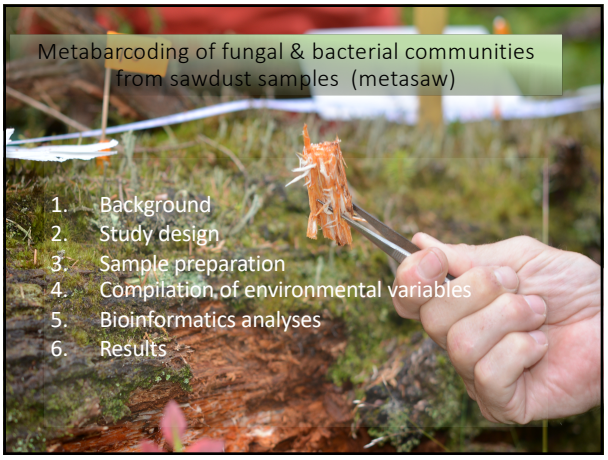
Case study: metabarcoding microbial communities inhabiting deadwood in boreal forests



Sundy Maurice
Oslo Mycology Group
Evolution and genetics
Faculty of Biosciences
University of Oslo

25th March 2019, Bioinformatics for DNA-metabarcoding and environmental sequencing



Forest practices leading to decline of wood substrates

Natural, protected, old-growth spruce forest with long continuity where dead wood is a characteristic feature



Old-growth forest (Kotinen, Hame)



Managed forest, Southern Finland

- Lack of dead wood resulting in population decline & species extinction

WP4.2: Metabarcoding of fungal & bacterial communities from sawdust samples (metasaw)

Overarching aim: Investigate the effect of forest management types on fungal & bacteria diversity and richness

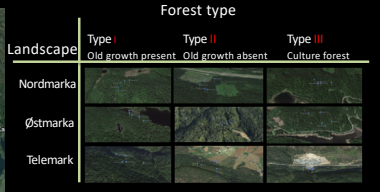
1. Are there any differences in community composition and diversity, for both bacteria & fungi, due to forest management types and/or geographically distinct landscapes?

2. What are the factors that correlate with community composition?

3. Any difference in composition within the same log?

Study design

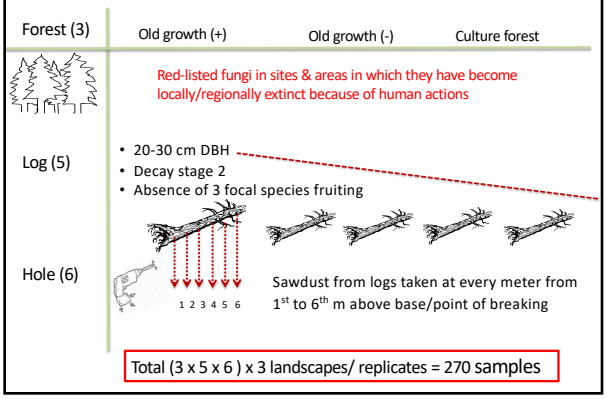




- Type I: Old growth forest (reserves) with focal species present
- Type II: Old growth forest with absence of focal species
- Type III: Cultured forest, mature forest, once clear-cut with logs removed but now with new logs

3 landscapes represented by 3 forest types each

Study design



Sample processing



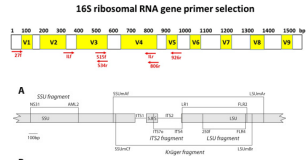
1. Mixing the sawdust to homogenize
2. Weighing 3g of fresh mass in 50 mL Falcon tubes
3. Freeze drying (30h)
4. Crushing in 50 mL Falcon using 2 * 6.8mm Zirconium oxide beads (Saveen Werner) using a Fast prep homogenizer (MoBIO). 2*2 Falcons at a time
5. CTAB & Phenol/Chloroform (40 mL) DNA extraction
6. Cleaning DNA extract with the E.Z.N.A Soil DNA kit [DNA] Qbit average 60 ng/µl
7. Diluting each sample to *ca.* 5 ng/µl



Wood particles of of different sizes & not only the sawdust
Freeze dry 48 hrs, ~1 g lost.

Sample processing : 12/2016

Sample processing



Primers used & region

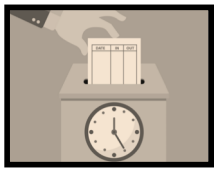
16S: 515F & 806R (240 pb)
modified by Caporaso et al. 2011

ITS: IT4 and gITS7 (ca. 300pb)
ITS4 White et al 1990
gITS7 from Ihrmark et al 2012

8. Preparing metabarcoding libraries
9. Normalising PCR product (SequalPrep plates)
10. Pooling *96 samples & AMPure beads cleaning
11. Sequencing barcoded libraries on 2 Miseq lanes: 16S & ITS by (StarSeq, Germany)

Library prep 01-02/2017
Sequencing 04/2017

Send for library preparation & sequencing, *ca.* 1 month



What can be done in the meanwhile?

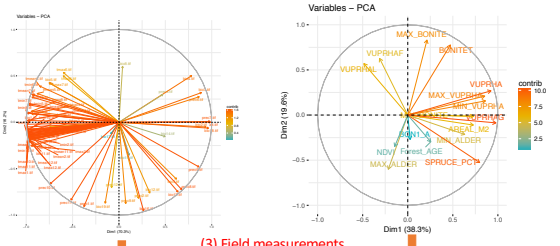
Take Home tips-1

- Learn & use Unix- useful beyond copy/paste cmd
- Talk to supervisors/collab. about organization of the analyses, ask for access to a cluster (Abel, Cod,
- Take a course in metabarcoding, **and** also practice
- Compile & prepare meta-data

Environmental variables

(1) Climatic variables
68 climatic variables (bioclim)

(2) Landscaped variables
43 landscape variables (Kilden + Satskog)



3 Selected variables:
Annual mean temperature
Annual precipitation
Temperature seasonality

(3) Field measurements
8 selected variables:
Decay stage,
Moisture content, Ground contact,
Bark cover, Epiphyte cover,
Length,
Spruce stand, Dead spruce

3 Selected variables
Area
Altitude
Forest quality

After 1 month …. data arrives

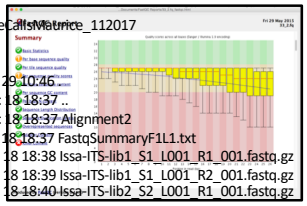


How/where do I get started with the data analyses?

Take Home tips-2

- Very important! Check quality of sequencing
e.g. FASTQC
- Encounter problems with upload/downloading data

```
/projects/clusteringits/rawdata/BaseCallMatrice_112017
-bash-4.1$ ls -lah
total 14G
drwxrwx--- 3 sandym users 2.0K Jan 29 10:46
drwxrwx--- 3 sandym users 2.0K Dec 18 18:37
drwxrwx--- 5 sandym users 2.0K Dec 18 18:37 Alignment2
-rwxrwx--- 1 sandym users 2.5K Dec 18 18:37 FastqSummaryF1L1.txt
-rwxrwx--- 1 sandym users 1.8G Dec 18 18:38 Issa-ITS-lib1_S1_L001_R1_001.fastq.gz
-rwxrwx--- 1 sandym users 2.0G Dec 18 18:39 Issa-ITS-lib1_S1_L001_R2_001.fastq.gz
-rwxrwx--- 1 sandym users 4.0G Dec 18 18:40 Issa-ITS-lib2_S2_L001_R1_001.fastq.gz
```



- Insufficient amount of data generated, problem with clustering e.g. 5 M sequences for reverse read (rather than 30-40 M for 16S)

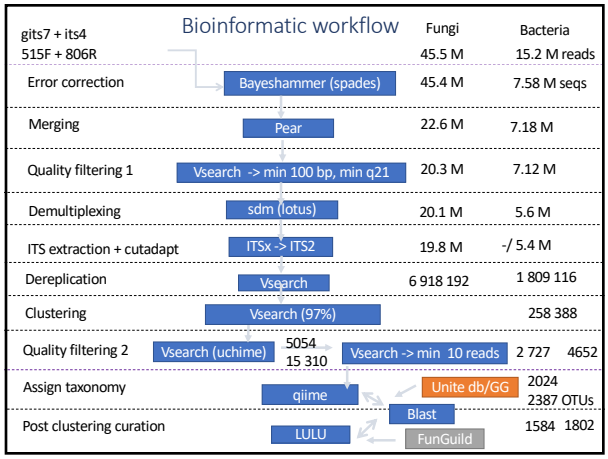
Bioinformatics workflow (16S & ITS)

Comparable steps

#	Analysis	Software	Bacteria	Fungi
1	Base correction	BayesHammer (Spades)	Yes	Yes
3	Merge reads 1 & 2	PEAR	Yes	Yes
4a	Quality filtering	fastx-toolkit	Yes	No
4b	Quality filtering	vsearch	Yes	Yes
5a	Re-orientation of the reads in same direction, 5´-> 3´	fqgrep	Yes	Yes
5b	Reverse complement the reverse read 3'-5' to 5'-3'	fastx-toolkit	Yes	Yes
6	Demultiplexing	sdm	Yes	Yes
8a	Remove primers both of forward and reverse ends	cutadapt	Yes	Yes
8b	ITS extraction	itsx	No	Yes
9	Dereplication & sort by abundance	vsearch	Yes	Yes
10	Clustering 97 %		Yes	Yes
11	*denovo* Chimera checking		Yes	Yes
12	Filtering OTU table & reset file with OTUs < 10 reads		Yes	Yes
13a	Generate OTUs table	Qiime	Yes	Yes
14	Assign taxonomy		Yes (NCBI, GG)	Yes (UNITE & Polypore DB)
15	Filter OTU tables to remove singleton otus		Yes	Yes
16	Convert biom to txt		Yes	Yes

Bio-info: 4-7/ 2017

Bioinformatic workflow

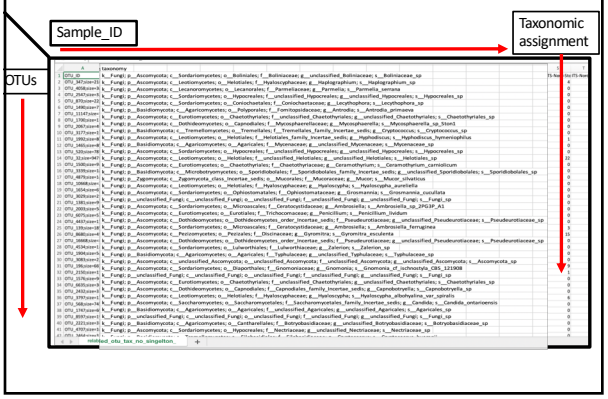


Importance of post bioinformatics filtering

Example from a different fungal dataset

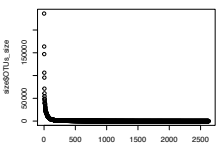
Filtering steps (post bio-informatics)	# of samples	# of OTUs	# of reads	Remarks
Raw OTU table	192	2188	16,196,863	
LULU	192	1632	16,196,863	rm erroneous OTUs by combining seq. similarity & co-occurence patterns
Remove replicates and no blast hit OTUs	181	1385	14,871,904	
Tag-switching correction owi_renormalized_10%	181	1385	13,511,207	rm **most** reads from tag-switching
Remove mock OTUs	181	1369	13,269,587	to rm tag leakage from mock community
Filter samples	176	1367	13,089,836	rm mock, PCR control & poscae12 & rhizaria OTUs
Remove tag-switching coming from host species	176	1367	12,816,168	rm manually the host leakage
Rarefy 10 000	176	1367	1,794,144	sample min 10,194 (Fompin3)

OTU= Operational Taxonomic Unit Table



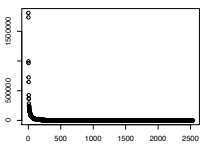
Number of OTUs obtained (288 inclu. replicates & -ve control)

16S = 2519 OTUs = **2** 953 094 reads



16S = **5** 607 492 read, post sdm

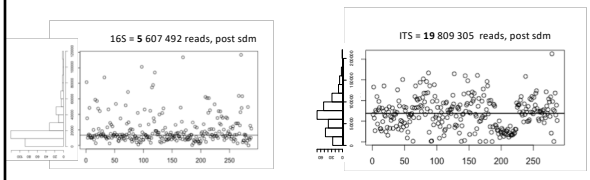
ITS = 2458 OTUs = **17** 990 661 reads



ITS = **19** 809 305 reads, post sdm

Number of sequences per sample

(288 inclu. replicates & -ve control)

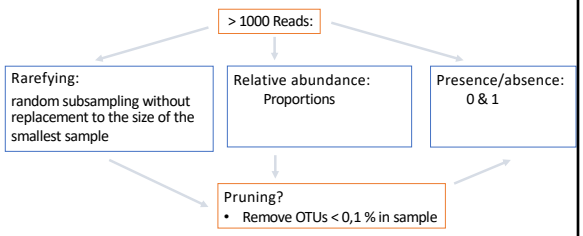


Min.	1st Quar.	Median	Mean	Max.
51	9464	13238	19470	116727

Min.	1st Quar.	Median	Mean	Max.
548	39006	69509	69022	213033

Conclusion: Large amount and high quality bacterial & fungal sequences

Normalisation methods



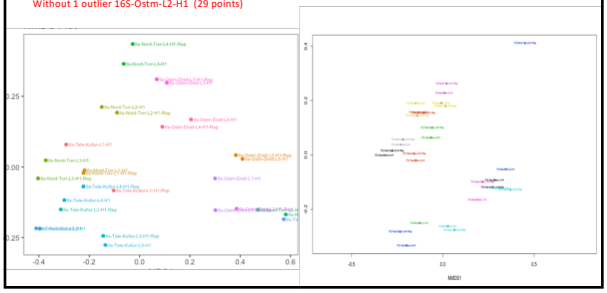
Rarified + Pruned
Protest correlation in procrustes's rotation

Are replicates more similar to each other (composition)

15 replicates (Culture forest, H1L1 to L5)

- 16S- Bacterial replicates
- ITS- Fungal replicates

Without 1 outlier 16S-Ostm-L2-H1 (29 points)



Take Home tips-3

- ❑ Play & learn your data, optimize threshold/filtering parameters … but learn to say STOP and move on with analyses
- ❑ Try to understand what you are doing
- ❑ Do the analyses in One-Go
- ❑ Account & correct tag hopping (0-10 % revealed!)

Tag switching/hitch-hacking/jumping/swapping/hopping can a big issue on Illumina, evidence with known samples/mock included

Try to make statistical sense of your data

WP4.2: Metabarcoding of fungal & bacterial communities from sawdust samples (metasaw)

Investigate the effect of forest management types on fungal & bacteria diversity and richness

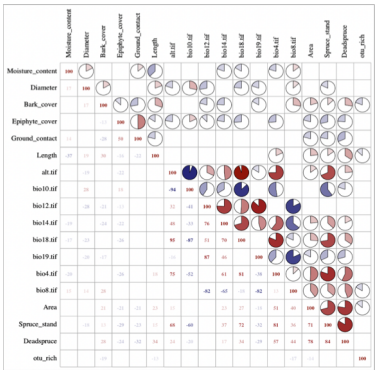
1. Are there any differences in community composition and diversity, for both bacteria and fungi, due to forest management types and/or geographically distinct landscapes?

2. What are the factors that correlate with community composition?

(landscape, forest type, hole number- whereas logs with similar diameter and DS are were selected, hence less likely to drive the structure

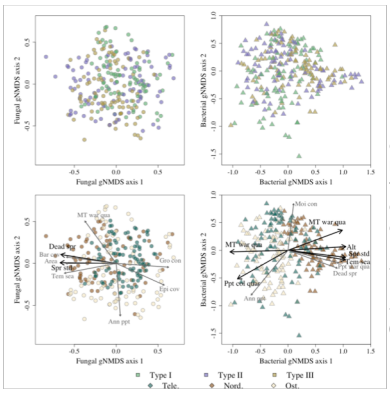
3. Any difference in composition among the 6 holes within the same log? I would expect to have a greater diversity at H1 (more decayed, higher moisture level) than H6 (sawdust with lower level of humidity)

Environmental variables- bacterial communities



- High collinearities still exist Bio10 (Mean Temp. of Warm. quar), alt, Bio 18= Ppt. of Wat. Quart.
- Select the ones that make sense biologically/ecologically

Main drivers of community structure

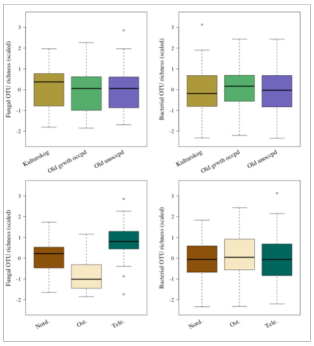


(a & b) Forest management
(c & d) Landscape
Left = Fungi
Right = Bacteria

Differences in the environmental and landscape correlates of fungal communities on logs compared to bacterial communities.

- Significant variables $|\text{Kendall } \tau| \geq 0.3$
- Marginally significant $0.2 \leq |\text{Kendall } \tau| < 0.3$

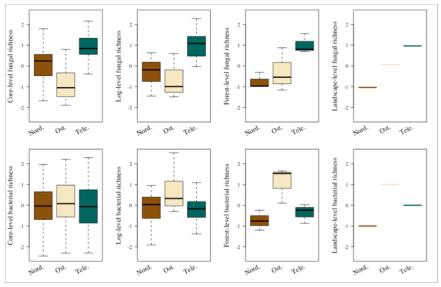
Forest management vs. Landscape



(a & b) Forest management
(c & d) Landscape
Left = Fungi
Right = Bacteria

Forest management patterns are not appreciably different in terms of richness

Diversity across spatial scales: from hole to landscape



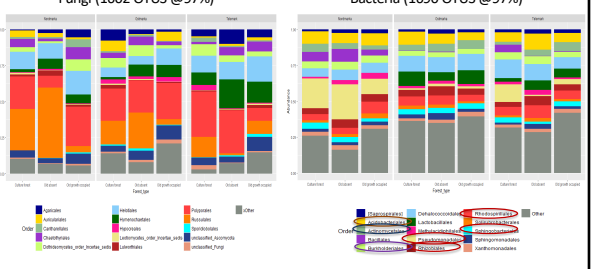
Total richness	Hole	Log	Forest	Landscape
Fungi (1602 OTUs)	57 to 656	486 to 882	946 to 1226	1200 to 1354
Bacteria (1690 OTUs)	84 to 766	474 to 1161	1267 to 1457	1507 to 1597

Try to make biological sense of your data

Community Composition Landscape & Forest types

Fungi (1602 OTUS @97%)

Bacteria (1690 OTUS @97%)



Bacteria	Fungi
❑ Make wood more water permeable ❑ Bacteria with direct decomposition ability ❑ Bacteria that stimulate fungal decomposition ❑ Bacteria that inhibit fungal decomposition	❑ Wood decay fungi ❑ Lichen-associated fungi present in wood? ❑ Knowledge about patterns of fungal growth in the logs

Generate some hypotheses, but be open for data exploration and ideas will pop-up

Perspectives

2. Interaction between the bacterial & fungal species? Are there any co-occurrence patterns observed?

2.1 Characterisation of the relationship

2.2 Correlation of main microbial taxonomic group with potential functions in ecosystem functioning, both for bacteria & fungi in wood

Inventory: records of fruiting body

❑ Onsite identification of fruiting body

107 polypores fruiting bodies were recorded

❑ Dried/dead specimens: 49 genus for a total of 268 fruiting bodies including polypores & corticioids

No always very easy to use all metadata compiled/gathered

Take Home tips-4

❑ Tag switching is a big issue on Illumina
❑ Be cautious with removal of outliers
❑ Generate some hypotheses… but be open for data exploration
❑ Downscale data output (OTUs & metadata). Learn how to overcome frustration
❑ Apply your skills to new cases, don´t let it rest just as knowledge

