

Notes S5 - MiSeq vs. SMRT-sequencing profiles

Bodenhausen et al. (ENTER YEAR and JOURNAL HERE)

2018-08-22

Contents

Readme	1
Summary	1
Setup: sample numbers and color code	2
Number of sequences per sample	3
Rarefaction analysis	4
Alpha Diversity	5
Filter data	6
Beta Diversity	7
Taxonomy	9
Abundant <i>Olpidium</i> OTU in <i>Arabidopsis</i> root samples	11
References	12

Readme

This script contains the code for the detailed comparison of MiSeq vs. SMRT-sequencing based ITS community profiles. It was build based on the main code, loads the same input data and functions but only returns this markdown output as a result.










Summary

The second technical goal of this study was to benchmark the applied MiSeq-based fungal community profiling method with a long-read sequencing approach. The drawback of MiSeq-based approaches is the technically required short amplicon length (up to ca. 550 bp) so that either the ITS1 or the ITS2 region can be sequenced, resulting in limited taxonomic resolution. We therefore compared the MiSeq-based method covering the ITS1 region (PCR primers ITS1F and ITS2; [1]) with the entire ITS region (PCR primers ITS1F and ITS4) using SMRT sequencing. Both approaches avoid amplifying plant sequences while abundantly capturing the AMF, reproduce similar taxonomic compositions and confirm each other in diversity patterns, probably because of the common forward primer. Of note, the representative OTU sequences of the dominant *O. brassicae* of the MiSeq approach (ITS1 region) matched to 100% on its counterpart from SMRT-sequencing (whole ITS region). Similar to Tedersoo et al. (2018) [2], we summarize the benchmarking of MiSeq- vs. SMRT sequencing approaches with their characteristics of high-throughput vs. high-resolution, respectively.

Setup: sample numbers and color code

Table 1: Samples

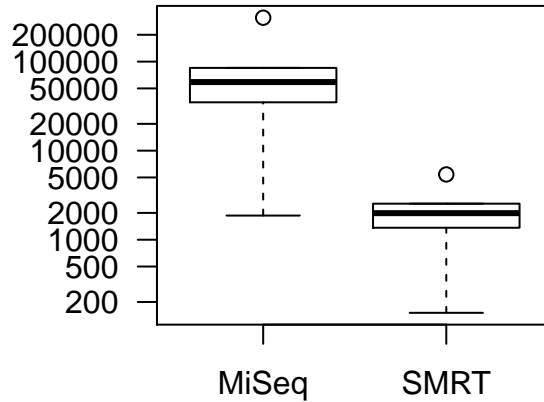
	low	medium	high
Arabidopsis	8	7	9
Petunia	7	8	9
Soil	4	4	4

	low	medium	high
Arabidopsis			
Petunia			
Soil			

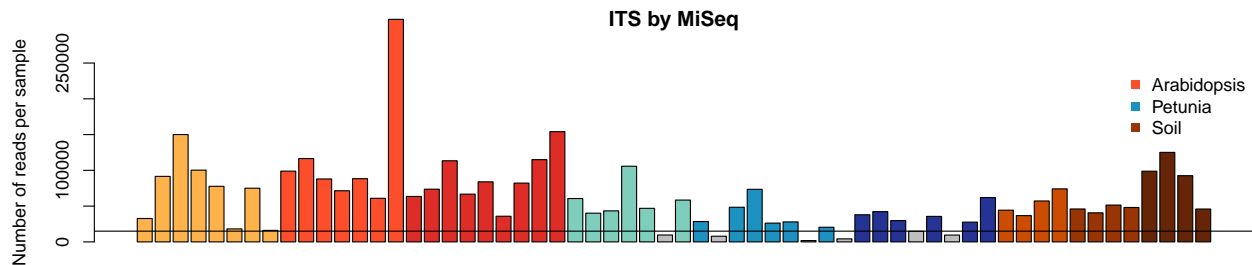
Number of sequences per sample

We first compared the throughputs of the MiSeq- and SMRT-sequencing based fungal ITS profiling approaches.

Method: The number of sequences per sample is displayed for the raw non-filtered data.

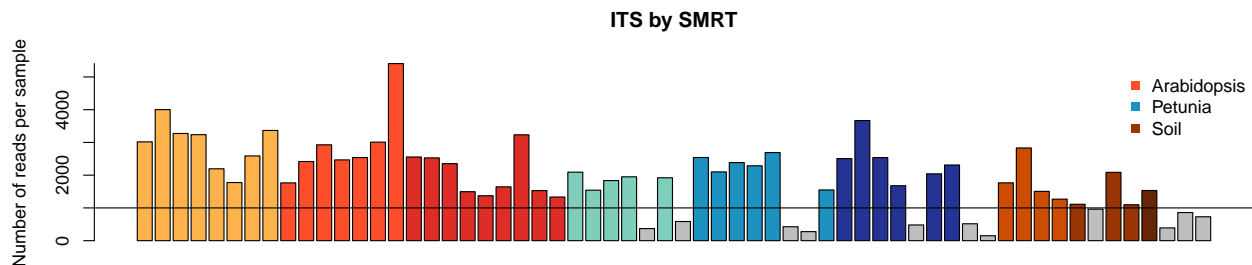


The *MiSeq-sequencing* based fungal ITS profiling approach yielded a total of 3'809'350 sequences with individual samples being sampled with a median of 54'337 sequences per sample.



Black horizontal line show the threshold used for rarefying (15'000). 6 samples colored in grey were removed for the alpha diversity analysis.

The *SMRT-sequencing* based fungal ITS profiling approach yielded a total of 118'607 sequences with individual samples being sampled with a median of 1'995.5 sequences per sample.



Black horizontal line show the threshold used for rarefying (1'000). 11 samples colored in grey were removed for the alpha diversity analysis.

CONCLUSION: The *MiSeq-sequencing* based fungal ITS profiling approach offers much higher throughput with a median of 54'337 sequences per sample vs. a median of 1'995.5 sequences per sample of the *SMRT-sequencing* approach. The higher sampling depth permitted to chose a higher threshold to sub-sample the data (MiSeq: 15'000, SMRT: 1'000).

Rarefaction analysis

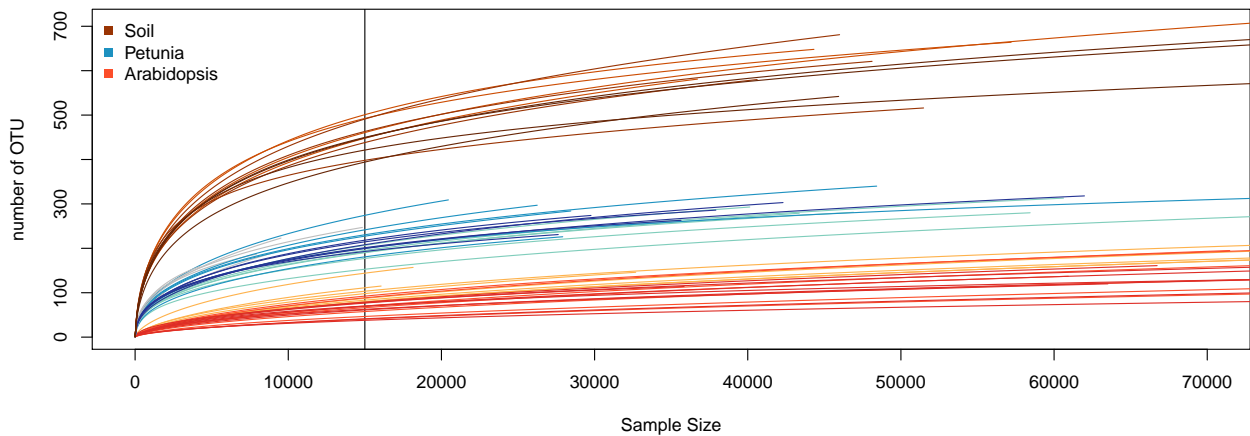
We then compared the sampling depths between the MiSeq- and SMRT-sequencing based fungal ITS profiling approaches.

Method: The sampling intensity analysis was conducted on the raw non-filtered data.

The *MiSeq-sequencing* based ITS profiling approach yielded a total of 1'688 fungal OTUs, while 334 fungal OTUs were obtained by *SMRT-sequencing* based ITS profiling.

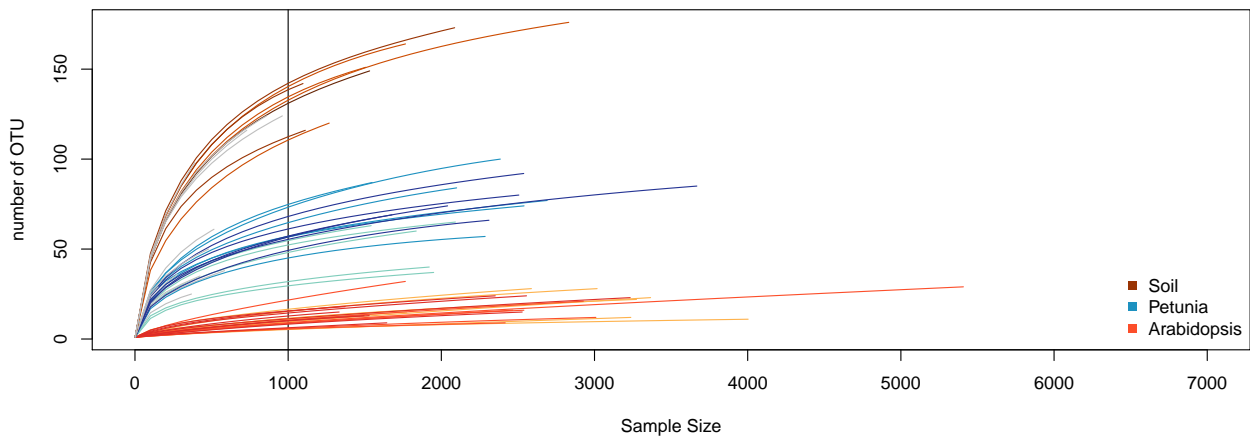
MiSeq

Diversity is highest in the soil followed by Petunia and then Arabidopsis:



SMRT

Diversity is highest in the soil followed by Petunia and then Arabidopsis:



CONCLUSION: With MiSeq there are 6 samples below threshold (15000) and with SMRT there are 11 samples below threshold (1000). The higher sampling depth of the *MiSeq-sequencing* based approach yields higher numbers fungal OTUs (MiSeq: 1'688, SMRT: 334) and rarefaction curves better point to sampling saturation. Both methods reveal the same biological richness pattern with highest richness in the soil followed by Petunia and then Arabidopsis.

Alpha Diversity

In a next step, we compared the MiSeq- and SMRT-sequencing based fungal ITS profiling approaches for how sampling intensity affected the OTU richness.

Method: The raw non-filtered dataset was rarefied 500x times and OTU richness was determined at each rarefaction event. Then a 2-factor ANOVA with factors ‘species’ (Arabidopsis, Petunia and Soil) and ‘treatment’ (low, medium, high) is conducted.

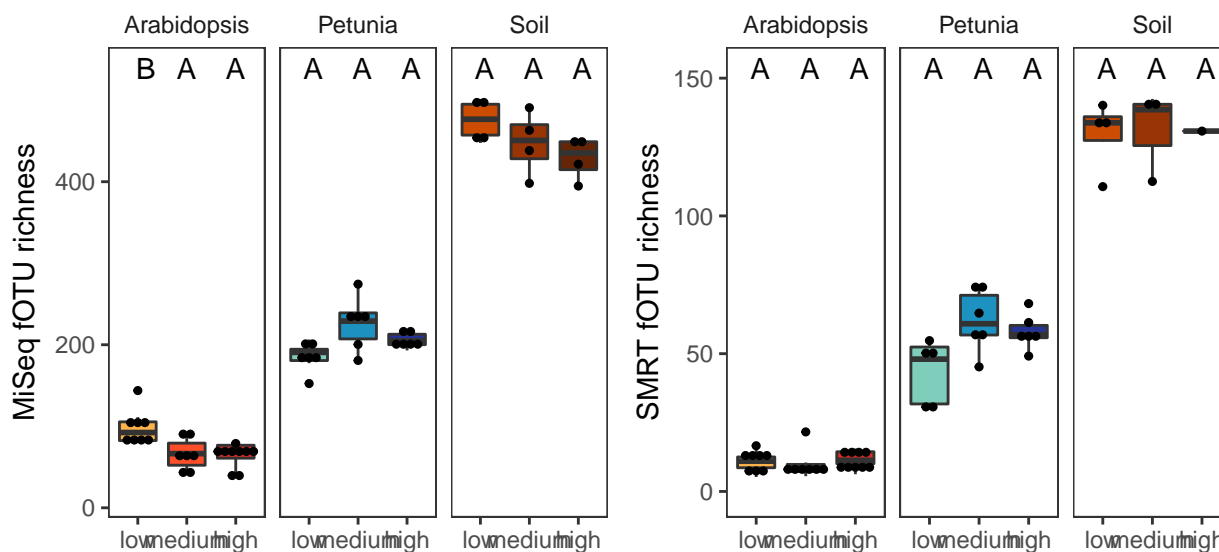
MiSeq, richness ANOVA (~ Species*Treatment)

	F value	Pr(>F)
Species	388.9	0.0000000000000000000000004002
Treatment	4.558	0.01576
Species:Treatment	5.317	0.001358

SMRT: richness ANOVA (~ Species*Treatment)

	F value	Pr(>F)
Species	264.9	0.000000000000000000000008436
Treatment	1.172	0.3203
Species:Treatment	1.002	0.4177

Richness plots



CONCLUSION: Although the *MiSeq-sequencing* based approach yielded higher numbers of fOTUs, both methods capture similar richness patterns with highest richness in the soil followed by Petunia and then Arabidopsis. The statistical evaluation found significant ‘species’ (Arabidopsis, Petunia vs. Soil) effects with both methods. The ‘treatment’ (low, medium, high) and ‘species x treatment’ effects found with the *MiSeq-sequencing* based approach were not found with *SMRT-sequencing* based ITS profiling.

Filter data

We utilized the raw non-filtered data for analyses involving random subsampling (rarefaction and alpha diversity). Here, the rarefying functions to normalize the data.

For the subsequent analyses we utilize data that we normalized either by total sum scaling (TSS) or trimmed mean of M values (TMM) methods. We use TSS and TMM normalized counts for data display and for differential abundance testing, respectively.

For these analyses we want to focus on the OTUs, which can be robustly quantified and we additionally filter the data. The rationale is to focus on abundant taxa but also to keep low abundant taxa (e.g., singletons) if they can be detected independently in replicate samples. The filtering threshold is to include all OTUs that don't have at least 4 assigned reads in at least 4 samples (the size of the smallest group in our data).

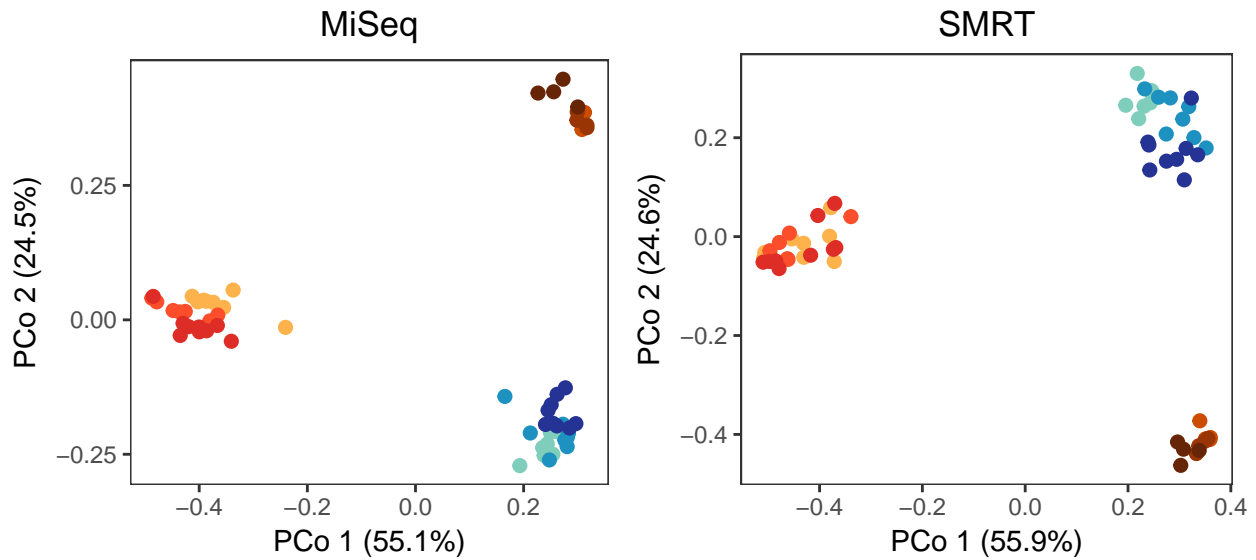
- For ITS, there are 508 OTU left for the analysis. This corresponds to 30.09% of the initial OTUs (1688). On average, the OTUs that pass the filtering threshold still account for 99.39% of the sequences.
- For SMRT, there are 109 OTU left for the analysis. This corresponds to 32.63% of the initial OTUs (334). On average, the OTUs that pass the filtering threshold still account for 96.21% of the sequences.

Beta Diversity

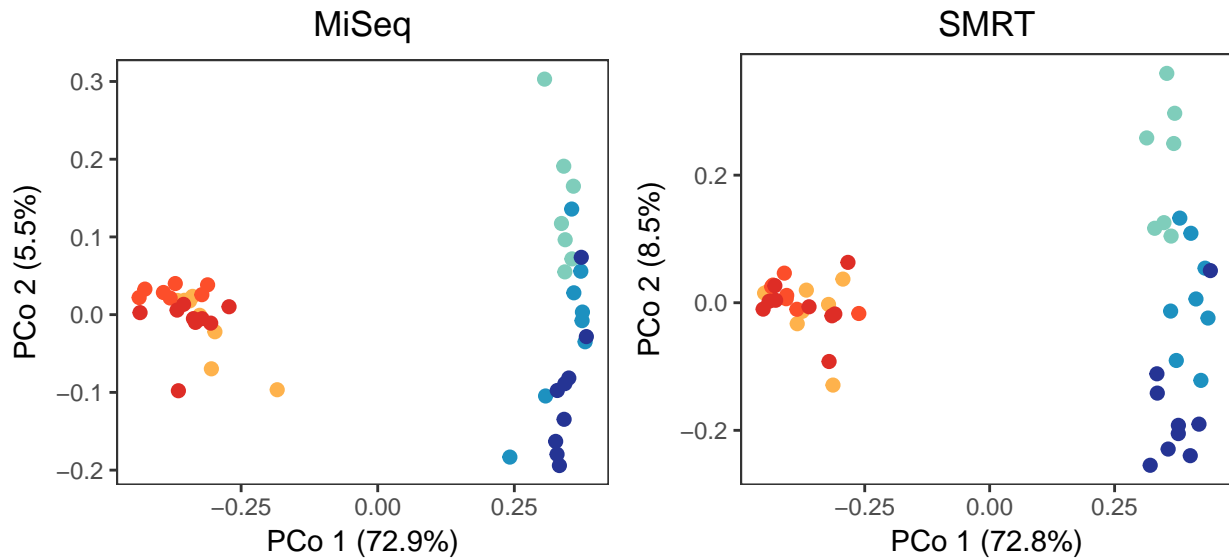
In a next step, we compared beta diversity patterns between MiSeq- and SMRT-sequencing based fungal ITS profiling approaches.

Method: A first analysis included all samples (including the soil samples) and the second one was limited to root samples. We utilize normalized data (filtered, TSS, log=FALSE, expressed as permille relative abundance) and follow Leff et al. by square-root transforming the counts. PCoA ordinations were computed based on Bray-Curtis dissimilarity (binary=FALSE) employing the 'Phyloseq' package. In a second step, we utilized PERMANOVA to test the effects of factors 'species' (Arabidopsis, Petunia and Soil) and 'treatment' (low, medium, high).

Principle Coordinate Analysis, with soils samples



Principle Coordinate Analysis, root samples only



PERMANOVA

	R2	Pr(>F)
sampleTable\$Species	0.7826	0.001
sampleTable\$Treatment	0.01833	0.02
sampleTableSpecies : sampleTableTreatment	0.0333	0.006
Residuals	0.1658	NA
Total	1	NA
NA	NA	NA

	R2	Pr(>F)
sampleTable\$Species	0.789	0.001
sampleTable\$Treatment	0.01883	0.007
sampleTableSpecies : sampleTableTreatment	0.03481	0.004
Residuals	0.1574	NA
Total	1	NA
NA	NA	NA

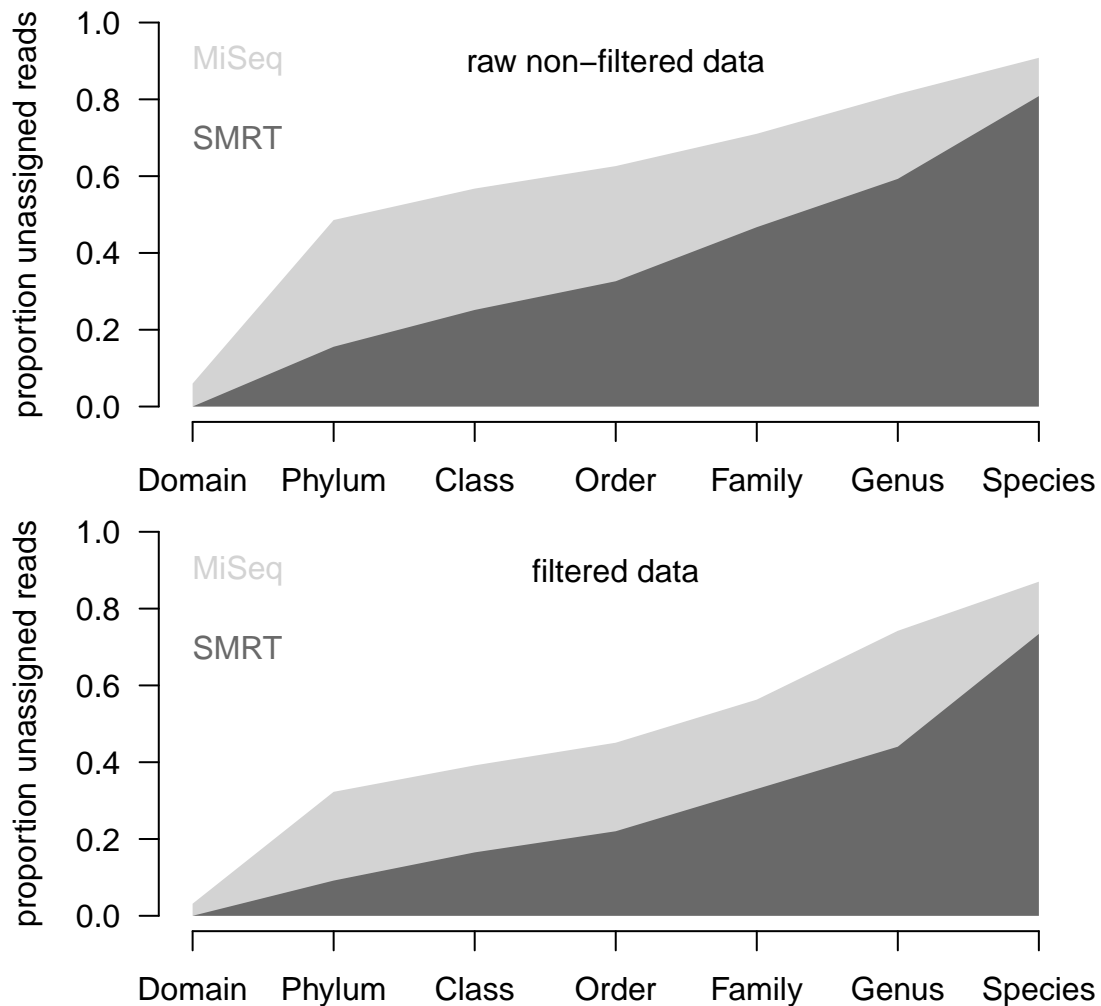
CONCLUSION: Both *MiSeq-sequencing* and *SMRT-sequencing* based ITS profiling methods produced similar beta diversity pattern with *Arabidopsis*, *Petunia* and soil samples clustering appart in an ordination space, where the PCo axes summarize similar amounts of variation for both methods. Zooming into the root samples only, both methods reveal a similar clustering of *Petunia* samples along the P gradient (along PCo axis 2). Consistent with the ordination patterns, PERMANOVA analysis of both methods confirmed significant ‘species’, ‘treatment’ and ‘species*treatment’ effects.

Taxonomy

Taxonomic resolution

In a first step, we compared the taxonomic resolution of the MiSeq- and SMRT-sequencing based fungal ITS profiling approaches.

Method: For this analysis, we determined the proportions of “unassigned” fungal OTUs at each rank in the taxonomy tables of both datasets. We conducted this analysis both with the raw non-filtered and the filtered taxonomy tables.

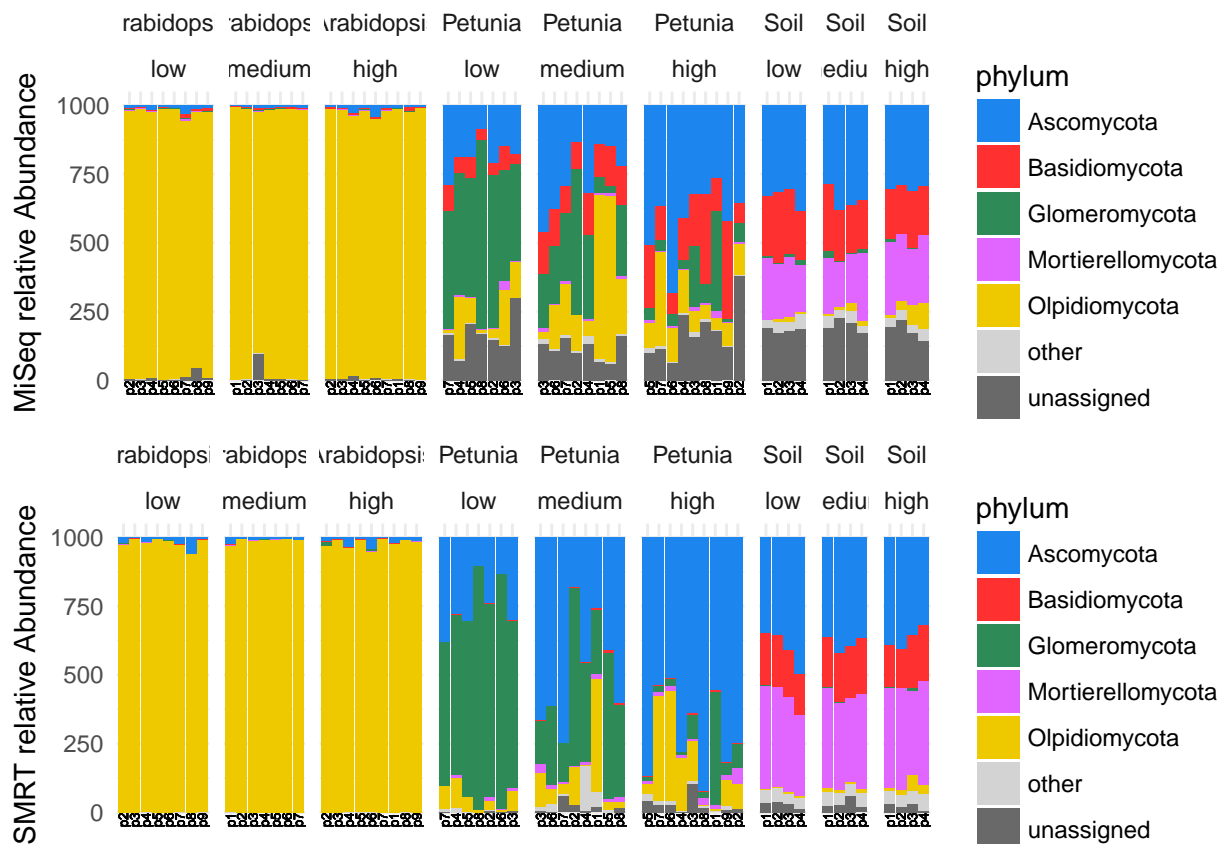


CONCLUSION: The proportions of “unassigned” fungal OTUs are in general lower in the *SMRT-sequencing* based ITS profiling method, where the longer ITS fragment is sequenced. Hence, taxonomic resolution is higher in the *SMRT-sequencing* compared to the *MiSeq-sequencing* based ITS profiling approach.

Taxonomic composition at Phylum level

Then, we graphed the taxonomic composition of the MiSeq- and SMRT-sequencing based fungal ITS profiling approaches at the Phylum level.

Method: Bargraphs were produced from normalized data (TSS, log=FALSE), which was expressed as permille relative abundance.



CONCLUSION: The *SMRT-sequencing* and the *MiSeq-sequencing* based ITS profiling approaches produce overall similar community profiles with a few distinctions: While both ITS profiling methods depict the abundance of the Olpidiummycota in Arabidospis roots in a similar manner, the *SMRT-sequencing* based method captures enhanced levels of Ascomycota and Glomeromycota in Petunia roots. Similarly, the *SMRT-sequencing* based method also quantifies higher levels of Mortierellomycota in soil samples. In contrast, petunia root and soil profiles of the *MiSeq-sequencing* based approach have higher levels of unassigned OTUs.

Abundant *Olpidium* OTU in *Arabidopsis* root samples

We found a dominant *Olpidium brassicae* OTU in the *Arabidopsis* root samples utilizing the *MiSeq-sequencing* based ITS profiling approach. To exclude that this was a technical artifact we developed the *SMRT-sequencing* based ITS profiling method and we confirmed the dominance of *Olpidium brassicae* by sequencing the entire ITS region (PCR primers ITS1F and ITS4).

Method: Boxplots were produced from normalized data (TSS, log=FALSE), which was expressed as permille relative abundance.

MiSeq data

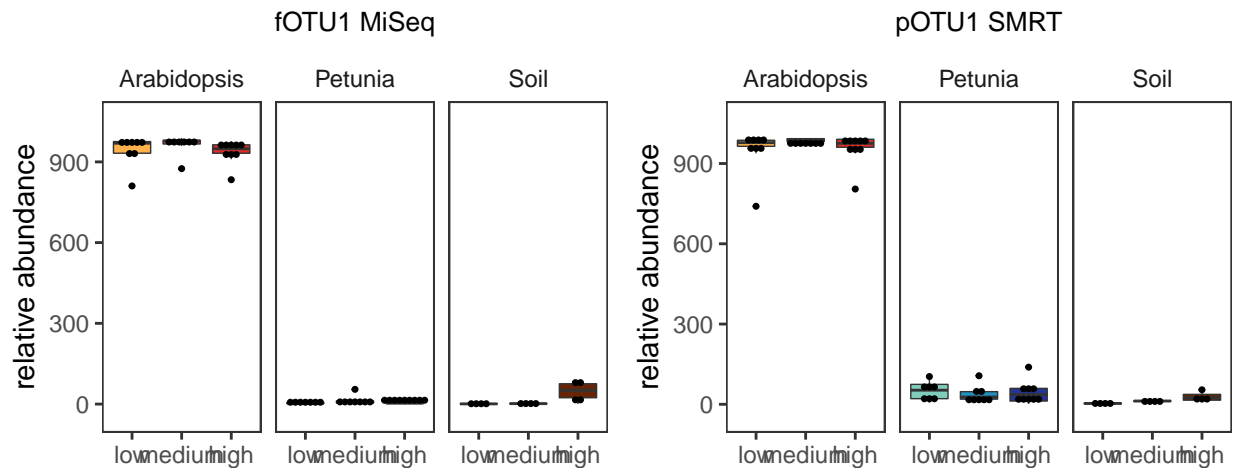
##	OTU	Arabidopsis	Petunia	Soil
## 1	fOTU_2	21.837255409	147.166892805	9.08474461
## 2	fOTU_1	946.385653888	8.256252202	16.81503811
## 3	fOTU_527	0.007009403	0.166890898	0.04063825
## 4	fOTU_318	0.012087718	0.008079231	0.37838979

The *MiSeq-sequencing* based ITS profiling approach captured 4 OTUs belonging to the Phylum *Olpidiomycota*, of which fOTU1 accounted for the dominance in *Arabidopsis* roots.

SMRT data

##	OTU	Arabidopsis	Petunia	Soil
## 1	pOTU_1	961.78861	43.79451	14.2364506
## 2	pOTU_3	20.26482	61.62742	0.4885352

The *SMRT-sequencing* based ITS profiling approach captured 2 OTUs belonging to the Phylum *Olpidiomycota*, with pOTU1 accounting for the dominance in *Arabidopsis* roots. The OTUs in the SMRT dataset have the prefix “p” to discriminate from the fungal ones of the MiSeq data (fOTUs).



CONCLUSION: Both ITS profiling methods capture the abundance of the dominant *Olpidiummycota* OTU in *Arabidopsis* roots in a highly similar manner.

References

1. McGuire SGAP Krista L. AND Payne. Digging the new york city skyline: Soil fungal communities in green roofs and city parks. PLoS ONE. Public Library of Science; 2013;8: 1–13. doi:10.1371/journal.pone.0058020
2. Tedersoo L, Tooming-Klunderud A, Anslan S. PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. New Phytologist. 2018; doi:10.1111/nph.14776