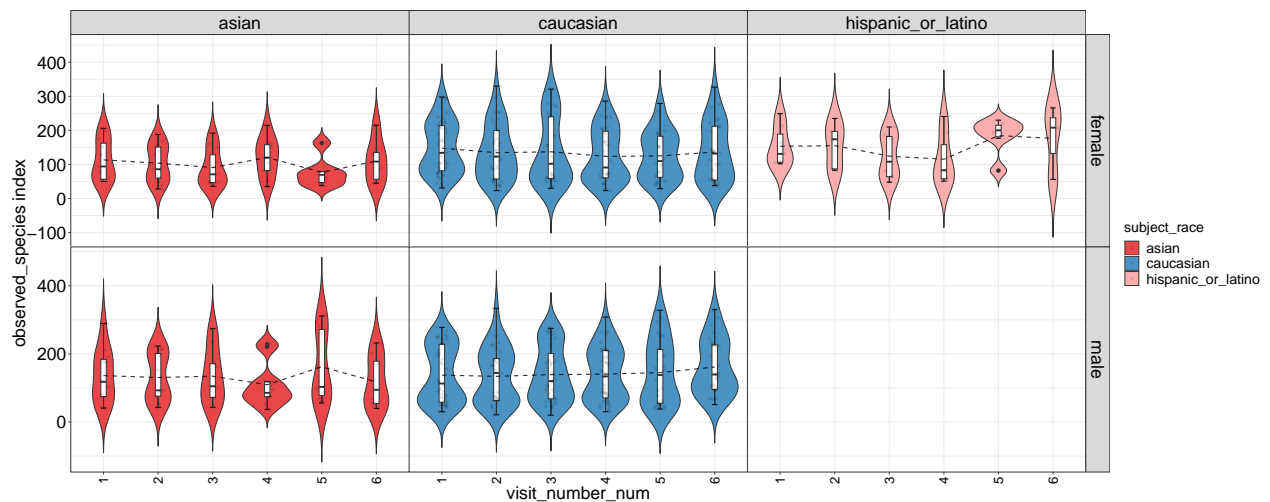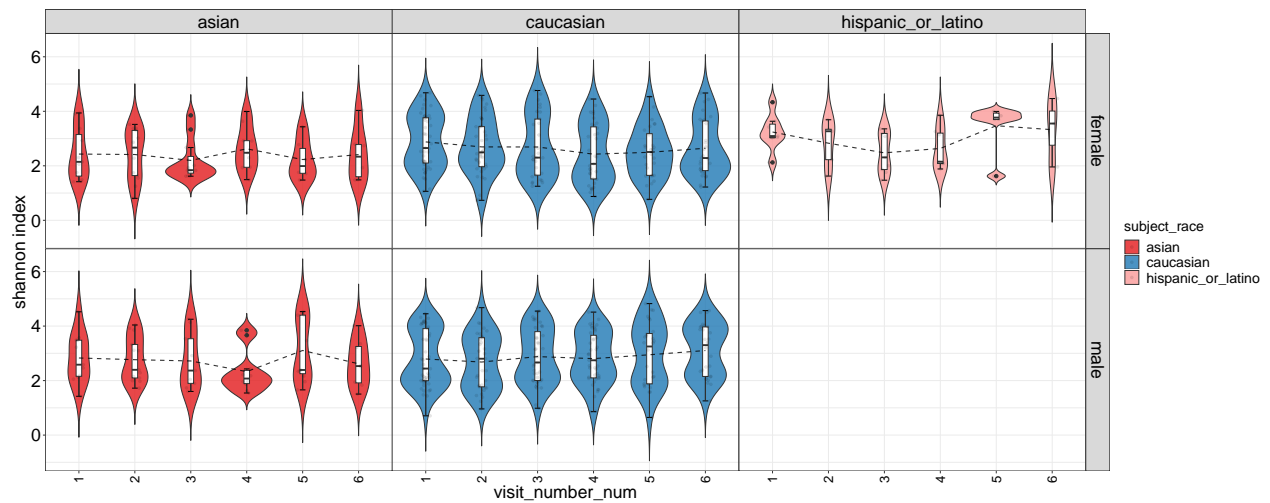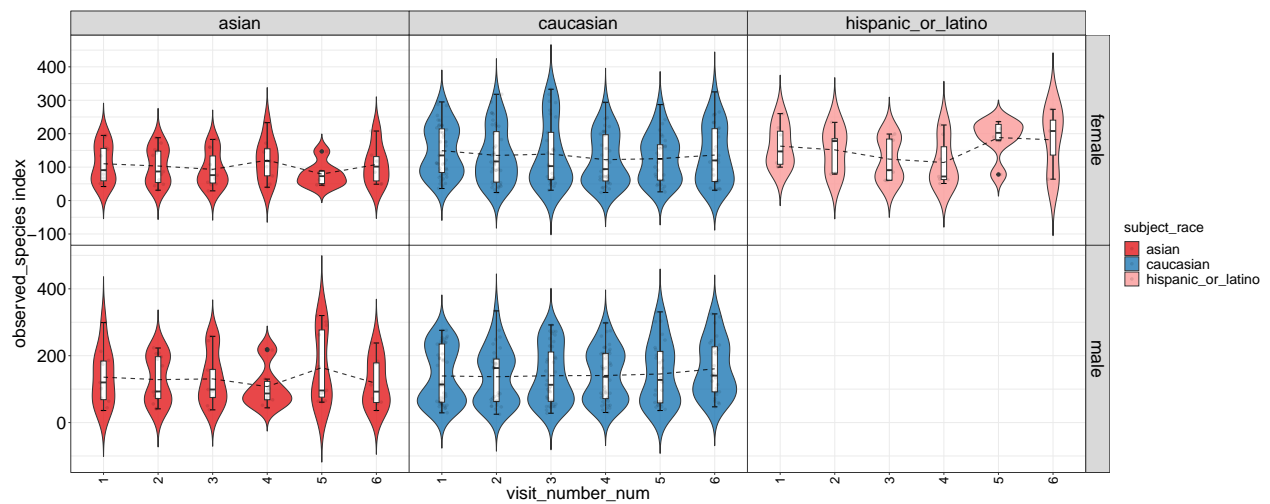
```
                                             file.ann = file.ann,
                                             pdf.wid = pdf.wid,
                                             pdf.hei = pdf.hei)
alpha_spaghettiplot_results
```
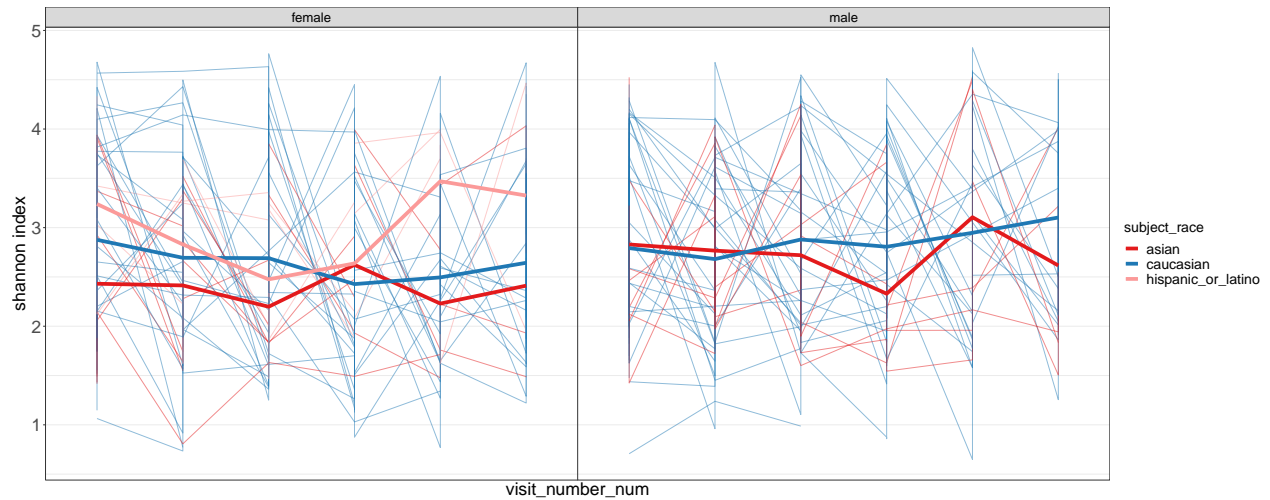
$shannon

$observed_species
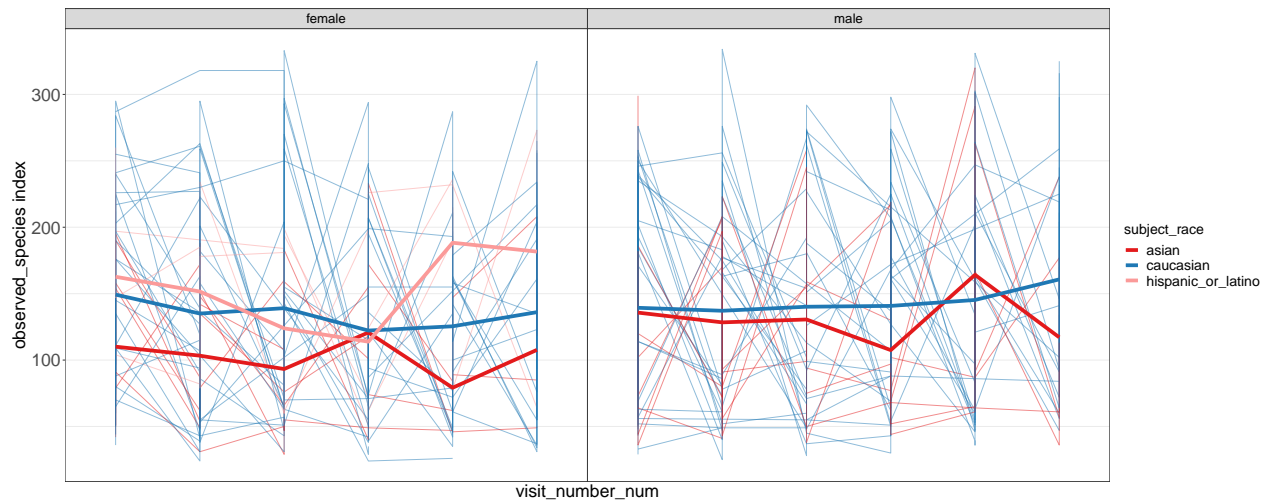
2.2 Trend test

```
alpha_trend_test_results <- generate_alpha_trend_test_long(
                                             data.obj = data.obj,
                                             alpha.obj = alpha.obj,
                                             alpha.name = alpha.name,
                                             depth = depth,
                                             time.var = time.var,
                                             subject.var = subject.var,
                                             group.var = group.var,
                                             adj.vars = test.adj.vars)
```

In this analysis, we utilized a linear mixed effects model with both a random intercept and a random slope to

investigate a potential difference in trend. Specifically, we included visit_number_num, subject_race and tested the interaction between the variables subject_race and visit_number_num.

2.2.1 Shannon index

An ANOVA test of the null hypothesis of no difference among the 3 groups produces a p-value of 0.830.

Table 3: Table continues below

Term	Estimate	Std.Error
(Intercept)	2.584	0.2041
subject_racecaucasian	0.2121	0.2307
subject_racehispanic_or_latino	0.1975	0.4402
visit_number_num	-0.007002	0.05486
subject_racecaucasian:visit_number_num	-0.01105	0.06188
subject_racehispanic_or_latino:visit_number_num	0.05312	0.1165
subject_race:visit_number_num	NA	NA

Statistic	P.Value
12.66	3.162e-27
0.9194	0.359
0.4487	0.6543
-0.1276	0.8985
-0.1786	0.8584
0.4558	0.6488
0.1862	0.8302

2.2.2 Observed_species index

An ANOVA test of the null hypothesis of no difference among the 3 groups produces a p-value of 0.881.

Table 5: Table continues below

Term	Estimate	Std.Error
(Intercept)	117.5	16.13
subject_racecaucasian	23.36	18.23
subject_racehispanic_or_latino	19.31	34.69
visit_number_num	0.1069	4.39
subject_racecaucasian:visit_number_num	-0.9141	4.946
subject_racehispanic_or_latino:visit_number_num	3.269	9.375
subject_race:visit_number_num	NA	NA

Statistic	P.Value
7.285	3.998e-12
1.281	0.2013
0.5566	0.5784
0.02435	0.9806
-0.1848	0.8535
0.3487	0.7275

Statistic	P.Value
0.1265	0.8812

2.3 Volatility test

```
alpha_volatility_test_results <- generate_alpha_volatility_test_long(
                                     data.obj = data.obj,
                                     alpha.obj = alpha.obj,
                                     alpha.name = alpha.name,
                                     depth = depth,
                                     time.var = time.var,
                                     subject.var = subject.var,
                                     group.var = group.var,
                                     adj.vars = test.adj.vars)
```

In this analysis, we employed a general linear model followed by ANOVA to test the effect of subject_race on volatility. The alpha diversity volatility is calculated by averaging the rate of change in alpha diversity across consecutive time points. Specifically, for each pair of adjacent time points, we compute the difference in alpha diversity, normalize it by the time difference, and then take the average over all such pairs.

2.3.1 Shannon index

Based on the general linear model, the level caucasian of the variable subject_race did not significantly differ from level asian, with a p-value of 0.407. Based on the general linear model, the level hispanic_or_latino of the variable subject_race did not significantly differ from level asian, with a p-value of 0.803. An ANOVA test of the null hypothesis of no difference in volatility among 3 groups produces a p-value of 0.705.

Term	Estimate	Std.Error	Statistic	P.Value
(Intercept)	1.043	0.1712	6.09	8.232e-08
subject_racecaucasian	0.1612	0.193	0.8354	0.4067
subject_racehispanic_or_latino	0.09899	0.3954	0.2504	0.8031
subject_race	NA	NA	0.3519	0.7048
Residuals	NA	NA	NA	NA

2.3.2 Observed_species index

Based on the general linear model, the level caucasian of the variable subject_race did not significantly differ from level asian, with a p-value of 0.296. Based on the general linear model, the level hispanic_or_latino of the variable subject_race did not significantly differ from level asian, with a p-value of 0.651. An ANOVA test of the null hypothesis of no difference in volatility among 3 groups produces a p-value of 0.576.

| Term	Estimate	Std.Error	Statistic	P.Value
(Intercept)	81.37	13.32	6.108	7.683e-08
subject_racecaucasian	15.84	15.02	1.055	0.2958
subject_racehispanic_or_latino	13.99	30.77	0.4548	0.6508
subject_race	NA	NA	0.5574	0.5756
Residuals	NA	NA	NA	NA

3. Beta diversity analysis

3.1 Data visualization

3.1.1 Beta diversity ordinationplot

```r
beta_ordination_results <- generate_beta_ordination_long(data.obj = data.obj,
                                                         dist.obj = dist.obj,
                                                         pc.obj = pc.obj,
                                                         subject.var = subject.var,
                                                         time.var = time.var,
                                                         t0.level = t0.level,
                                                         ts.levels = ts.levels,
                                                         group.var = group.var,
                                                         strata.var = NULL,
                                                         adj.vars = vis.adj.vars,
                                                         dist.name = dist.name,
                                                         base.size = base.size,
                                                         theme.choice = theme.choice,
                                                         custom.theme = custom.theme,
                                                         palette = palette,
                                                         pdf = pdf,
                                                         file.ann = file.ann,
                                                         pdf.wid = pdf.wid,
                                                         pdf.hei = pdf.hei)
beta_ordination_results
```

$BC
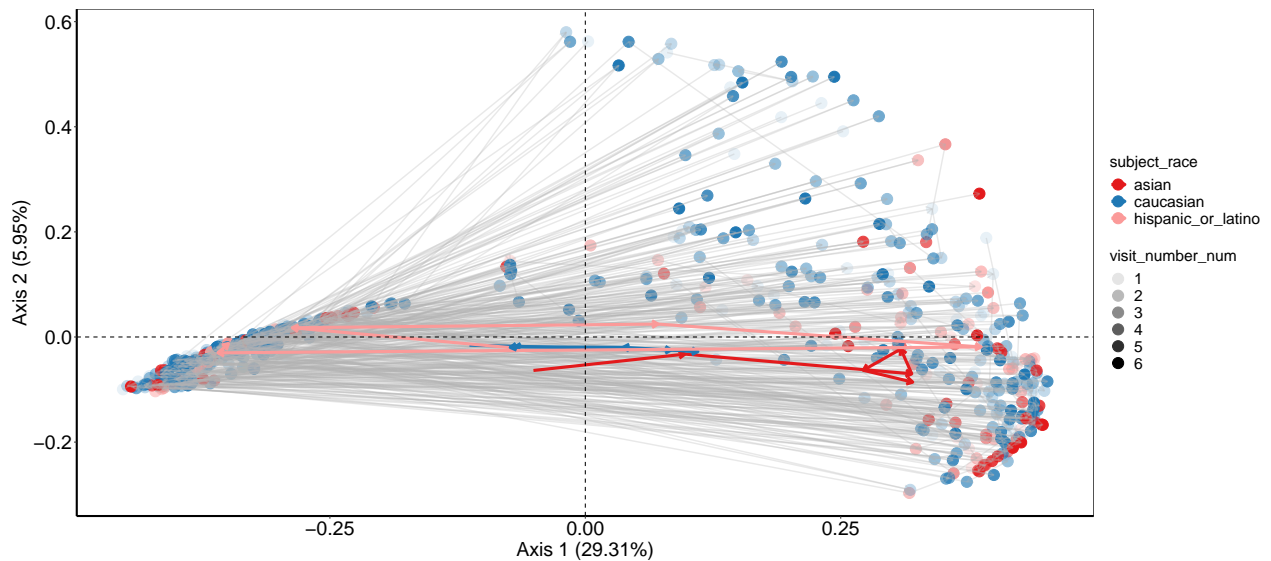
$Jaccard

```
if (!is.null(strata.var)){
  beta_ordination_stratified_results <- generate_beta_ordination_long(data.obj = data.obj,
                                          dist.obj = dist.obj,
                                          pc.obj = pc.obj,
                                          subject.var = subject.var,
                                          time.var = time.var,
                                          t0.level = t0.level,
                                          ts.levels = ts.levels,
                                          group.var = group.var,
                                          strata.var = strata.var,
                                          adj.vars = vis.adj.vars,
                                          dist.name = dist.name,
                                          base.size = base.size,
                                          theme.choice = theme.choice,
                                          custom.theme = custom.theme,
                                          palette = palette,
                                          pdf = pdf,
                                          file.ann = file.ann,
                                          pdf.wid = pdf.wid,
                                          pdf.hei = pdf.hei)
beta_ordination_stratified_results
}
```
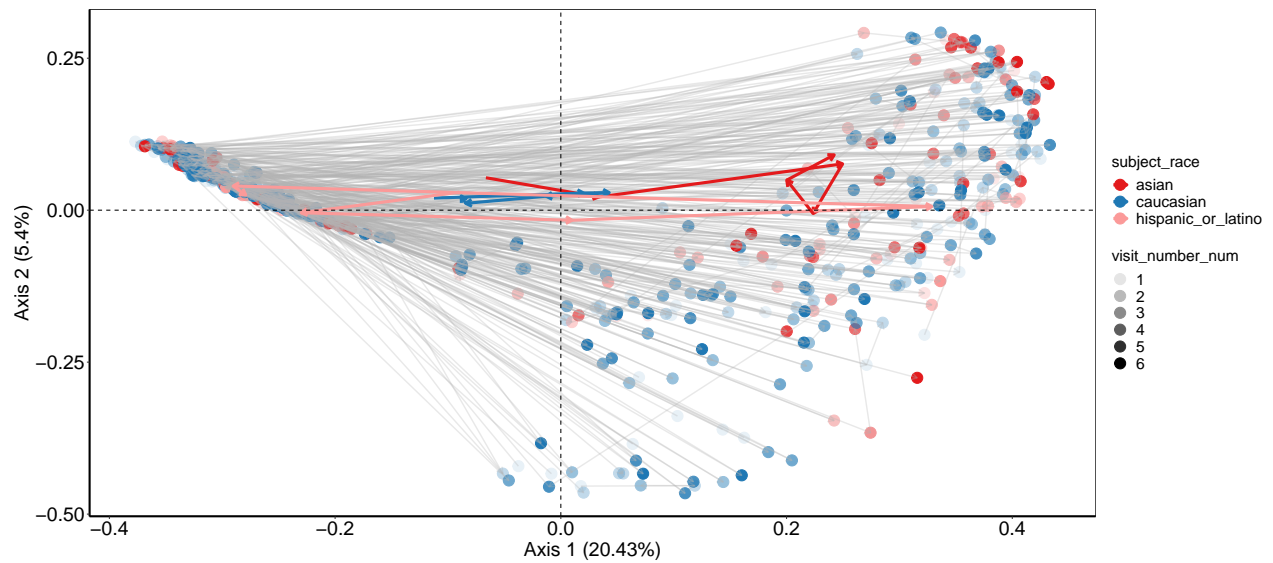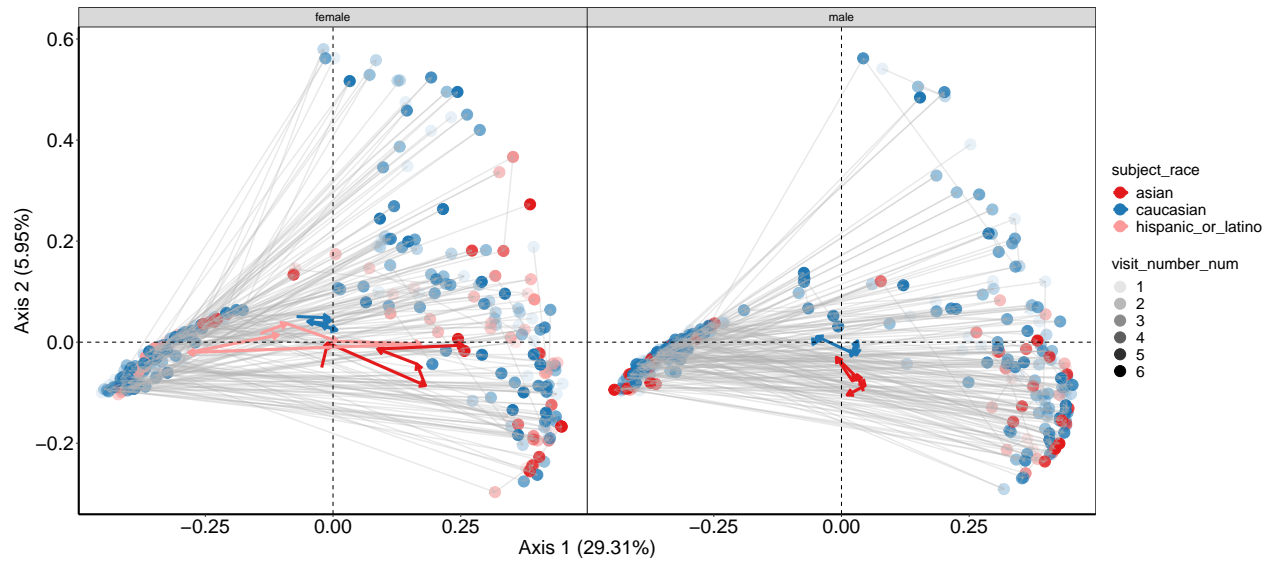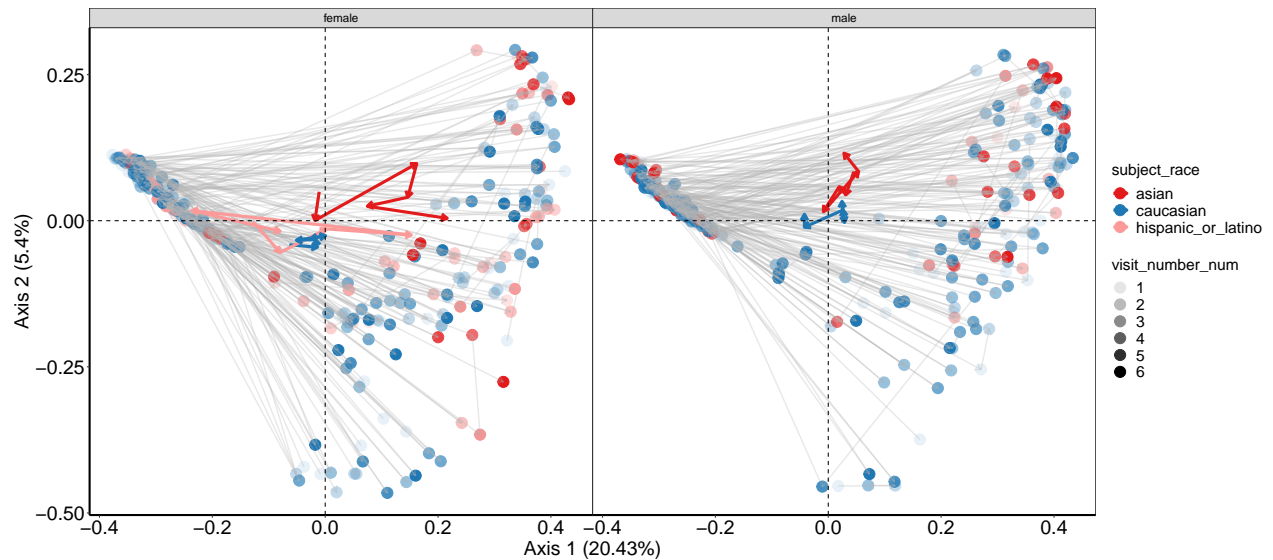
$BC

18

$Jaccard

3.1.2 Beta diversity principal coordinate spaghettiplot

```
pc_boxplot_longitudinal_results <- generate_beta_pc_spaghettiplot_long(
  data.obj = data.obj,
  dist.obj = dist.obj,
  pc.obj = pc.obj,
  pc.ind = c(1, 2),
  subject.var = subject.var,
  time.var = time.var,
  t0.level = t0.level,
  ts.levels = ts.levels,
  group.var = group.var,
  strata.var = strata.var,
  adj.vars = vis.adj.vars,
  dist.name = dist.name,
  base.size = base.size,
```
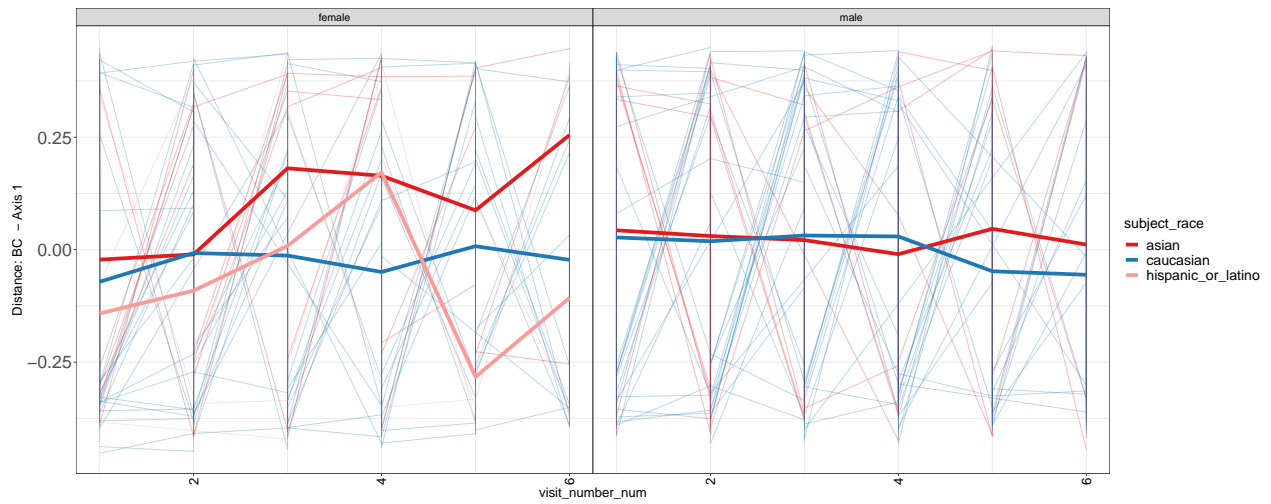
```
  theme.choice = theme.choice,
  custom.theme = custom.theme,
  palette = palette,
  pdf = pdf,
  file.ann = file.ann,
  pdf.wid = pdf.wid,
  pdf.hei = pdf.hei
)

pc_boxplot_longitudinal_results
```

$BC *BC*PC1

*BC*PC2

$Jaccard *Jaccard*PC1

*Jaccard*PC2
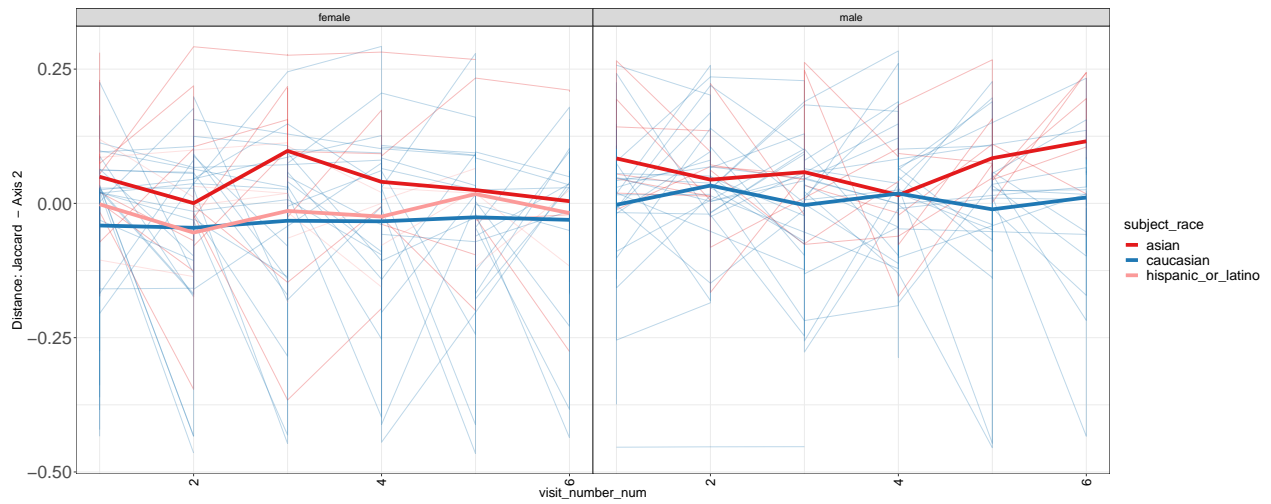
3.1.3 Beta diversity change spaghettiplot

```
spaghettiplot_longitudinal_results <- generate_beta_change_spaghettiplot_long(
  data.obj = data.obj,
  dist.obj = dist.obj,
  subject.var = subject.var,
  time.var = time.var,
  t0.level = t0.level,
  ts.levels = ts.levels,
  group.var = group.var,
  strata.var = strata.var,
  adj.vars = vis.adj.vars,
  dist.name = dist.name,
  base.size = base.size,
  theme.choice = theme.choice,
  custom.theme = custom.theme,
  palette = palette,
  pdf = pdf,
  file.ann = file.ann,
  pdf.wid = pdf.wid,
```
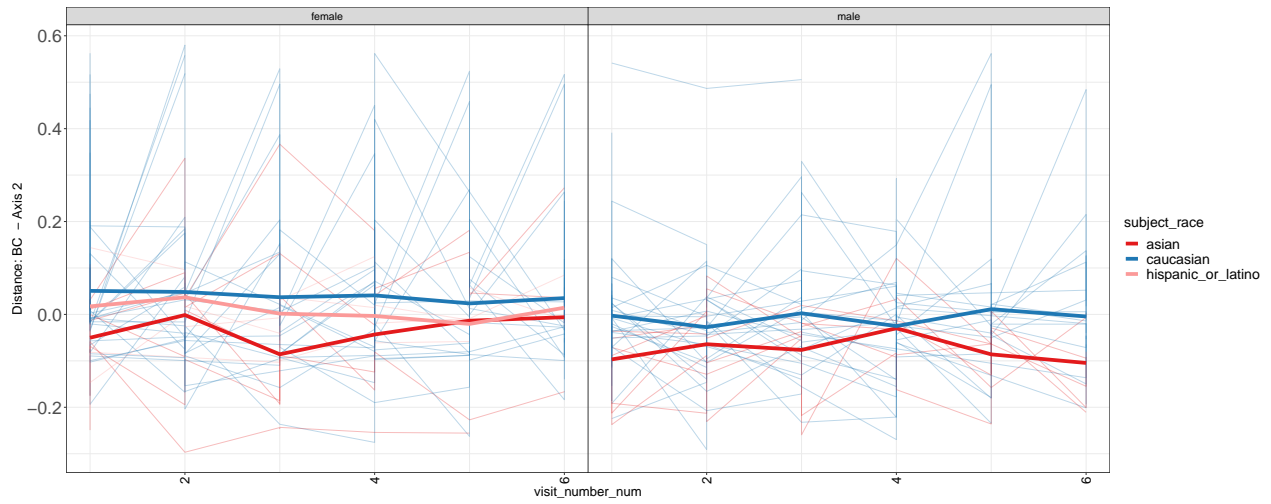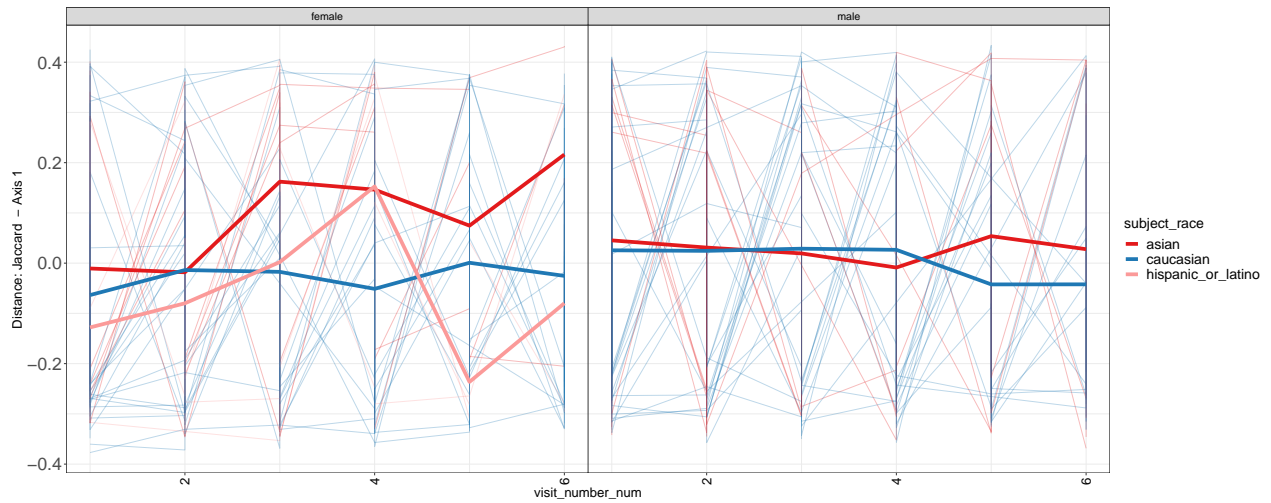
```
    pdf.hei = pdf.hei
)
```

In this visualization, the beta change represents the distance of each subject from their first/reference time point.

$BC
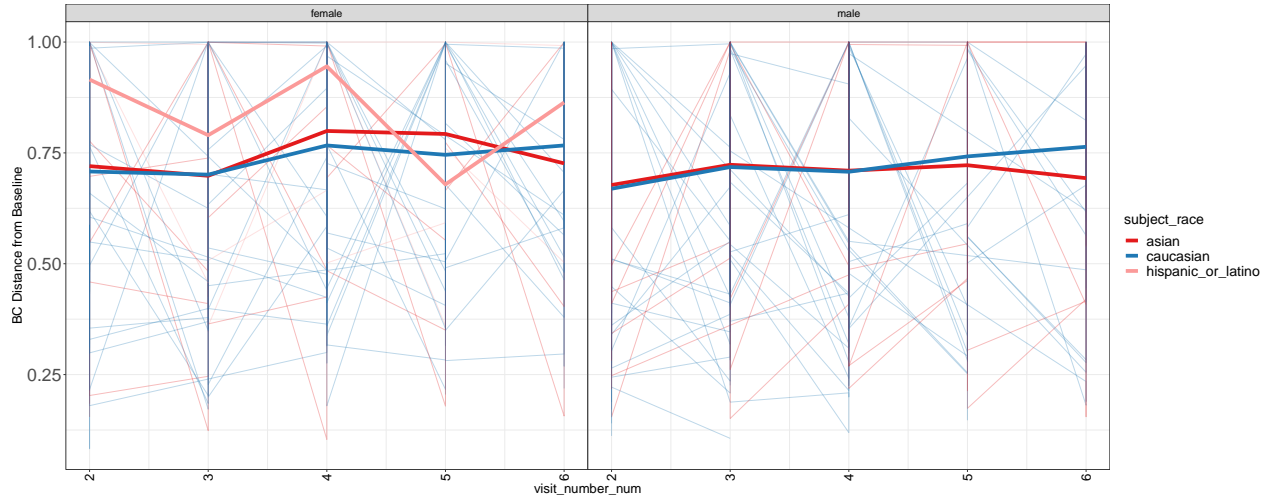
$Jaccard
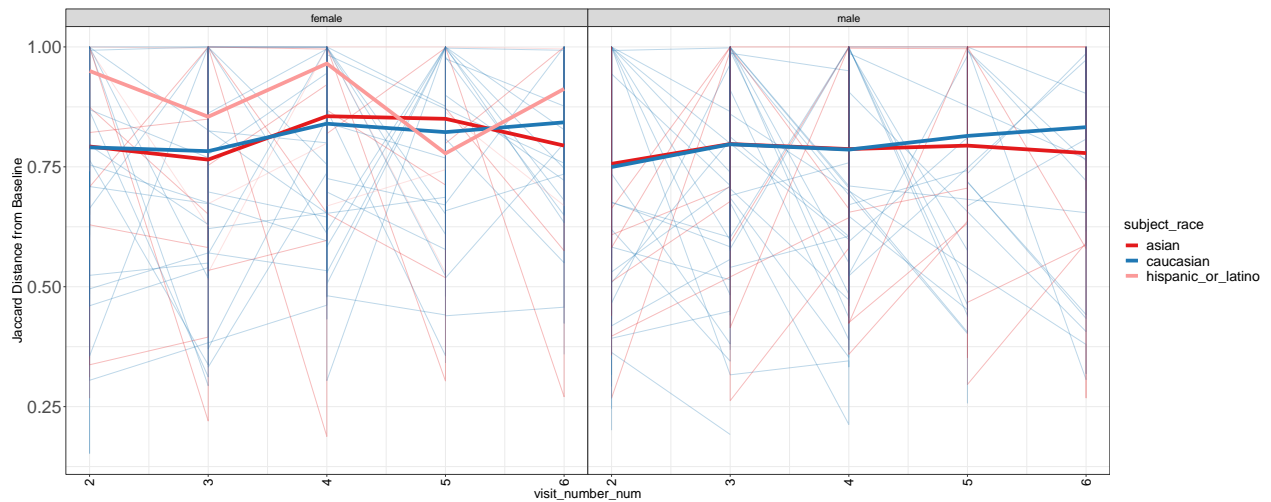
3.2 Distance-based trend test

```
beta_trend_test_longitudinal_results <- generate_beta_trend_test_long(
                                        data.obj = data.obj,
                                        dist.obj = dist.obj,
                                        subject.var = subject.var,
                                        time.var = time.var,
                                        group.var = group.var,
                                        adj.vars = test.adj.vars,
                                        dist.name = dist.name)
```

In this analysis, we utilized a linear mixed effects model with both a random intercept and a random slope to investigate a potential difference in trend. Specifically, we included visit_number_num, subject_race and

tested the interaction between the variables subject_race and visit_number_num.

3.2.1 Bray–Curtis distance

An ANOVA test of the null hypothesis of no difference among the 3 groups produces a p-value of 0.341.

Table 9: Table continues below

Term	Estimate	Std.Error
(Intercept)	0.6958	0.06534
subject_racecaucasian	-0.04149	0.07489
subject_racehispanic_or_latino	0.2256	0.1311
visit_number_num	0.007063	0.01618
subject_racecaucasian:visit_number_num	0.01135	0.01847
subject_racehispanic_or_latino:visit_number_num	-0.03062	0.03271
subject_race:visit_number_num	NA	NA

Statistic	P.Value
10.65	5.589e-24
-0.5541	0.5798
1.721	0.08617
0.4366	0.6625
0.6146	0.539
-0.9361	0.3495
1.078	0.3406

3.2.2 Jaccard distance

An ANOVA test of the null hypothesis of no difference among the 3 groups produces a p-value of 0.342.

Table 11: Table continues below

Term	Estimate	Std.Error
(Intercept)	0.7717	0.05137
subject_racecaucasian	-0.03381	0.05891
subject_racehispanic_or_latino	0.1697	0.1037
visit_number_num	0.005663	0.01251
subject_racecaucasian:visit_number_num	0.01065	0.01429
subject_racehispanic_or_latino:visit_number_num	-0.02057	0.02526
subject_race:visit_number_num	NA	NA

Statistic	P.Value
15.02	5.329e-37
-0.5739	0.5666
1.636	0.1039
0.4527	0.6509
0.7454	0.4562
-0.8144	0.4157
1.075	0.3418

3.3 Distance-based volatility test

```
beta_volatility_test_longitudinal_results <- generate_beta_volatility_test_long(
                                    data.obj = data.obj,
                                    dist.obj = dist.obj,
                                    subject.var = subject.var,
                                    time.var = time.var,
                                    group.var = group.var,
                                    adj.vars = test.adj.vars,
                                    dist.name = dist.name)
```

In this analysis, we employed a general linear model followed by ANOVA to test the effect of subject_race on beta diversity volatility. The beta diversity volatility is calculated by averaging the rate of change in beta diversity across consecutive time points. Specifically, for each pair of adjacent time points, we compute the difference in beta diversity, normalize it by the time difference, and then take the average over all such pairs.

3.3.1 Bray–Curtis distance

Based on the general linear model, the level caucasian of the variable subject_race did not significantly differ from level asian, with a p-value of 0.896. Based on the general linear model, the level hispanic_or_latino of the variable subject_race did not significantly differ from level asian, with a p-value of 0.272. An ANOVA test of the null hypothesis of no difference in volatility among 3 groups produces a p-value of 0.517.

Term	Estimate	Std.Error	Statistic	P.Value
(Intercept)	0.3254	0.06347	5.126	3.332e-06
subject_racecaucasian	-0.009375	0.07172	-0.1307	0.8964
subject_racehispanic_or_latino	-0.1626	0.1466	-1.109	0.2717
subject_race	NA	NA	0.6662	0.5174
Residuals	NA	NA	NA	NA

3.3.2 Jaccard distance

Based on the general linear model, the level caucasian of the variable subject_race did not significantly differ from level asian, with a p-value of 0.846. Based on the general linear model, the level hispanic_or_latino of the variable subject_race did not significantly differ from level asian, with a p-value of 0.240. An ANOVA test of the null hypothesis of no difference in volatility among 3 groups produces a p-value of 0.480.

Term	Estimate	Std.Error	Statistic	P.Value
(Intercept)	0.3765	0.06838	5.505	8.089e-07
subject_racecaucasian	-0.01502	0.07726	-0.1945	0.8465
subject_racehispanic_or_latino	-0.1873	0.1579	-1.186	0.2402
subject_race	NA	NA	0.7432	0.4799
Residuals	NA	NA	NA	NA

4. Feature-level Analysis

Rarefaction has been enabled for feature-level analysis.

Reason: The observed abundance of rare/low-abundance features can be strongly influenced by sequence depth. Rarefaction is an effective method to control the effect of sequence depth variation. By employing rarefaction, we can potentially increase the power of detecting those rare/low-abundance features, even though it introduces some variation due to under-sampling. In essence, this step helps to ensure more accurate and consistent results across samples with varying sequence depths.

If you do not wish to perform rarefaction during feature-level analysis, please turn feature.analysis.rarefy to FALSE.

4.1 Trend test

```
taxa_trend_test_results <- generate_taxa_trend_test_long(
                                          data.obj = data.obj,
                                          subject.var = subject.var,
                                          time.var = time.var,
                                          group.var = group.var,
                                          adj.vars = test.adj.vars,
                                          prev.filter = prev.filter,
                                          abund.filter = abund.filter,
                                          feature.level = test.feature.level,
                                          feature.dat.type = feature.dat.type,
                                          ...)
```

$Family $Family$caucasian vs asian (Reference)

$Family$hispanic_or_latino vs asian (Reference)
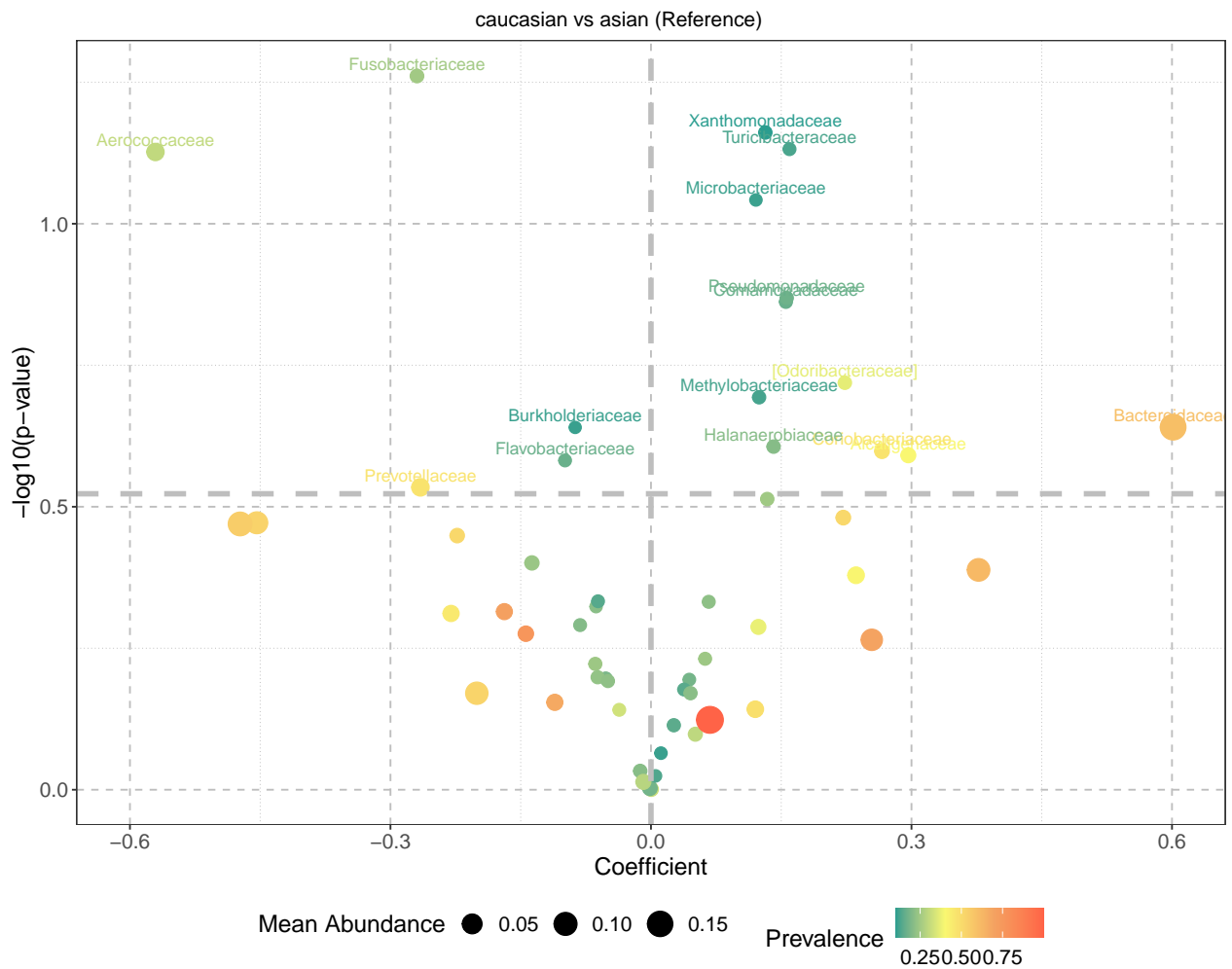
hispanic_or_latino vs asian (Reference)
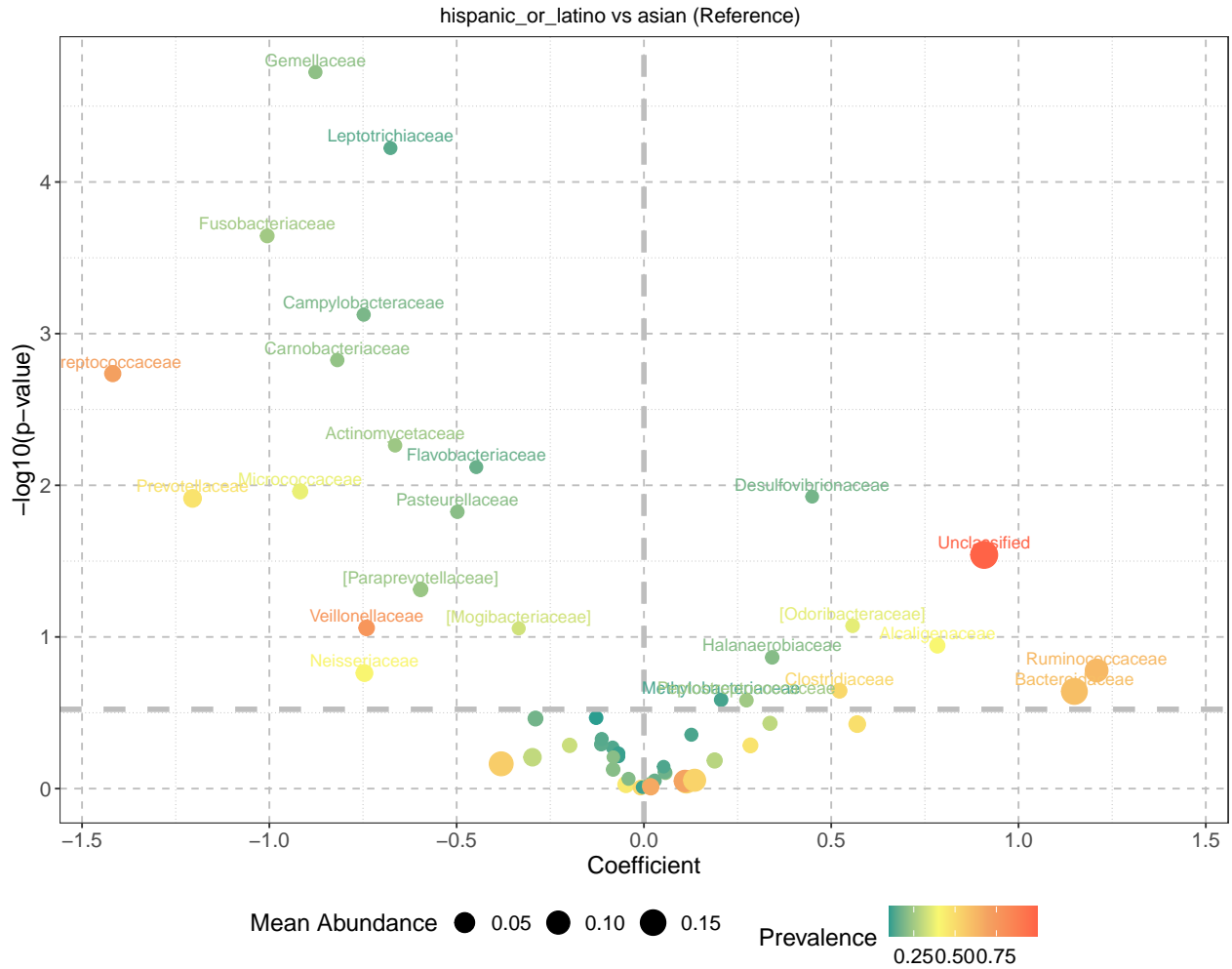
In this analysis, we utilized the LinDA linear mixed effects model to investigate potential differences in trend. Specifically, we tested the interaction between the variables subject_race and visit_number_num, for different taxa, while adjusting for other covariates.

For the taxon Family in comparison caucasian vs asian (Reference), significant results were identified using the none method for p-value adjustment, based on a threshold of 0.3.

Table 15: Table continues below

Variable	Coefficient	SE	P.Value	Adjusted.P.Value
[Odoribacteraceae]	0.2233	0.1705	0.1909	0.883
Aerococcaceae	-0.5707	0.3195	0.07468	0.883
Alcaligenaceae	0.2963	0.2609	0.2565	0.883
Bacteroidaceae	0.6013	0.4991	0.2288	0.883
Burkholderiaceae	-0.08727	0.07213	0.229	0.883
Comamonadaceae	0.1554	0.1043	0.1374	0.883
Coriobacteriaceae	0.266	0.2321	0.2522	0.883
Flavobacteriaceae	-0.09895	0.0881	0.2619	0.883
Fusobacteriaceae	-0.2695	0.1397	0.05482	0.883
Halanaerobiaceae	0.1412	0.1219	0.2476	0.883
Methylobacteriaceae	0.1245	0.09714	0.2026	0.883
Microbacteriaceae	0.1209	0.07123	0.09078	0.883
Prevotellaceae	-0.2655	0.2498	0.2922	0.883

Variable	Coefficient	SE	P.Value	Adjusted.P.Value
Pseudomonadaceae	0.1564	0.1045	0.1352	0.883
Turicibacteraceae	0.1595	0.0871	0.07382	0.883
Xanthomonadaceae	0.1317	0.07095	0.06898	0.883

Mean.Abundance	Prevalence
0.001367	0.367
0.02322	0.3043
0.006531	0.407
0.1566	0.6017
0.0001716	0.1148
0.0007825	0.1652
0.005378	0.4748
0.0004532	0.167
0.001891	0.2522
0.001353	0.2087
0.001122	0.127
0.0002028	0.1165
0.02444	0.4765
0.0009125	0.1739
0.000539	0.1304
0.00121	0.1043

For the taxon Family in comparison hispanic_or_latino vs asian (Reference), significant results were identified using the none method for p-value adjustment, based on a threshold of 0.3.

Table 17: Table continues below

Variable	Coefficient	SE	P.Value	Adjusted.P.Value
[Mogibacteriaceae]	-0.3342	0.1953	0.0878	0.2841
[Odoribacteraceae]	0.557	0.3222	0.08445	0.2841
[Paraprevotellaceae]	-0.5961	0.3016	0.04862	0.191
Actinomycetaceae	-0.6642	0.2286	0.005457	0.04287
Alcaligenaceae	0.7834	0.4947	0.1139	0.3479
Bacteroidaceae	1.149	0.9537	0.2287	0.547
Campylobacteraceae	-0.7481	0.2063	0.0007499	0.01031
Carnobacteriaceae	-0.8185	0.2438	0.001494	0.01644
Clostridiaceae	0.5221	0.4309	0.2263	0.547
Desulfovibrionaceae	0.449	0.1779	0.01188	0.06102
Flavobacteriaceae	-0.4475	0.167	0.007587	0.05216
Fusobacteriaceae	-1.006	0.2672	0.0002268	0.004157
Gemellaceae	-0.877	0.2019	1.886e-05	0.001038
Halanaerobiaceae	0.3424	0.2295	0.1363	0.3946
Leptotrichiaceae	-0.6767	0.1669	5.976e-05	0.001643
Methylobacteriaceae	0.2061	0.1819	0.2597	0.5753
Micrococcaceae	-0.9176	0.3594	0.01097	0.06102
Neisseriaceae	-0.7461	0.547	0.1731	0.4534
Pasteurellaceae	-0.498	0.2016	0.01495	0.06854
Peptostreptococcaceae	0.2736	0.2432	0.2615	0.5753
Prevotellaceae	-1.205	0.4647	0.0122	0.06102

Variable	Coefficient	SE	P.Value	Adjusted.P.Value
Ruminococcaceae	1.208	0.8722	0.1664	0.4534
Streptococcaceae	-1.418	0.4528	0.001834	0.01681
Veillonellaceae	-0.7403	0.4316	0.08719	0.2841
Unclassified	0.9082	0.4031	0.02875	0.1216

Mean.Abundance	Prevalence
0.0004627	0.3304
0.001367	0.367
0.00396	0.24
0.0008527	0.2452
0.006531	0.407
0.1566	0.6017
0.0006594	0.193
0.0008457	0.2296
0.005265	0.513
0.0003345	0.1843
0.0004532	0.167
0.001891	0.2522
0.0006681	0.2226
0.001353	0.2087
0.0004367	0.1478
0.001122	0.127
0.006502	0.3757
0.01687	0.4122
0.001104	0.2157
0.001255	0.2504
0.02444	0.4765
0.09302	0.6209
0.0128	0.7096
0.009015	0.7617
0.1839	0.9983

The trend test results for features have been saved in the current working directory. Each taxa rank and its corresponding comparison have their own file named with the prefix: taxa_trend_test_results_ followed by the taxon rank, the comparison, and the file extension .csv. Please refer to these files for more detailed data.

4.2 Volatility test

```
taxa_volatility_test_results <- generate_taxa_volatility_test_long(
                                    data.obj = data.obj,
                                    subject.var = subject.var,
                                    time.var = time.var,
                                    group.var = group.var,
                                    adj.vars = test.adj.vars,
                                    prev.filter = prev.filter,
                                    abund.filter = abund.filter,
                                    feature.level = test.feature.level,
                                    feature.dat.type = feature.dat.type
                                    )
```

$Family $Family$caucasian vs asian (Reference)

caucasian vs asian (Reference)

$Family$hispanic_or_latino vs asian (Reference)
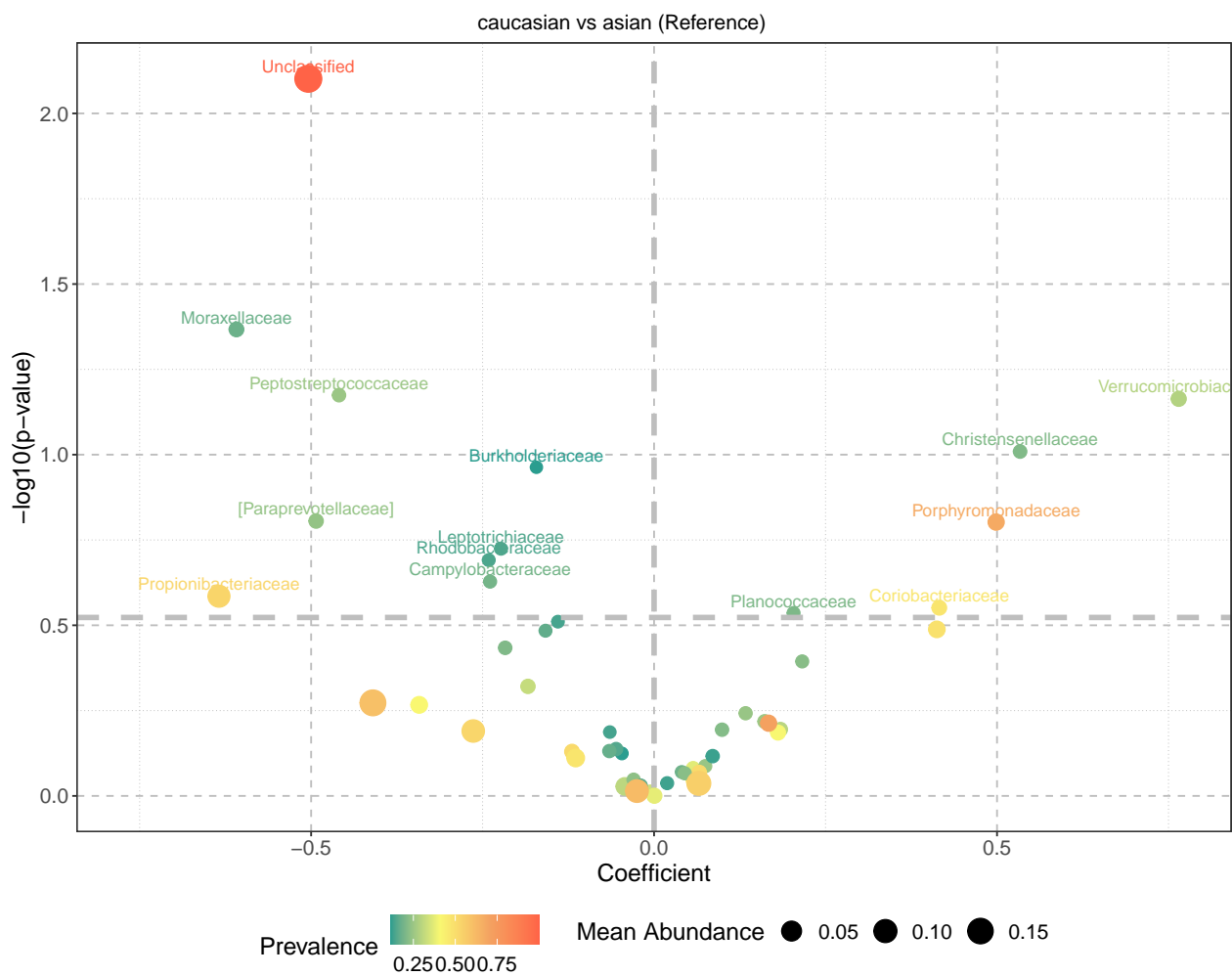
hispanic_or_latino vs asian (Reference)
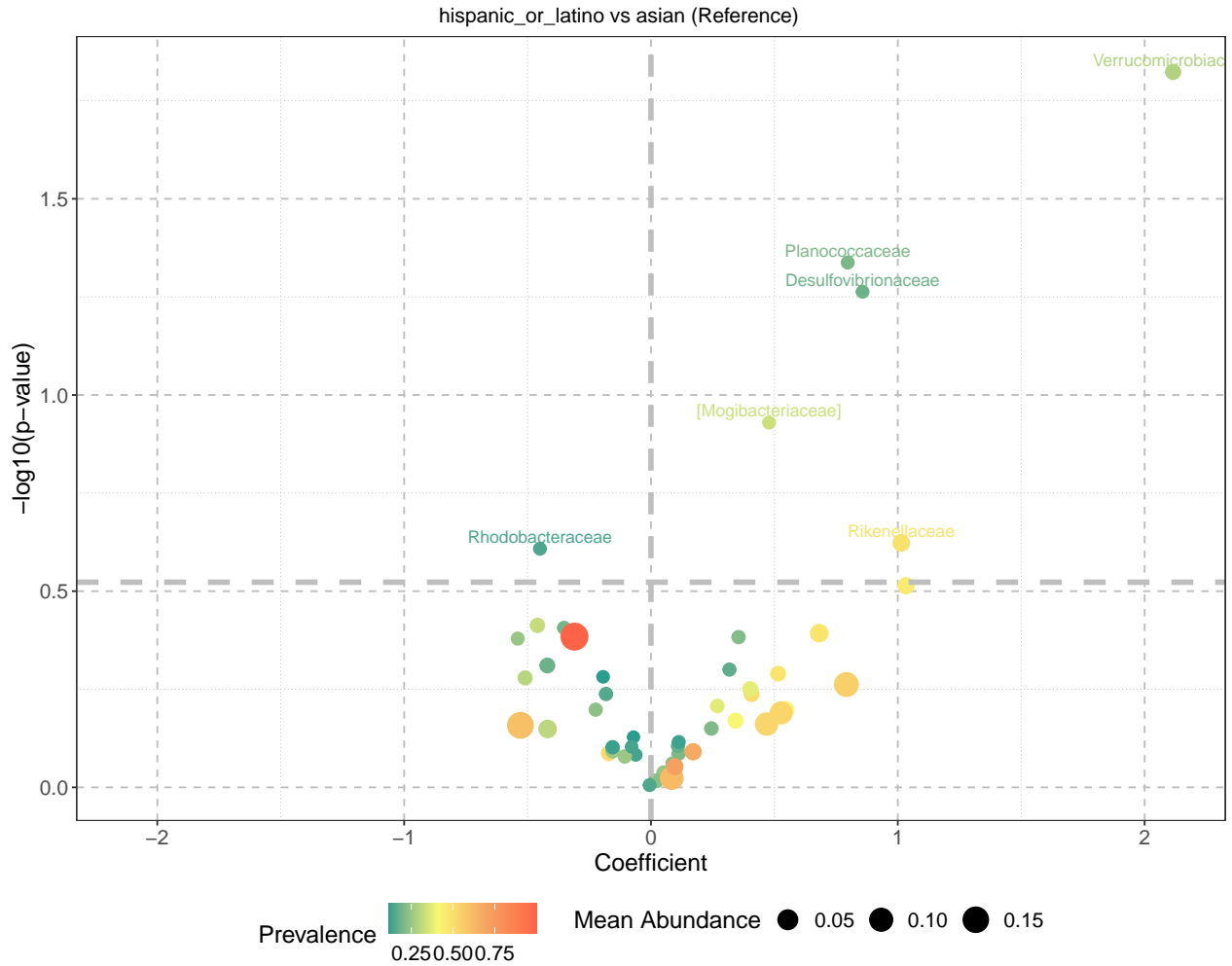
In this analysis, a general linear model followed by ANOVA was employed to test the effect of subject_race on the volatility of various taxa abundances.Feature abundances were transformed using the centered log-ratio (CLR) transformation. For count data, 0.5 was added to all counts before performing the CLR. For proportion data, zeros were replaced by half the minimum non-zero proportion for each taxon.

For the taxon Family in comparison caucasian vs asian (Reference), significant results were identified using the none method for p-value adjustment, based on a threshold of 0.3.

Table 19: Table continues below

Variable	Coefficient	SE	P.Value	Adjusted.P.Value
[Paraprevotellaceae]	-0.4927	0.3434	0.1565	0.9648
Burkholderiaceae	-0.1715	0.1054	0.1089	0.9075
Campylobacteraceae	-0.2391	0.1996	0.2354	0.9648
Christensenellaceae	0.5336	0.3175	0.09791	0.9075
Coriobacteriaceae	0.4158	0.3822	0.2809	0.9648
Leptotrichiaceae	-0.2232	0.1679	0.1886	0.9648
Moraxellaceae	-0.6089	0.2945	0.04294	0.8574
Peptostreptococcaceae	-0.4593	0.2463	0.06695	0.8574
Planococcaceae	0.2033	0.191	0.2912	0.9648
Porphyromonadaceae	0.4987	0.3486	0.1576	0.9648
Propionibacteriaceae	-0.6344	0.5579	0.2599	0.9648
Rhodobacteraceae	-0.241	0.1876	0.2037	0.9648

Variable	Coefficient	SE	P.Value	Adjusted.P.Value
Verrucomicrobiaceae	0.7649	0.4126	0.06859	0.8574
Unclassified	-0.5041	0.1836	0.007914	0.3957

Mean.Abundance	Prevalence
0.00396	0.24
0.0001716	0.1148
0.0006594	0.193
0.001458	0.2104
0.005378	0.4748
0.0004367	0.1478
0.00552	0.1809
0.001255	0.2504
0.0004896	0.2052
0.01341	0.6765
0.08086	0.5322
0.0004272	0.1443
0.006963	0.2887
0.1839	0.9983

For the taxon Family in comparison hispanic_or_latino vs asian (Reference), significant results were identified using the none method for p-value adjustment, based on a threshold of 0.3.

Table 21: Table continues below

Variable	Coefficient	SE	P.Value	Adjusted.P.Value
[Mogibacteriaceae]	0.4783	0.3014	0.1176	0.9839
Desulfovibrionaceae	0.8573	0.4374	0.05456	0.9094
Planococcaceae	0.797	0.3912	0.04597	0.9094
Rhodobacteraceae	-0.4497	0.3843	0.2465	0.9839
Rikenellaceae	1.014	0.8512	0.2382	0.9839
Verrucomicrobiaceae	2.115	0.8452	0.01502	0.7508

Mean.Abundance	Prevalence
0.0004627	0.3304
0.0003345	0.1843
0.0004896	0.2052
0.0004272	0.1443
0.01598	0.4887
0.006963	0.2887

For the taxon Family in comparison subject_race, significant results were identified using the none method for p-value adjustment, based on a threshold of 0.3.

Table 23: Table continues below

Variable	Coefficient	SE	P.Value	Adjusted.P.Value
[Mogibacteriaceae]	NA	NA	0.2302	0.9952
[Paraprevotellaceae]	NA	NA	0.2846	0.9952
Burkholderiaceae	NA	NA	0.2605	0.9952
Christensenellaceae	NA	NA	0.2414	0.9952
Desulfovibrionaceae	NA	NA	0.1303	0.9952
Moraxellaceae	NA	NA	0.1261	0.9952
Peptostreptococcaceae	NA	NA	0.1598	0.9952
Planococcaceae	NA	NA	0.1286	0.9952
Verrucomicrobiaceae	NA	NA	0.0333	0.8325
Unclassified	NA	NA	0.0278	0.8325

Mean.Abundance	Prevalence
0.0004627	0.3304
0.00396	0.24
0.0001716	0.1148
0.001458	0.2104
0.0003345	0.1843
0.00552	0.1809
0.001255	0.2504
0.0004896	0.2052
0.006963	0.2887
0.1839	0.9983

The volatility test results for individual feature have been saved in the current working directory. Each taxa rank and its corresponding comparison have their own file named with the prefix: taxa_volatility_test_results_ followed by the taxon rank, the comparison, and the file extension .csv. Please refer to these files for more detailed data.

4.3 Data visualization(significant features)

4.3.1 Significant features boxplot

```
if (length(combined_significant_taxa) != 0){

taxa_indiv_boxplot_results <- generate_taxa_indiv_boxplot_long(
                        data.obj = data.obj,
                        subject.var = subject.var,
                        time.var = time.var,
                        t0.level = t0.level,
                        ts.levels = ts.levels,
                        group.var = group.var,
                        strata.var = strata.var,
                        feature.level = test.feature.level,
                        features.plot = combined_significant_taxa,
                        transform = feature.box.axis.transform,
                        feature.dat.type = feature.dat.type,
                        top.k.plot = top.k.plot,
                        top.k.func = top.k.func,
```

```
                                        prev.filter = prev.filter,
                                        abund.filter = abund.filter,
                                        base.size = 10,
                                        theme.choice = theme.choice,
                                        custom.theme = custom.theme,
                                        palette = palette,
                                        pdf = TRUE,
                                        file.ann = file.ann,
                                        pdf.wid = pdf.wid,
                                        pdf.hei = pdf.hei)
}
```

The boxplot results for individual taxa or features can be found in the current working directory. The relevant file is named: taxa_indiv_boxplot_long_subject_subject_id_time_visit_number_num_feature_level_Family_transform_sqrt_prev_filter_0.1_abund_filter_1e-04_group_subject_race_strata_subject_gender. Please refer to this file for more detailed visualizations.

4.3.2 Significant features spaghettiplot

```
if (length(combined_significant_taxa) != 0){
  taxa_spaghettiplot_results <- generate_taxa_spaghettiplot_long(
                                        data.obj = data.obj,
                                        subject.var = subject.var,
                                        time.var = time.var,
                                        group.var = group.var,
                                        strata.var = strata.var,
                                        t0.level = t0.level,
                                        ts.levels = ts.levels,
                                        feature.level = test.feature.level,
                                        feature.dat.type = feature.dat.type,
                                        features.plot = combined_significant_taxa,
                                        top.k.plot = top.k.plot,
                                        top.k.func = top.k.func,
                                        prev.filter = prev.filter,
                                        abund.filter = abund.filter,
                                        base.size = 10,
                                        theme.choice = theme.choice,
                                        custom.theme = custom.theme,
                                        palette = palette,
                                        pdf = pdf,
                                        file.ann = file.ann,
                                        pdf.wid = pdf.wid,
                                        pdf.hei = pdf.hei)

taxa_indiv_spaghettiplot_results <- generate_taxa_indiv_spaghettiplot_long(
                                        data.obj = data.obj,
                                        subject.var = subject.var,
                                        time.var = time.var,
                                        t0.level = t0.level,
                                        ts.levels = ts.levels,
                                        group.var = group.var,
                                        strata.var = strata.var,
                                        change.base = change.base,
                                        feature.change.func = feature.change.func,
```

```
                            feature.level = test.feature.level,
                            features.plot = combined_significant_taxa,
                            feature.dat.type = feature.dat.type,
                            top.k.plot = top.k.plot,
                            top.k.func = top.k.func,
                            prev.filter = prev.filter,
                            abund.filter = abund.filter,
                            base.size = 10,
                            theme.choice = theme.choice,
                            custom.theme = custom.theme,
                            palette = palette,
                            pdf = TRUE,
                            file.ann = file.ann,
                            pdf.wid = pdf.wid,
                            pdf.hei = pdf.hei)
}
```
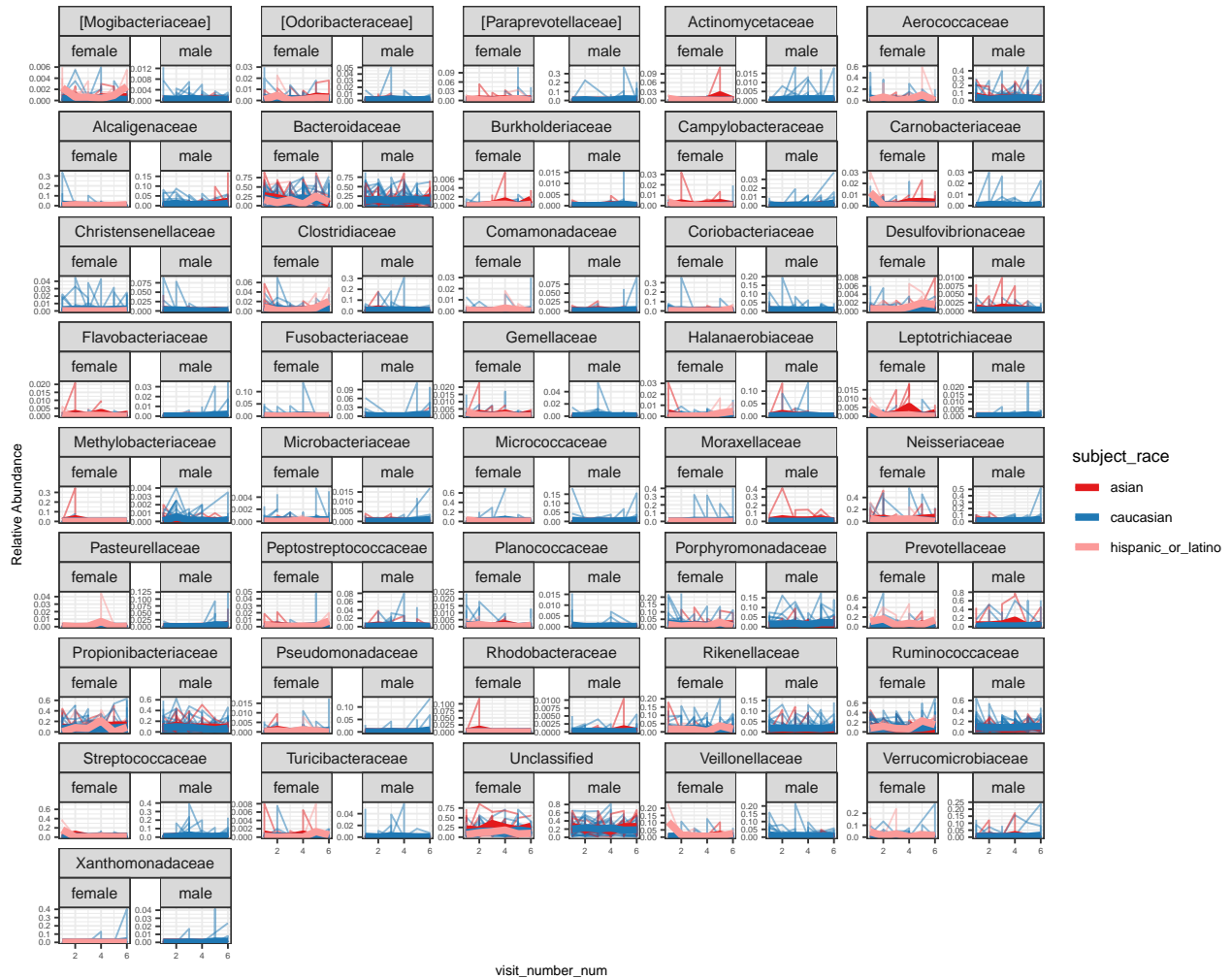
$Family

The spaghettiplot results for individual taxa or features can be found in the current working directory.
The relevant file is named: taxa_indiv_spaghettiplot_long_subject_subject_id_time_visit_num-

ber_num_group_subject_race_strata_subject_gender_feature_level_Family_prev_filter_0.1_abund_filter_1e-04_base_size_20_theme_choice_bw_pdf_wid_11_pdf_hei_8.5.pdf. Please refer to this file for more detailed visualizations.