

Feature-Level Phylogenetic Assessment

Nate Olson

2017-11-17

0.1 relative abundanceDADA

```
dada_signal <- signal_pipe(rel_abu_feature_metrics, dada_tree)

## Adding missing grouping variables: `pipe`
## Joining, by = "feature_id"
## Warning: Column `feature_id` joining factor and character vector, coercing
## into character vector
## Found more than one class "phylo" in cache; using the first, from namespace 'treeio'
## Also defined by 'RNeXML'
## Found more than one class "phylo" in cache; using the first, from namespace 'treeio'
## Also defined by 'RNeXML'
## Joining, by = "feature_id"
## Joining, by = "feature_id"
## Joining, by = "feature_id"
## Warning: Column `feature_id` joining character vector and factor, coercing
## into character vector

# dada_sim <- phyloSim(mothur_signal$p4d, method = c("Cmean"))
# dada_sim %>% plot()

dada_signal$signal %>%
  knitr::kable()



| Metric      | Method | stat      | pvalue    |
|-------------|--------|-----------|-----------|
| sc_bias     | Cmean  | 0.2917834 | 0.0010000 |
| sc_bias     | I      | 0.1515487 | 0.0010000 |
| sc_bias     | K      | 0.0213141 | 0.1040000 |
| sc_bias     | K.star | 0.0245915 | 0.0710000 |
| sc_bias     | Lambda | 0.6644439 | 0.0010000 |
| sc_variance | Cmean  | 0.1238726 | 0.0050000 |
| sc_variance | I      | 0.0509330 | 0.0020000 |
| sc_variance | K      | 0.0244555 | 0.0280000 |
| sc_variance | K.star | 0.0284928 | 0.0010000 |
| sc_variance | Lambda | 0.1309146 | 0.4906464 |



# dada_bias_corr <- phyloCorrelogram(dada_signal$p4d, trait = "sc_bias")
# plot(dada_bias_corr)

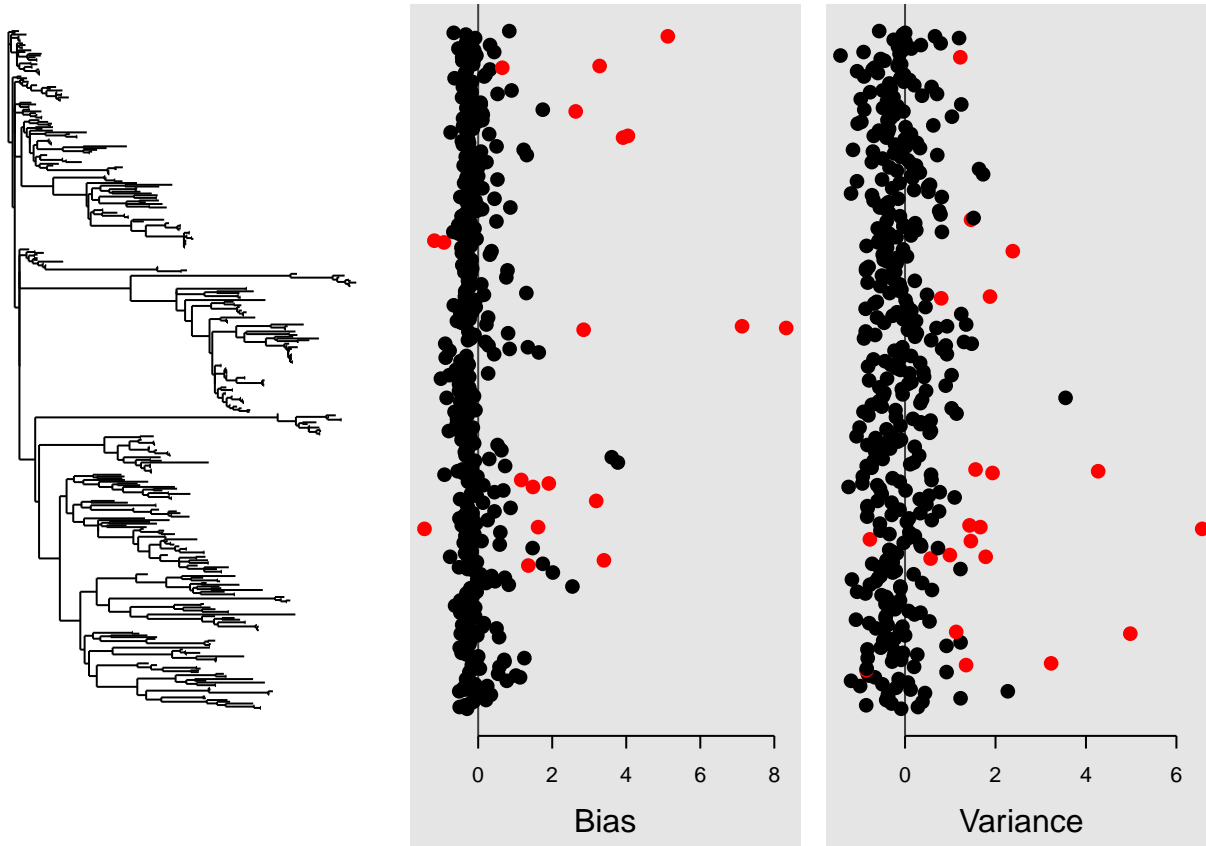
# dada_variance_corr <- phyloCorrelogram(dada_signal$p4d, trait = "sc_variance")
# plot(dada_variance_corr)

dotplot(dada_signal$p4d, dot.col = (dada_signal$lipa$p.value < 0.05) + 1,
        center = FALSE, scale = FALSE,
```

```

grid.horizontal = FALSE,
trait.labels = c("Bias", "Variance"),
show.tip = FALSE,
tree.ladderize = TRUE)

```



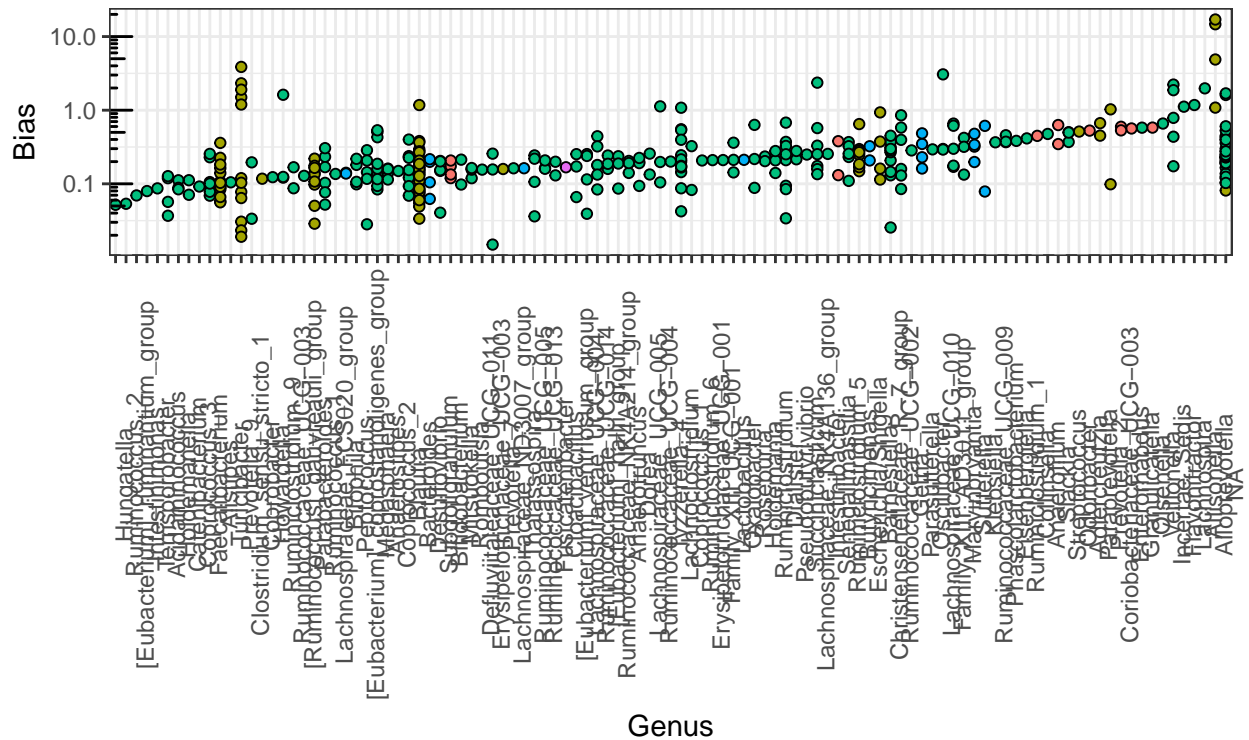
```

dada_signal$df %>% left_join(dada_tax) %>%
  mutate(Rank6 = fct_reorder(Rank6, Bias)) %>%
  select(Rank2, Rank6, Bias, Variance) %>%
  # gather(Metric, Value, -Rank2, -Rank6) %>%
  ggplot() + geom_point(aes(x = Rank6, y = Bias, fill = Rank2), shape = 21) +
  # facet_wrap(~Metric, scales = "free_y", ncol = 1) +
  geom_hline(aes(yintercept = 0), linetype = 2) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90), legend.position = "bottom") +
  labs(x = "Genus", fill = "Phylum") + scale_y_log10() + annotation_logticks(sides = "l")

```

```
## Joining, by = "feature_id"
```

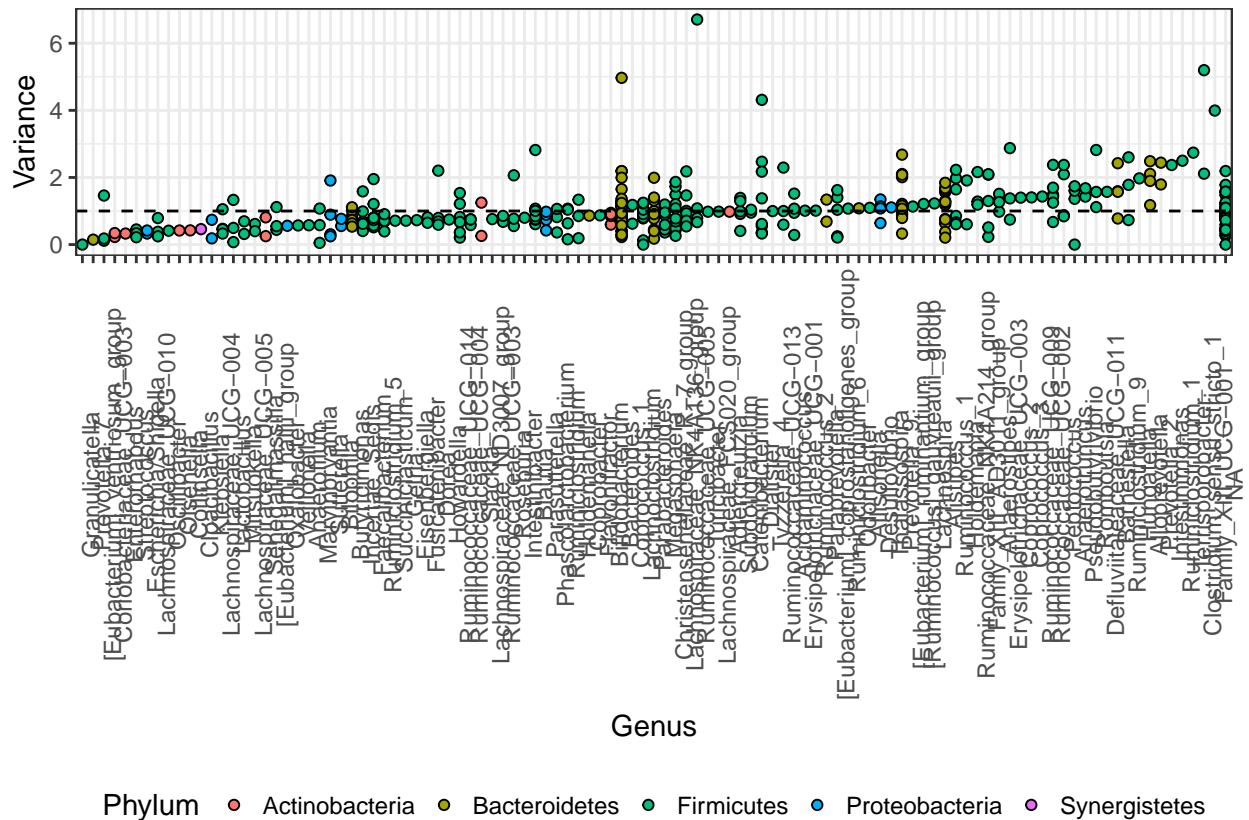
```
## Warning: Transformation introduced infinite values in continuous y-axis
```



Phylum ● Actinobacteria ● Bacteroidetes ● Firmicutes ● Proteobacteria ● Synergistetes

```
dada_signal$df %>% left_join(dada_tax) %>%
  mutate(Rank6 = fct_reorder(Rank6, Variance)) %>%
  select(Rank2, Rank6, Bias, Variance) %>%
  # gather(Metric, Value, -Rank2, -Rank6) %>%
  ggplot() + geom_point(aes(x = Rank6, y = Variance, fill = Rank2), shape = 21) +
  # facet_wrap(~Metric, scales = "free_y", ncol = 1) +
  geom_hline(aes(yintercept = 1), linetype = 2) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90), legend.position = "bottom") +
  labs(x = "Genus", fill = "Phylum")
```

```
## Joining, by = "feature_id"
```



0.2 relative abundance Mothur

```
mothur_signal <- signal_pipe(rel_abu_feature_metrics, mothur_tree)

## Adding missing grouping variables: `pipe`
## Joining, by = "feature_id"
## Warning: Column `feature_id` joining factor and character vector, coercing
## into character vector
## Found more than one class "phylo" in cache; using the first, from namespace 'treeio'
## Also defined by 'RNeXML'
## Found more than one class "phylo" in cache; using the first, from namespace 'treeio'
## Also defined by 'RNeXML'
## Joining, by = "feature_id"
## Joining, by = "feature_id"
## Joining, by = "feature_id"
## Warning: Column `feature_id` joining character vector and factor, coercing
## into character vector

# mothur_sim <- phyloSim(mothur_signal$p4d, method = c("Cmean"))
# mothur_sim %>% plot()

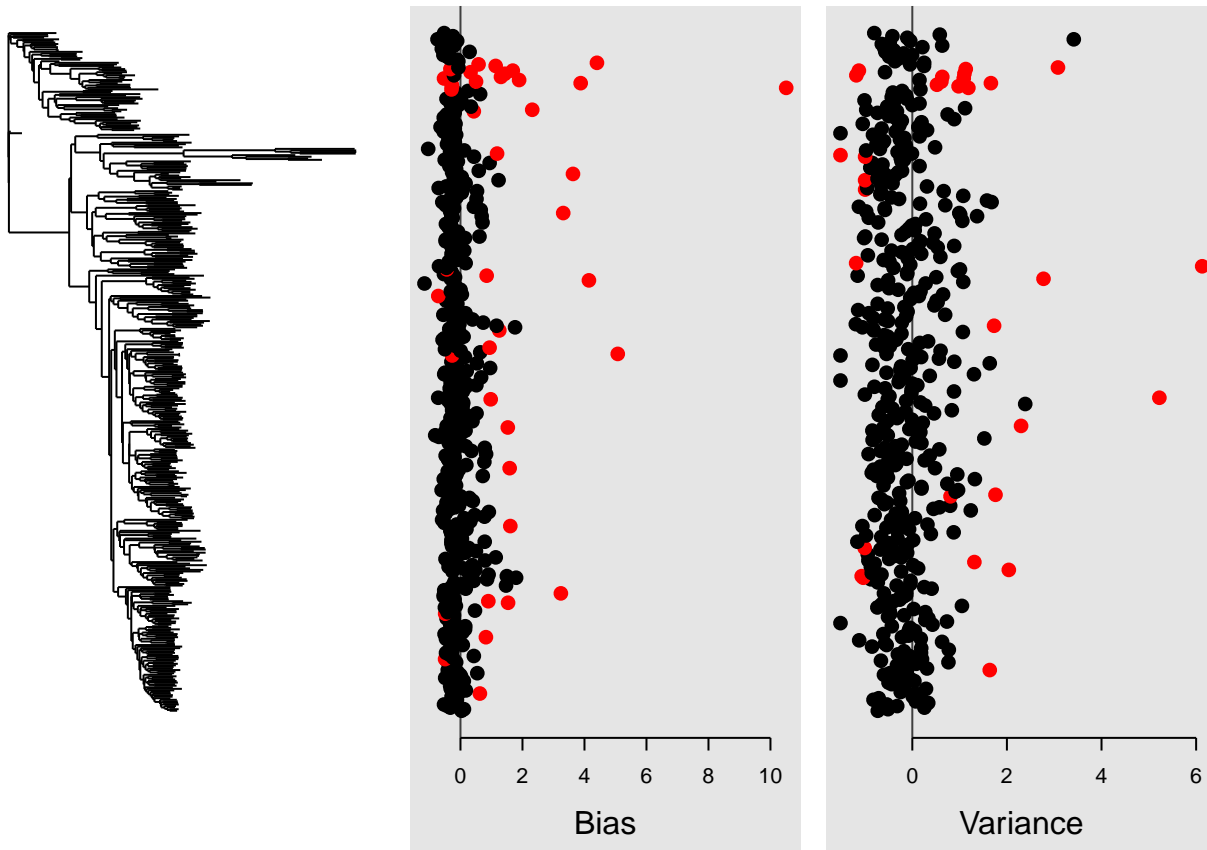
mothur_signal$signal %>%
  knitr::kable()
```

Metric	Method	stat	pvalue
sc_bias	Cmean	0.0618102	0.035
sc_bias	I	0.0012909	0.018
sc_bias	K	0.1650671	0.894
sc_bias	K.star	0.2000974	0.952
sc_bias	Lambda	0.6240029	0.001
sc_variance	Cmean	0.0177774	0.260
sc_variance	I	0.0007990	0.018
sc_variance	K	0.2563364	0.294
sc_variance	K.star	0.3658755	0.040
sc_variance	Lambda	0.6682174	0.001

```
# mothur_bias_corr <- phyloCorrelogram(mothur_signal$p4d, trait = "sc_bias")
# plot(mothur_bias_corr)
```

```
# mothur_variance_corr <- phyloCorrelogram(mothur_signal$p4d, trait = "sc_variance")
# plot(mothur_variance_corr)
```

```
dotplot(mothur_signal$p4d, dot.col = (mothur_signal$lipa$p.value < 0.05) + 1,
        center = FALSE, scale = FALSE,
        grid.horizontal = FALSE,
        trait.labels = c("Bias", "Variance"),
        show.tip = FALSE,
        tree.ladderize = TRUE)
```

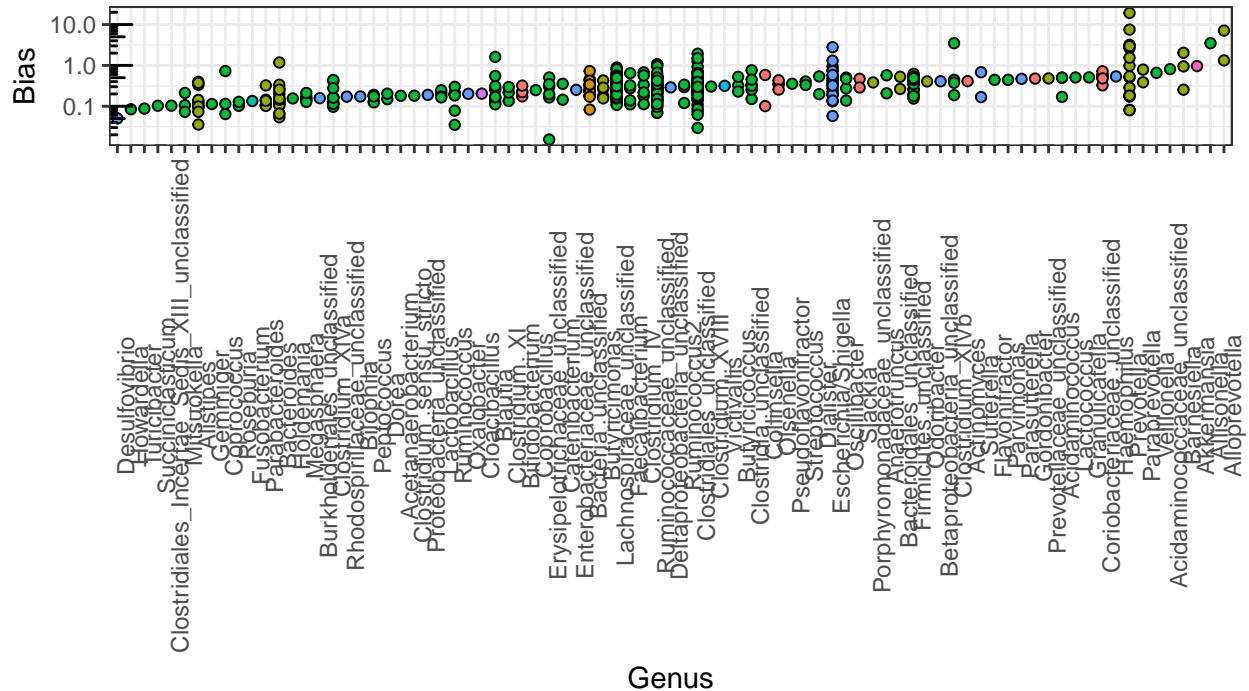


```
mothur_signal$df %>% left_join(mothur_tax) %>%
  mutate(Rank6 = fct_reorder(Rank6, Bias)) %>%
  select(Rank2, Rank6, Bias, Variance) %>%
```

```
# gather(Metric, Value, -Rank2, -Rank6) %>%
ggplot() + geom_point(aes(x = Rank6, y = Bias, fill = Rank2), shape = 21) +
# facet_wrap(~Metric, scales = "free_y", ncol = 1) +
geom_hline(aes(yintercept = 0), linetype = 2) +
theme_bw() +
theme(axis.text.x = element_text(angle = 90), legend.position = "bottom") +
labs(x = "Genus", fill = "Phylum") + scale_y_log10() + annotation_logticks(sides = "l")
```

```
## Joining, by = "feature_id"
```

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

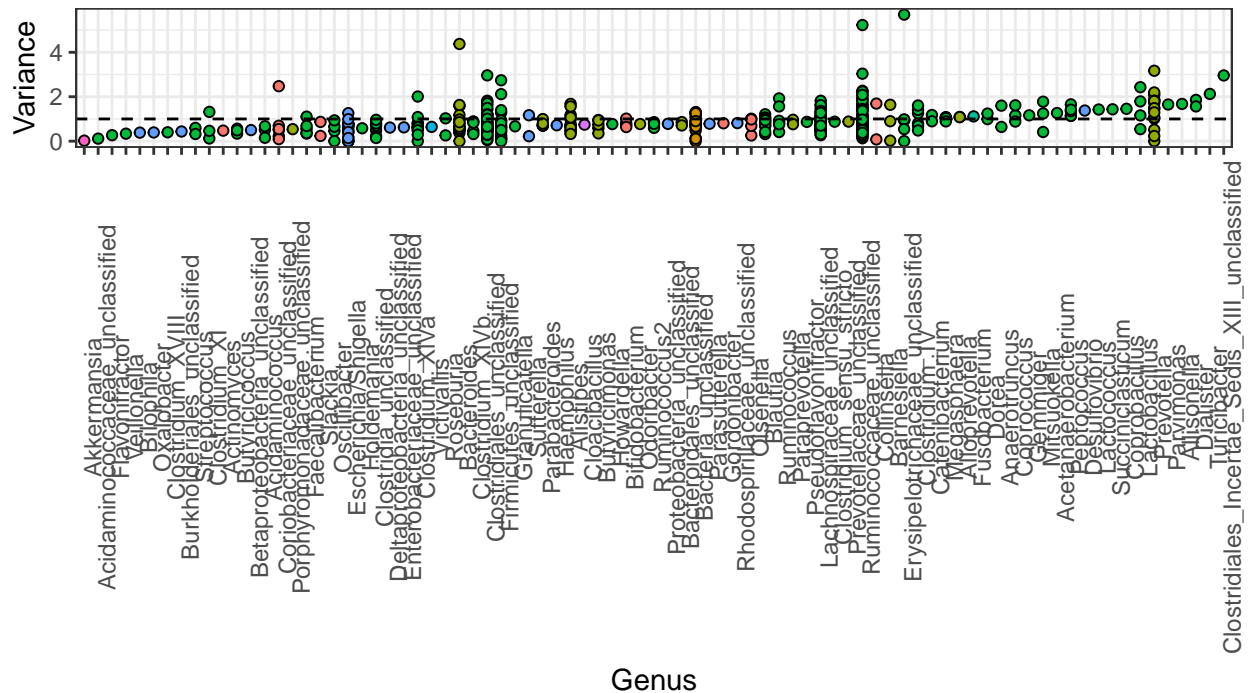


Phylum

- Actinobacteria ● Bacteroidetes ● Fusobacteria ● Proteobacteria ● Verrucomicrobia
- Bacteria_unclassified ● Firmicutes ● Lentisphaerae ● Synergistetes

```
mothur_signal$df %>% left_join(mothur_tax) %>%
mutate(Rank6 = fct_reorder(Rank6, Variance)) %>%
select(Rank2, Rank6, Bias, Variance) %>%
# gather(Metric, Value, -Rank2, -Rank6) %>%
ggplot() + geom_point(aes(x = Rank6, y = Variance, fill = Rank2), shape = 21) +
# facet_wrap(~Metric, scales = "free_y", ncol = 1) +
geom_hline(aes(yintercept = 1), linetype = 2) +
theme_bw() +
theme(axis.text.x = element_text(angle = 90), legend.position = "bottom") +
labs(x = "Genus", fill = "Phylum")
```

```
## Joining, by = "feature_id"
```



0.3 log fold-change Mothur

```

mothur_signal <- signal_pipe(logFC_feature_metrics, mothur_tree)

## Joining, by = "feature_id"
## Warning: Column `feature_id` joining factor and character vector, coercing
## into character vector
## Found more than one class "phylo" in cache; using the first, from namespace 'treeio'
## Also defined by 'RNeXML'
## Found more than one class "phylo" in cache; using the first, from namespace 'treeio'
## Also defined by 'RNeXML'
## Joining, by = "feature_id"
## Joining, by = "feature_id"
## Joining, by = "feature_id"
## Warning: Column `feature_id` joining character vector and factor, coercing
## into character vector

# mothur_sim <- phyloSim(mothur_signal$p4d, method = c("Cmean"))
# mothur_sim %>% plot()

mothur_signal$signal %>%
  knitr::kable()

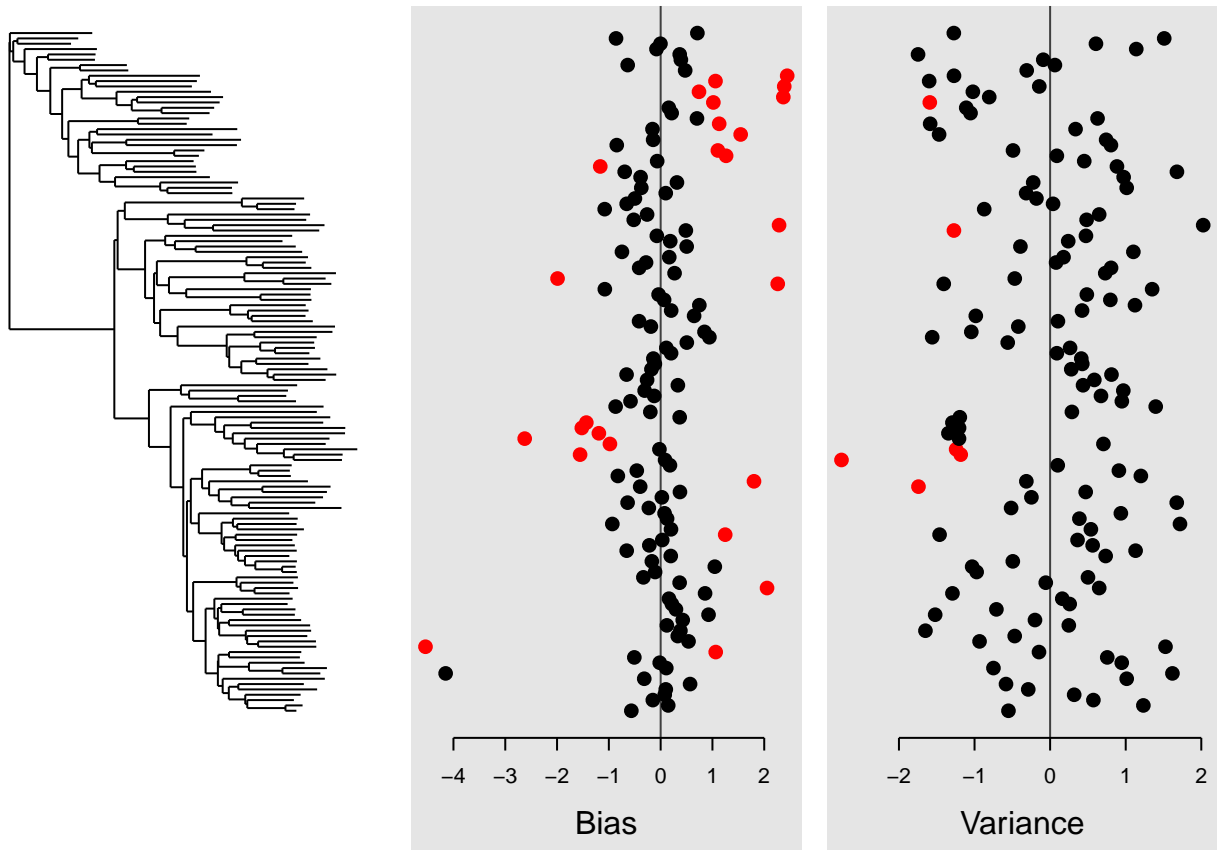
```

Metric	Method	stat	pvalue
sc_bias	Cmean	0.1340650	0.0170000
sc_bias	I	0.0039903	0.0110000
sc_bias	K	0.3465087	0.3920000
sc_bias	K.star	0.4747649	0.1850000
sc_bias	Lambda	0.3454275	0.0122141
sc_variance	Cmean	0.1758399	0.0030000
sc_variance	I	-0.0013723	0.0660000
sc_variance	K	0.3780044	0.1710000
sc_variance	K.star	0.5049729	0.0140000
sc_variance	Lambda	0.7189168	0.0010000

```
# mothur_bias_corr <- phyloCorrelogram(mothur_signal$p4d, trait = "sc_bias")
# plot(mothur_bias_corr)
```

```
# mothur_variance_corr <- phyloCorrelogram(mothur_signal$p4d, trait = "sc_variance")
# plot(mothur_variance_corr)
```

```
dotplot(mothur_signal$p4d, dot.col = (mothur_signal$lipa$p.value < 0.05) + 1,
        center = FALSE, scale = FALSE,
        grid.horizontal = FALSE,
        trait.labels = c("Bias", "Variance"),
        show.tip = FALSE,
        tree.ladderize = TRUE)
```

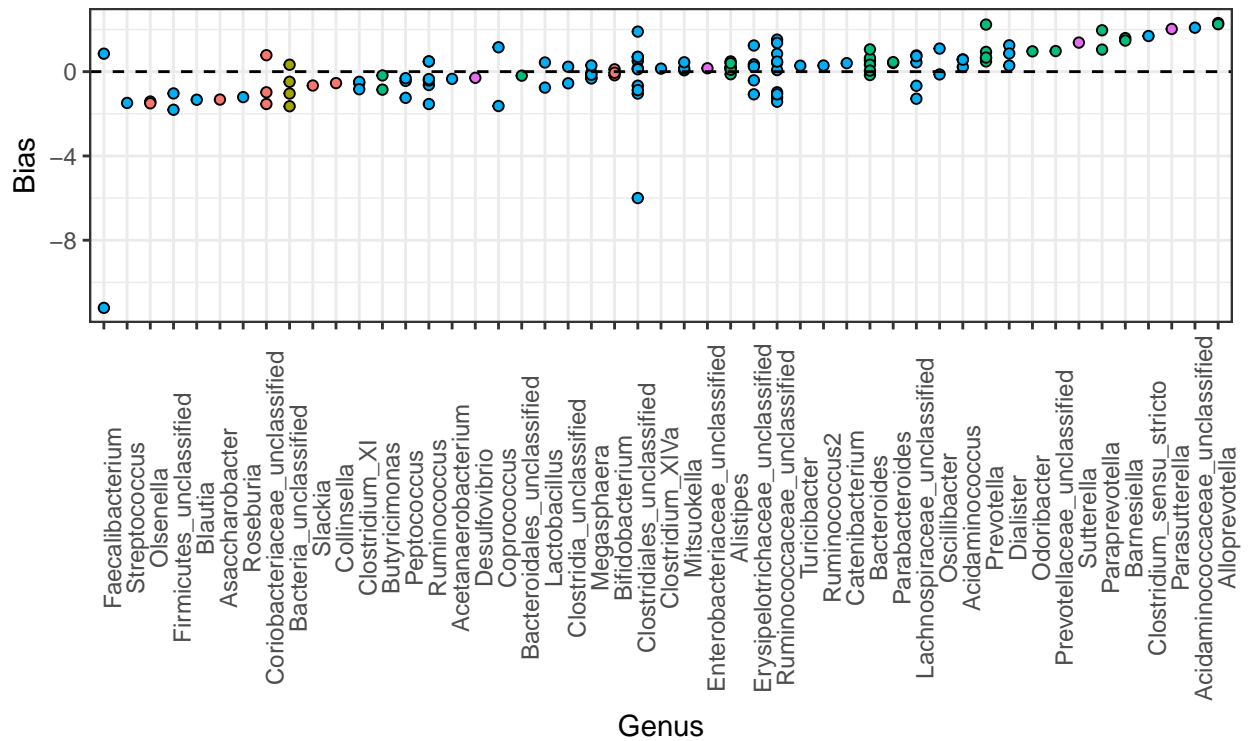


```
mothur_signal$df %>% left_join(mothur_tax) %>%
  mutate(Rank6 = fct_reorder(Rank6, Bias)) %>%
  select(Rank2, Rank6, Bias, Variance) %>%
```



```
# gather(Metric, Value, -Rank2, -Rank6) %>%
ggplot() + geom_point(aes(x = Rank6, y = Bias, fill = Rank2), shape = 21) +
# facet_wrap(~Metric, scales = "free_y", ncol = 1) +
geom_hline(aes(yintercept = 0), linetype = 2) +
theme_bw() +
theme(axis.text.x = element_text(angle = 90), legend.position = "bottom") +
labs(x = "Genus", fill = "Phylum")
```

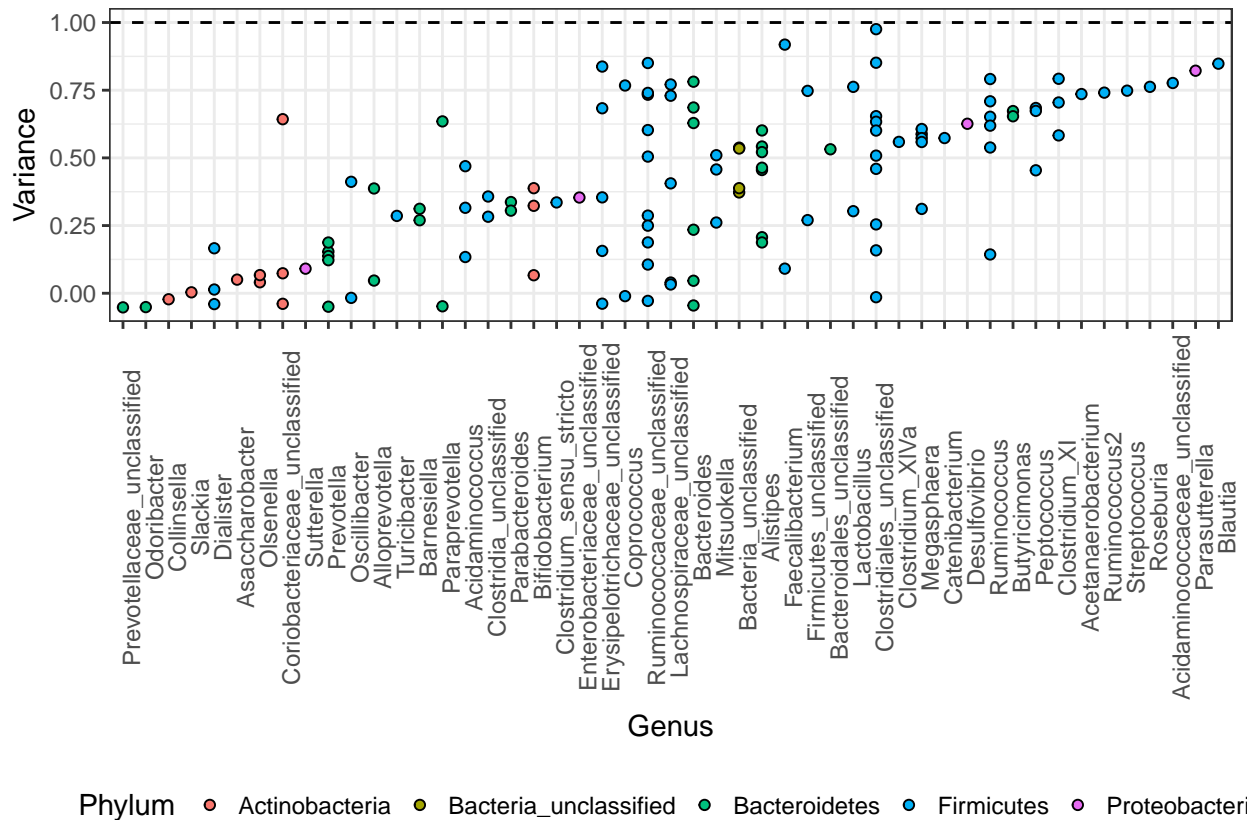
```
## Joining, by = "feature_id"
```



Phylum ● Actinobacteria ● Bacteria_unclassified ● Bacteroidetes ● Firmicutes ● Proteobacteria

```
mothur_signal$df %>% left_join(mothur_tax) %>%
mutate(Rank6 = fct_reorder(Rank6, Variance)) %>%
select(Rank2, Rank6, Bias, Variance) %>%
# gather(Metric, Value, -Rank2, -Rank6) %>%
ggplot() + geom_point(aes(x = Rank6, y = Variance, fill = Rank2), shape = 21) +
# facet_wrap(~Metric, scales = "free_y", ncol = 1) +
geom_hline(aes(yintercept = 1), linetype = 2) +
theme_bw() +
theme(axis.text.x = element_text(angle = 90), legend.position = "bottom") +
labs(x = "Genus", fill = "Phylum")
```

```
## Joining, by = "feature_id"
```



0.4 log fold-change DADA

```
dada_signal <- signal_pipe(logFC_feature_metrics, dada_tree)
```

```
## Joining, by = "feature_id"
```

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

```
## Found more than one class "phylo" in cache; using the first, from namespace 'treeio'
```

```
## Also defined by 'RNeXML'
```

```
## Found more than one class "phylo" in cache; using the first, from namespace 'treeio'
```

```
## Also defined by 'RNeXML'
```

```
## Joining, by = "feature_id"
```

```
## Joining, by = "feature_id"
```

```
## Joining, by = "feature_id"
```

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

```
# dada_sim <- phyloSim(mothur_signal$p4d, method = c("Cmean"))
```

```
# dada_sim %>% plot()
```

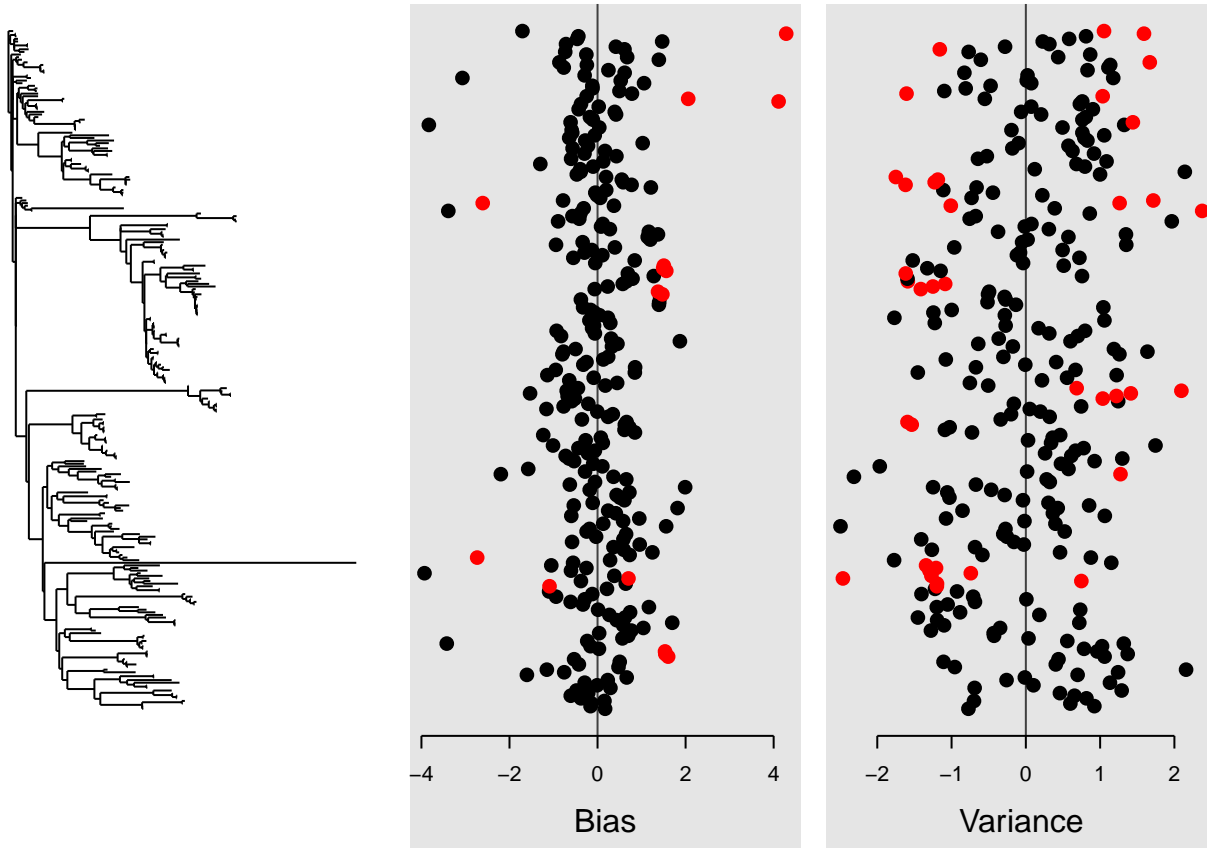
```
dada_signal$signal %>%  
  knitr::kable()
```

Metric	Method	stat	pvalue
sc_bias	Cmean	0.1536775	0.001
sc_bias	I	0.0744907	0.004
sc_bias	K	0.0207588	0.071
sc_bias	K.star	0.0247799	0.007
sc_bias	Lambda	NA	NA
sc_variance	Cmean	0.2438078	0.001
sc_variance	I	0.0983126	0.001
sc_variance	K	0.0342892	0.002
sc_variance	K.star	0.0263989	0.001
sc_variance	Lambda	0.7579744	0.001

```
# dada_bias_corr <- phyloCorrelogram(dada_signal$p4d, trait = "sc_bias")
# plot(dada_bias_corr)
```

```
# dada_variance_corr <- phyloCorrelogram(dada_signal$p4d, trait = "sc_variance")
# plot(dada_variance_corr)
```

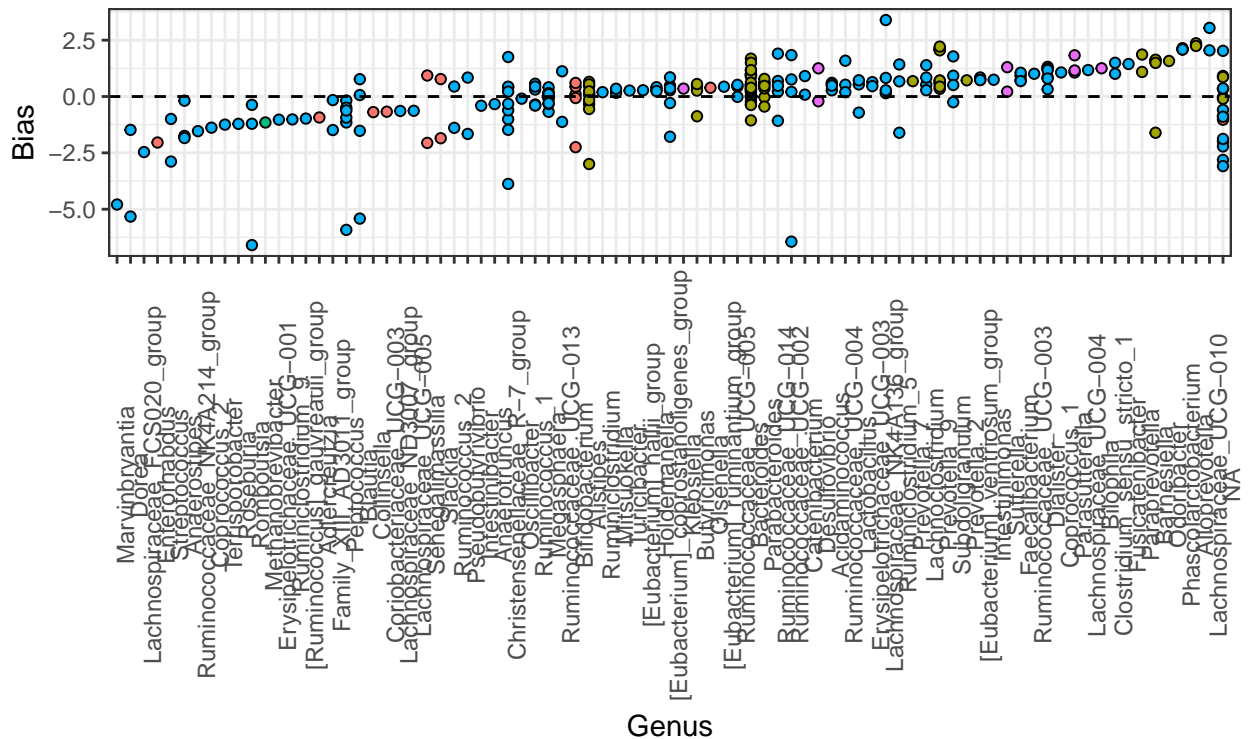
```
dotplot(dada_signal$p4d, dot.col = (dada_signal$lipa$p.value < 0.05) + 1,
        center = FALSE, scale = FALSE,
        grid.horizontal = FALSE,
        trait.labels = c("Bias", "Variance"),
        show.tip = FALSE,
        tree.ladderize = TRUE)
```



```
dada_signal$df %>% left_join(dada_tax) %>%
  mutate(Rank6 = fct_reorder(Rank6, Bias)) %>%
  select(Rank2, Rank6, Bias, Variance) %>%
```

```
# gather(Metric, Value, -Rank2, -Rank6) %>%
ggplot() + geom_point(aes(x = Rank6, y = Bias, fill = Rank2), shape = 21) +
# facet_wrap(~Metric, scales = "free_y", ncol = 1) +
geom_hline(aes(yintercept = 0), linetype = 2) +
theme_bw() +
theme(axis.text.x = element_text(angle = 90), legend.position = "bottom") +
labs(x = "Genus", fill = "Phylum")
```

```
## Joining, by = "feature_id"
```



Phylum ● Actinobacteria ● Bacteroidetes ● Euryarchaeota ● Firmicutes ● Proteobacteria

```
dada_signal$df %>% left_join(dada_tax) %>%
mutate(Rank6 = fct_reorder(Rank6, Variance)) %>%
select(Rank2, Rank6, Bias, Variance) %>%
# gather(Metric, Value, -Rank2, -Rank6) %>%
ggplot() + geom_point(aes(x = Rank6, y = Variance, fill = Rank2), shape = 21) +
# facet_wrap(~Metric, scales = "free_y", ncol = 1) +
geom_hline(aes(yintercept = 1), linetype = 2) +
theme_bw() +
theme(axis.text.x = element_text(angle = 90), legend.position = "bottom") +
labs(x = "Genus", fill = "Phylum")
```

```
## Joining, by = "feature_id"
```

