Microbial activity in the soil and its seasonal changes: The picture provided by metatranscriptomics and metaproteomics

Petr Baldrian

Institute of Microbiology, Prague, Czech Republic

Ecology of forest topsoil

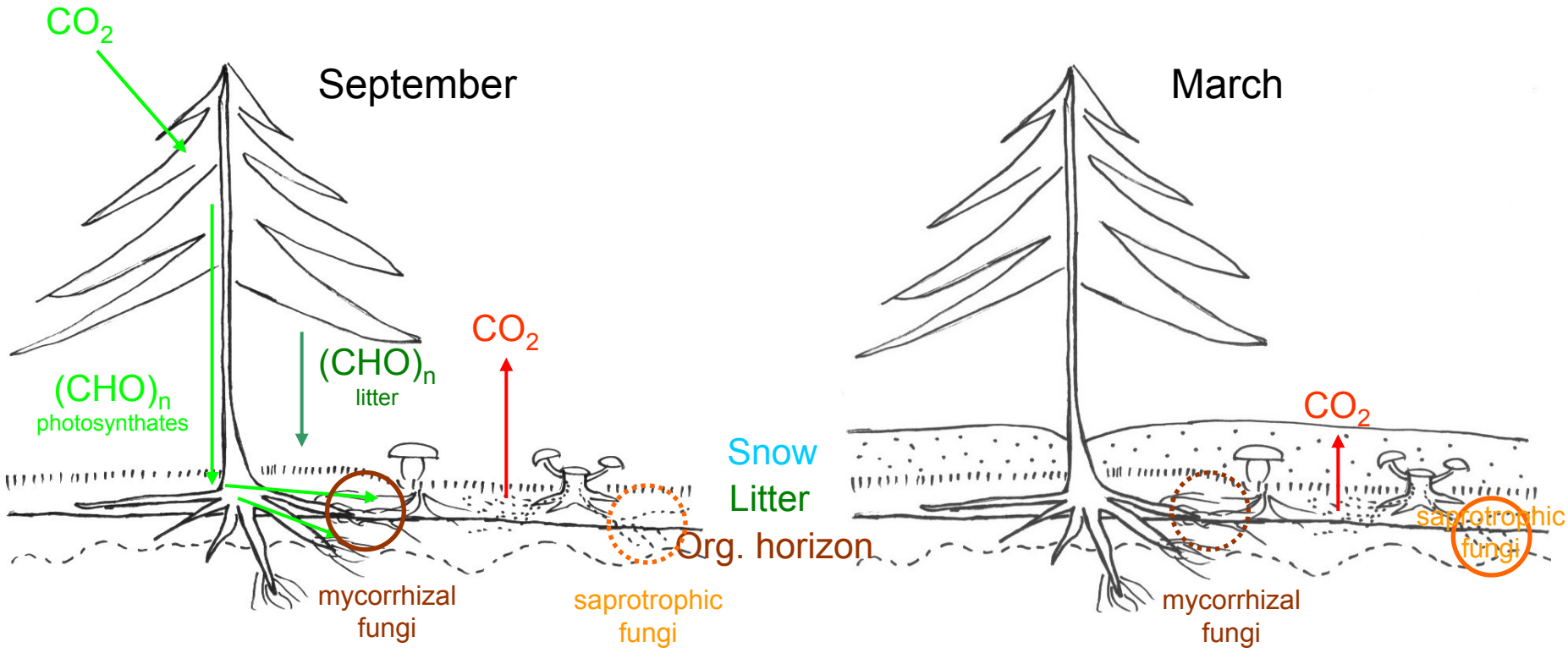
Nutrients

- leaf and root litter	- major source of organic matter in soil - cellulose, lignin, hemicellulose, pectin, proteins - numerous enzymes required for degradation
- root exudates	- nutrient source in mineral and organic soil - more readily available C

Microorganisms

- saprotrophic fungi	- obtain carbon by organic matter decomposition - production of extracellular degradative enzymes
- mycorrhizal fungi	- obtain carbohydrates from symbiotic plants - provide soil nutrients to plants - saprotrophic abilities vary
- bacteria	- preferential utilization of easily available compounds - degradation of biopolymers by some taxa

Fungal activity in litter and soil in contrasting seasons



- High share of fungi in the ecosystem is reflected by high fungal contribution to transcription and protein production, especially in litter
- Fungal activity is important for decomposition of complex organic matter
- Seasonal differences in rhizodeposition affect the share of root-associated / saprotrophic fungi
- Activity of root-associated fungi decreases in winter

Methods choice and methodology

Exploring ecosystem functioning

Stable isotope probing – indication of microorganisms involved in certain processes containing incorporation of C (stable isotope 13C) or N (stable isotope 15N), need for labelled substrate. Microorganisms building biomass on incorporated substrate get their biomolecules (PLFA, DNA, RNA) labelled. There is danger of cheating (comensalism) and cross-feeding (predation on microorganisms accummulating labelled isotope). Leads to „label dillution“ over time.

Metaproteomics – isolation of total protein and „sequencing“ of peptides. Peptides can be identified, but carry less information than nucleic acids (degenerated code). Difficulty with sequence assignment + different stability of proteins in the environment over time. Can be extremely powerfull in combination with metatranscriptomics or metagenomics but difficult to be used alone.

Microarray technologies – hybridization of DNA or RNA with chips can avoid reverse transcription bias or PCR bias. High resolution across levels of expression. However, microarrays only cover the genes used for their desing. Excellent for study of individual taxa in the environment where sequenced genomes are available.

Isolation and analysis of microbial strains – challenging but can be very useful.

Exploring ecosystem functioning - metagenomics

Pros:

- can indicate genetic potential of the community
- DNA community relatively stable over time (good representation of the ecosystem)
- can theoretically reveal co-occurrence of genes (when longer contigs of DNA covering several genes are assebled)
- sometimes gives exact identity of the gene source (when the gene sequence and 16S co-occur on an assebled molecule)
- powerfull for exploration of bacteria / archaea

Cons:

- eukaryota (e.g. fungi) contain much of noncoding DNA (up to over 90%)
- eukaryotic genes contain introns
- due to the above, short reads of eukaryotic genes can give no information (noncoding DNA) or be difficult to assign (contains both exons and introns)
- need for assembly
- potential is not the function, presence is not activity (extracellular DNA, pseudogenes, levels of expression…)
- for diverse ecosystems, high depth of sequencing required

Metatranscriptomics - opportunities

- can indicate real activity in the studied ecosystem; fast response to disturbance / experimental treatment
- little danger of „ancient“ RNA from dead cells – such RNA decomposes rapidly
- avoids the problem with noncoding DNA
- for gene-coding sequences, functional and taxonomic assignment is more simple than for DNA, even for shorter reads
- powerfull for exploration of both prokaryota and eukaryota
- in the case eukaryota, ease of isolation and purification by „fishing“ the poly-A tails of mRNA molecules (but does not always work)
- metatranscriptomes much less complex than metagenomes - results in easier assembly (higher coverage with the same number of reads)
- with sufficient depth of sequencing, relative importance of individual processes can be analysed to some depth by comparison of transcription level
- while metagenomics tell which genes **may be involved**, metatranscriptomics tell which genes actually **are involved** (expressed)

Metatranscriptomics - limitations

- expression is highly regulated and corresponds to „actual“ conditions, not „mean“ conditions of the site; for example, transcription increases by orders of magnitude when dry soil is moistened
- mRNA is short-lived so the metatranscriptome reveals what happened within last tens of minutes
- the amount of extracted RNA usually makes amplification necessary; PCR amplification of cDNA brings bias
- extracted RNA contains much rRNA that can be difficult to remove
- genes with low level of expression are difficult to recover
- there is little (if any useful) information on mRNA stability in time and translation rate and thus the amount of protein molecules synthesized per mRNA molecule in its lifetime
- metagenomics can theoretically deliver long contigs - chromosome fragments with multiple genes that are from the same genome; this is impossible for metatranscriptomics

Workflow

Sampling
- sample collection
- stabilization
- storage

Library preparation
- RNA isolation
- RNA purification
- DNA removal
- rRNA removal (to recover all mRNA) or capture of eukaryotic mRNA (poly A)
- RNA fragmentation
- cDNA synthesis
- RNA removal
- adapter ligation (with barcodes when multiple samples will be sequenced)
- amplification
- selection of molecules of approppriate size

Sequencing

Data analysis
- long sequences – direct annotation (function, binning to microbial taxa)
- short sequences – sequence assembly and annotation of contigs (scaffolds)

Sample collection and storage

Each study site 64 m^2

Eight 4.5 cm soil cores at defined locations, approx. 2.0 m from each other

After core collection (within 30 min, on site):

- collection of core material

- vegetation, roots and mineral soil discarded

- litter material and soil organic horizon separated into two composite samples (each containing materials from all cores)

- sieving of soil (0.5 mm sieve), cutting of litter (scissors)

- aliquoting into tubes

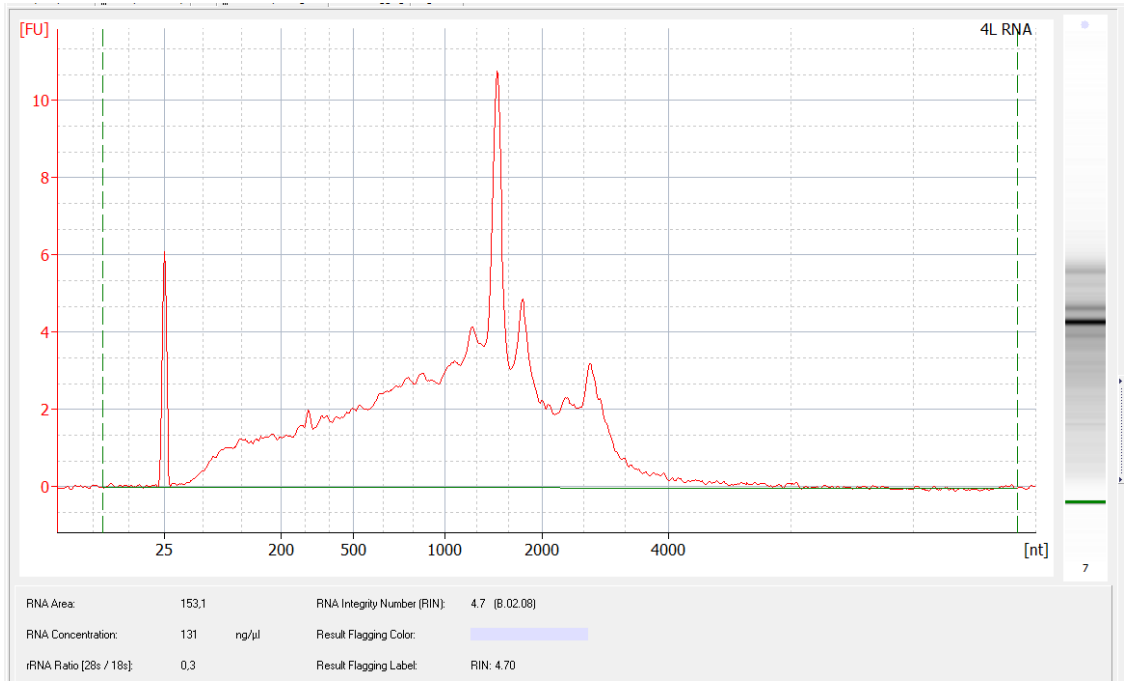
- flash-freezing in liquid nitrogen

Storage for transport on dry ice

Storage upon arrival at -80 °C

RNA isolation and purification

- sample homogenization by mortar and pestle in liquid nitrogen
- MoBio RNA PowerSoil Kit used for isolation in combination with Zymo Research OneStep PCR Inhibitor Removal Kit (PVPP removal of humic and fulvic acids)
- DNA removal (DNAse)
- verification of DNA removal (no PCR amplification of 16S)
- check of RNA yield (Qubit)
- check of RNA quality (BioAnalyzer)
- storage of isolated total RNA (-80 °C)

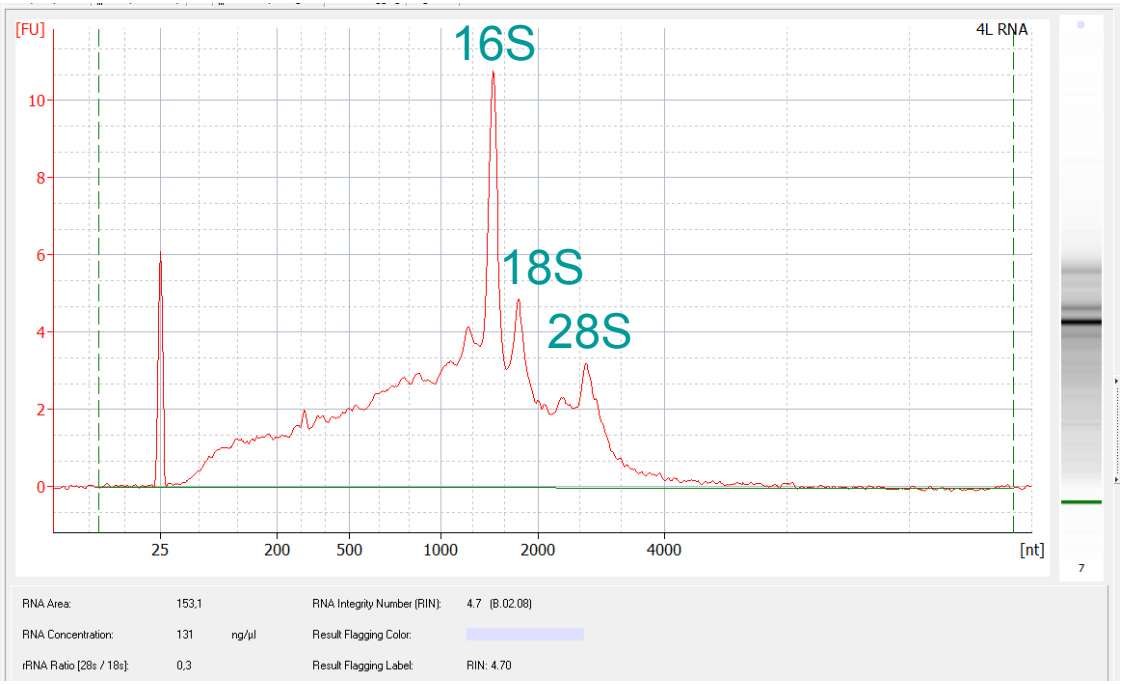


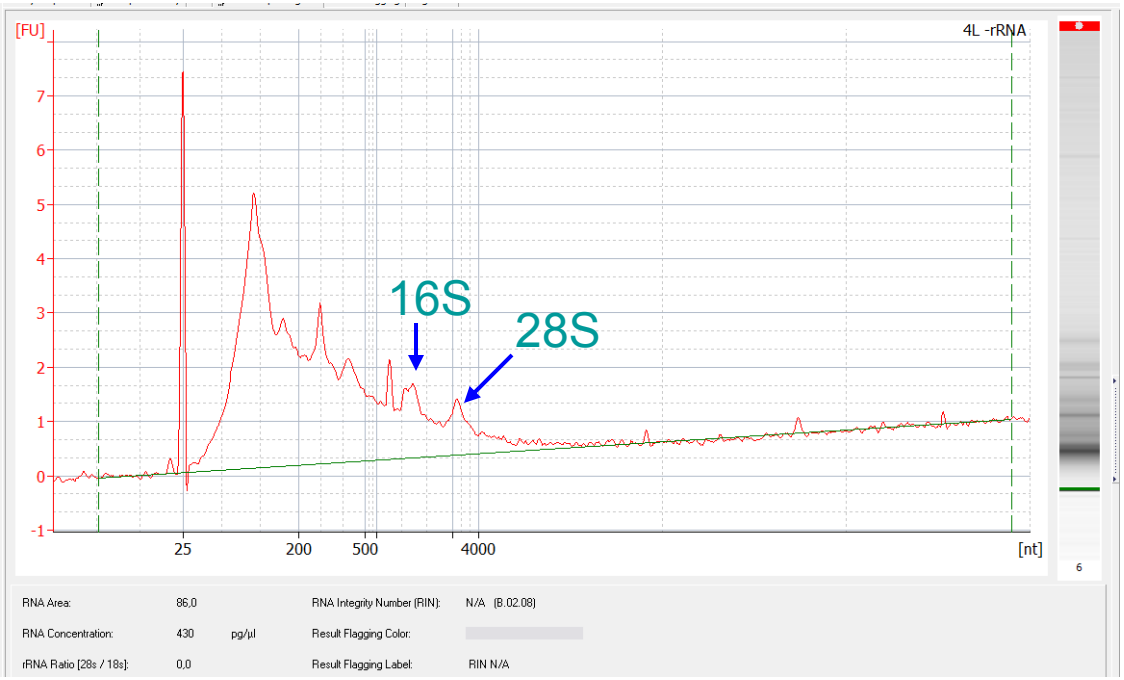
Removal of ribosomal RNA

- use of Epicentre RiboZero kits

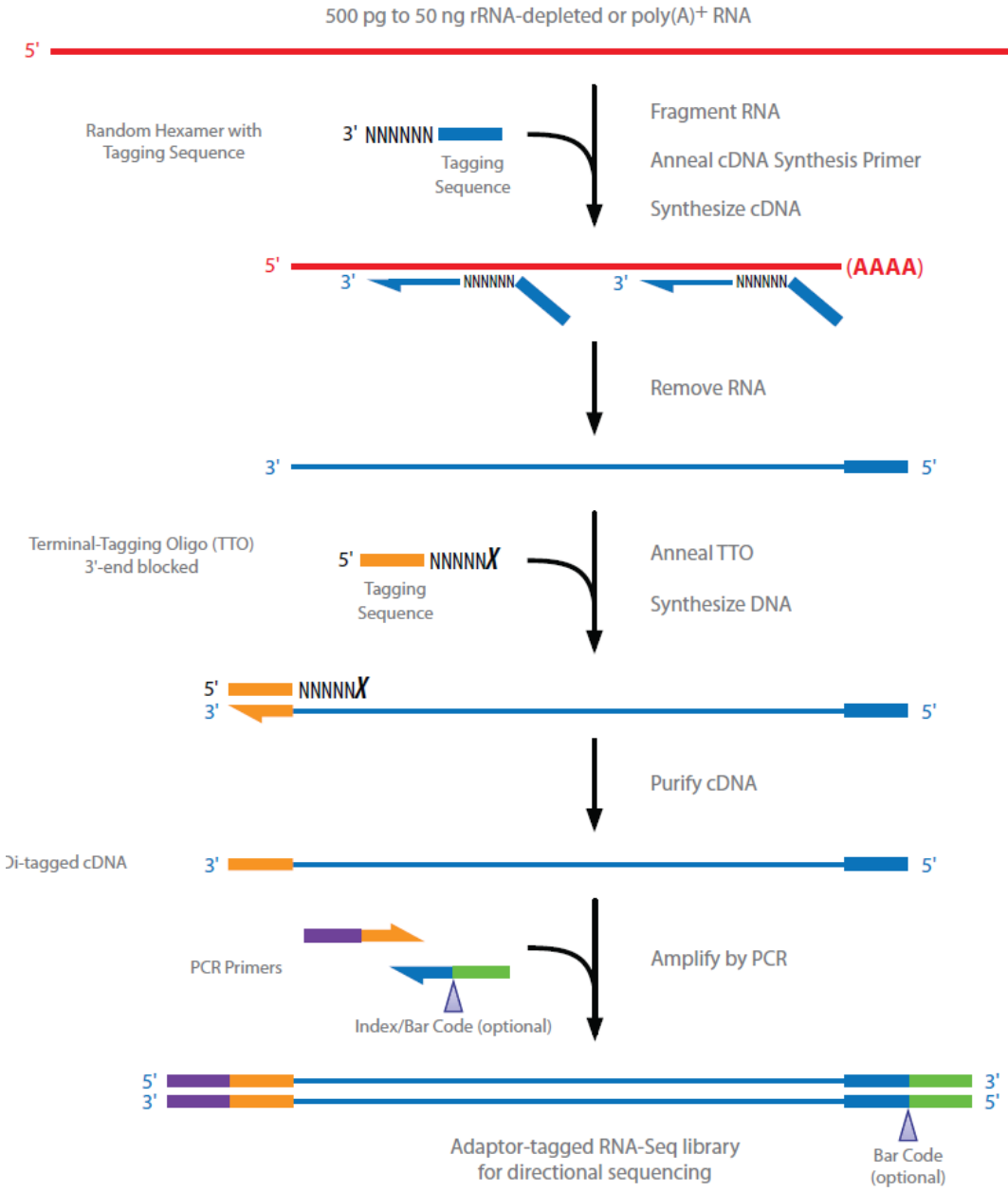
- to remove both bacterial and eukaryotic rRNA, RiboZero Metabacterial and RiboZero Human/Mouse/Rat kits were combined

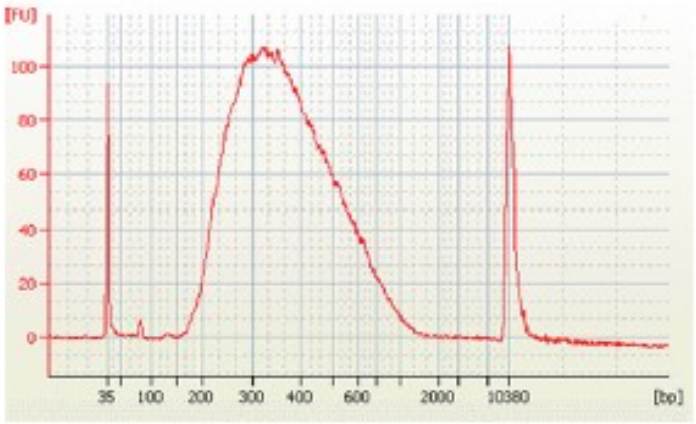
- efficiency of rRNA removal verified on Bioanalyzer





Library preparation: ScriptSeq™ v2 RNA-Seq Kit (Epicentre)





500 pg RNA

Library preparation

- RNA fragmentation to get desired size distribution of molecules

- cDNA synthesis (reverse transcription); success check by amplification of 16S from cDNA

- RNA removal (RNAse treatment)

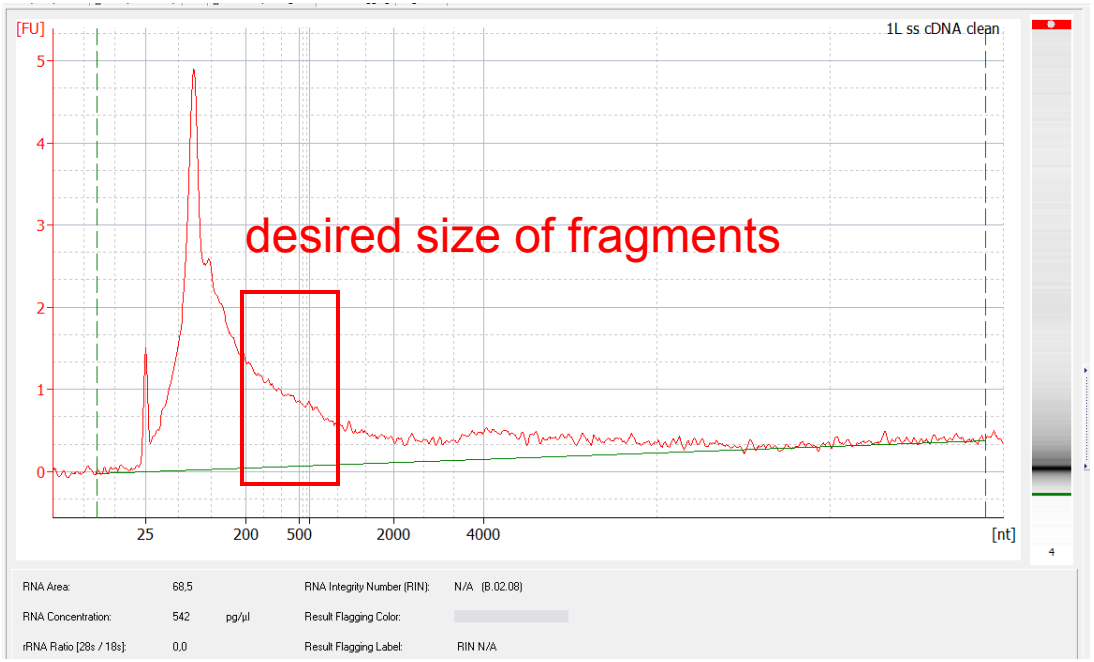
- Illumina adaptor ligation to cDNA

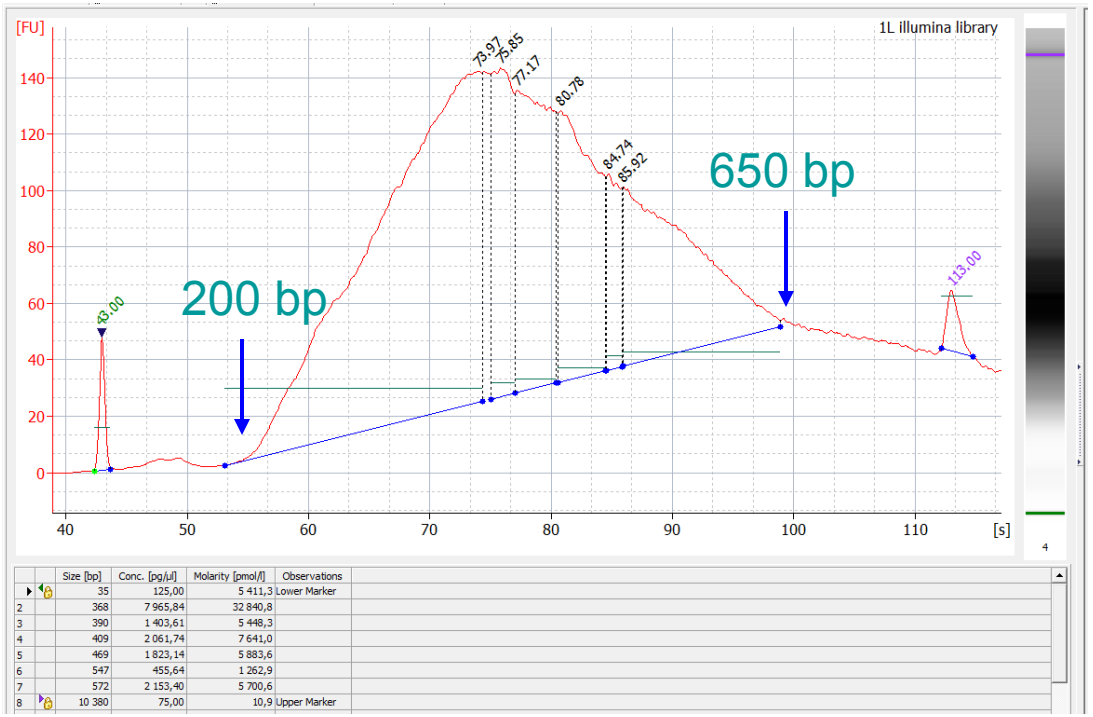
- amplification from Illumina adaptors with barcodes for each sample (10-15 cycles)

- check of DNA yield and size distribution

- removal of short fragments (Agencourt AMPure beads) and long fragments (cut from gel)

- equimolar combination of samples into one common library





Sequencing

Samples were sequenced on Illumina HiSeq (2x150 bases pair-end reads)

Library for one lane contained 12 samples (6x sample from litter + 6x sample from soil – summer); another lane for winter samples

Theoretically, one lane should deliver up to 350 millions pair-end sequences (some 30 millions per sample).

Data analysis and interpretation

Pair end 2x 150-base reads were used for assembly of contigs (scaffolds) – assembly (Velvet) performed externally, we do not have enough computing power at the moment

Contigs were taxonomically binned and functionally annotated (in MG Rast)

Individual sequence reads were mapped to contigs

Transcript abundance in each sample expressed as coverage per base = sequence count x sequence length/ contig length

The seasonality project

Analysis of microbial activity in summer and winter

Sampling

6 sites x 2 horizons (litter, soil) x 2 seasons (September, March) = 24 samples

Community analysis

Amplicon sequencing of DNA and RNA-deriver ITS2 sequences (MiSeq)

Metatranscriptomics: Shotgun sequencing of rRNA-depleted RNA

Isolation of total RNA
Deletion of bacterial rRNA and eukaryotic rRNA
(communities analysed by 16S and ITS sequening of DNA and RNA)

Sequencing on Illumina HiSeq – 2 lanes, 2x150 b *673 000 000 sequences*
Assembly of reads from all samples together *4 500 000 contigs >200 bases*
Annotation using MG-RAST and GenBank

44% reads mapped to contigs, 21% to identified contigs (taxon, function)

Metadata: Microbial biomass, enzyme activity, chemistry

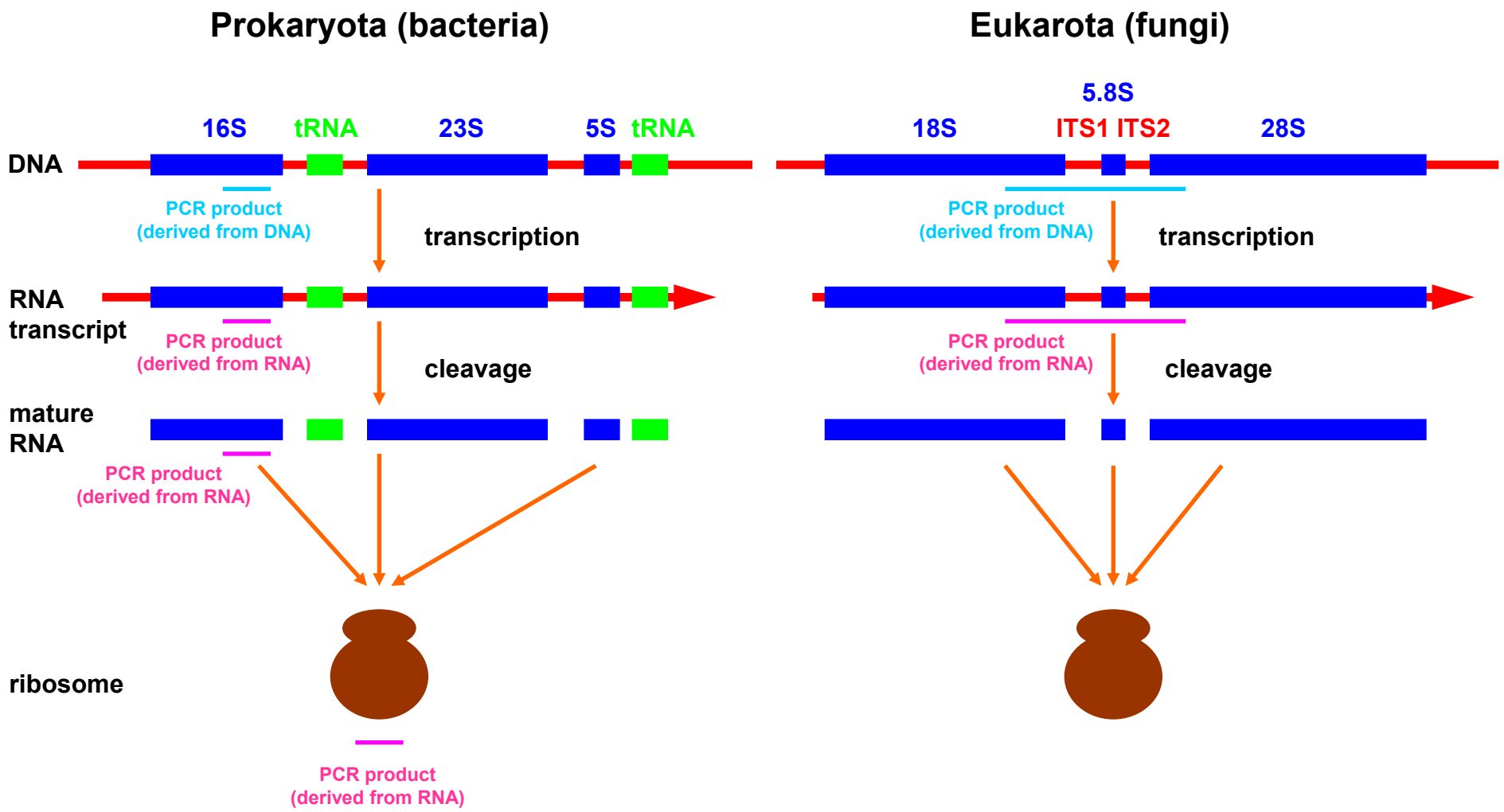
Metaproteomics: Identification of fungal / bacterial proteins

Sampling

September - soil temperature 15°C

March - soil temperature 2°C

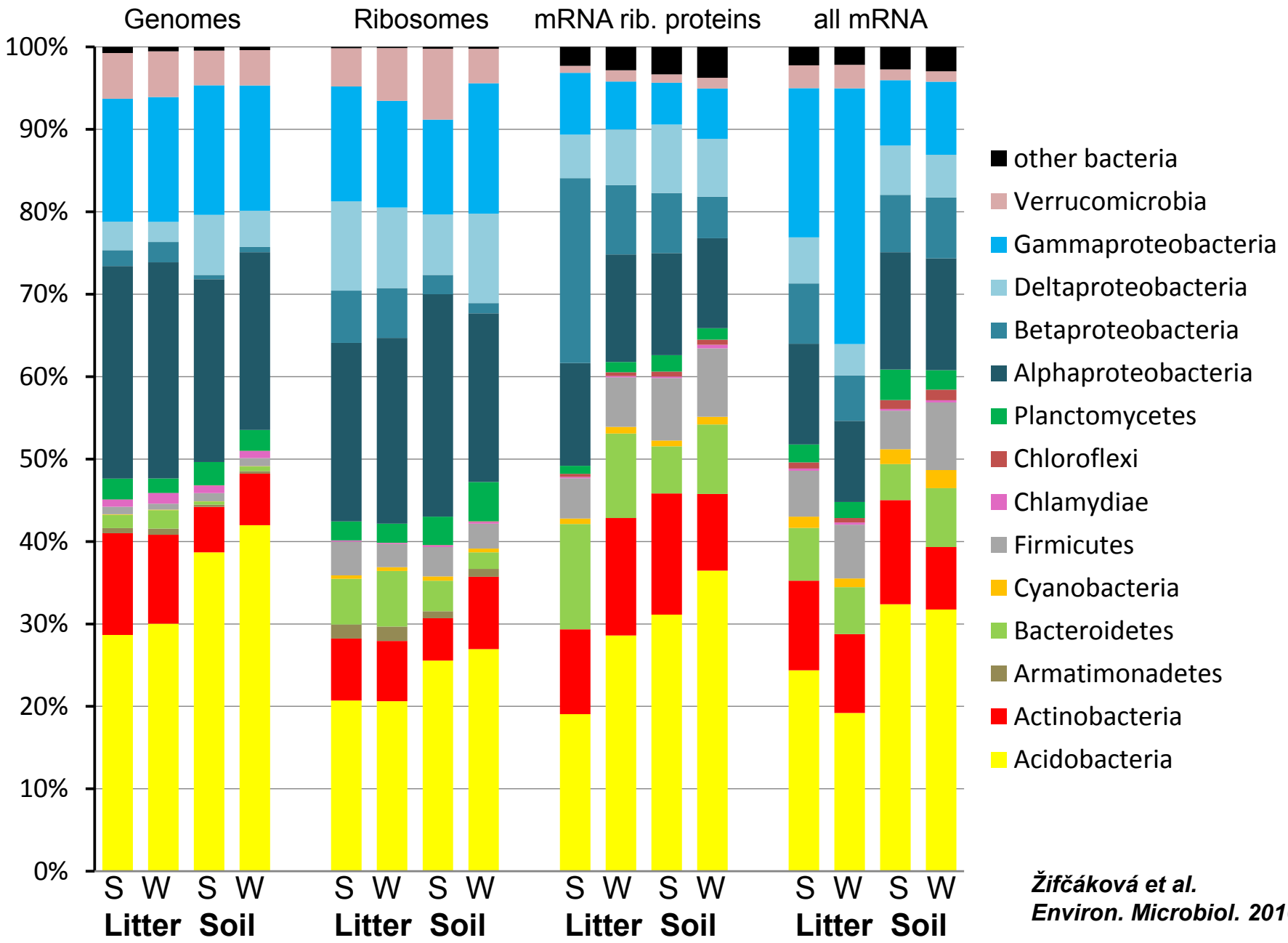
Identification of active microbes by 16S / ITS sequencing from RNA



RNA amplicons:
microorganisms possessing ribosomes

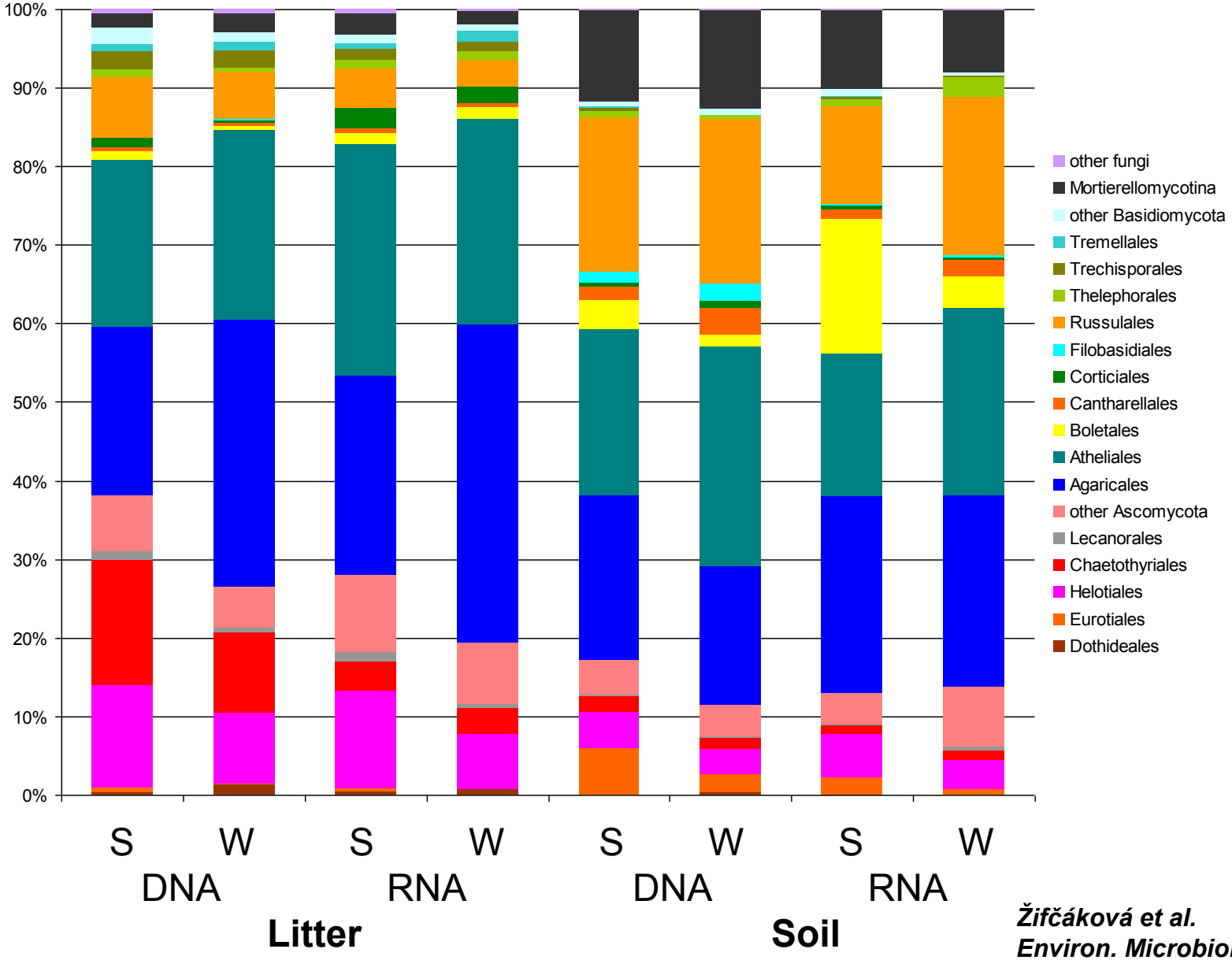
RNA amplicons:
microorganisms producing ribosomes

Community composition and activity of bacteria



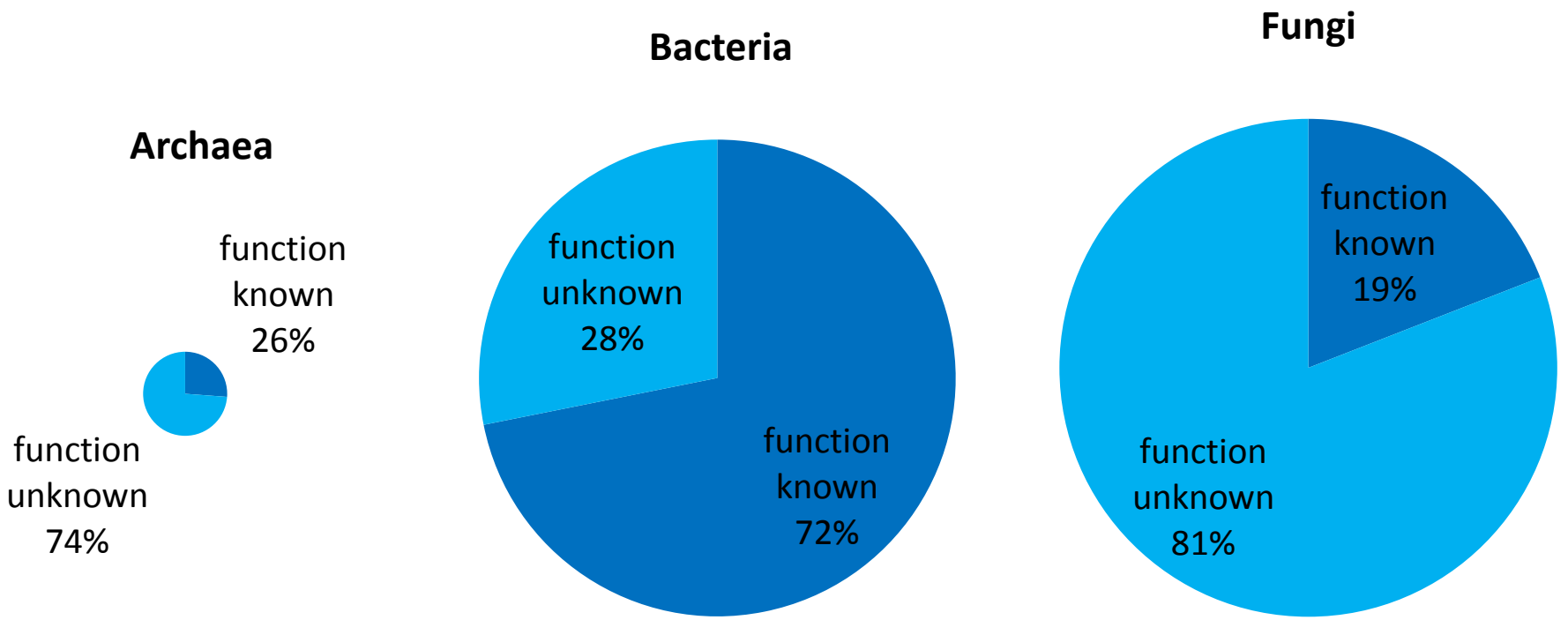
Žifčáková et al.
Environ. Microbiol. 2016

Composition of total (DNA) and active (RNA) communities of fungi



Žifčáková et al.
Environ. Microbiol. 2016

Exploring microbial activity: assignment of mRNA taxonomy and function



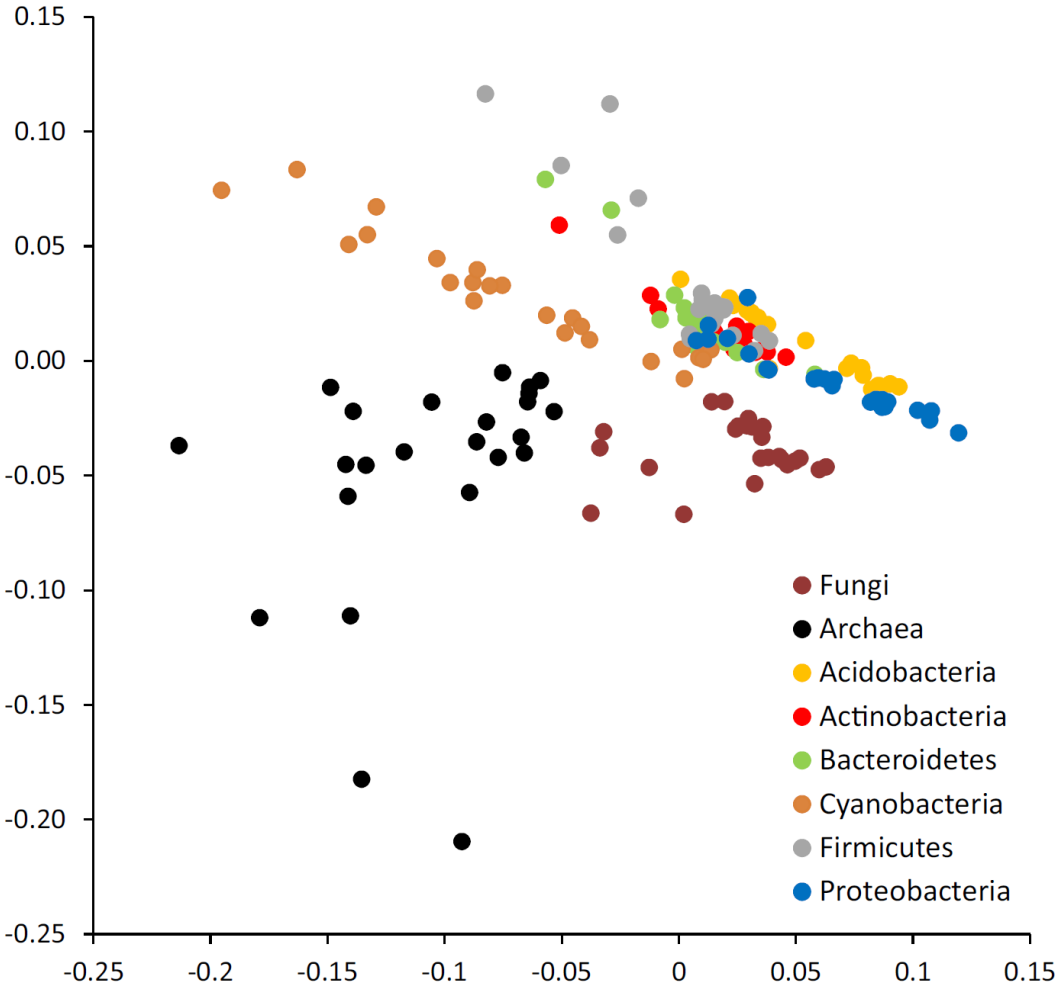
Functional annotation of predicted genes works well for bacteria but far less well for fungi and archaea. There, many hits are to „hypothetical proteins“.

The situation is even worse for nonmicrobial sequences (protozoa, invertebrates…)

Size of charts corresponds to numbers of transcripts.

Žifčáková et al. Environ. Microbiol. 2016

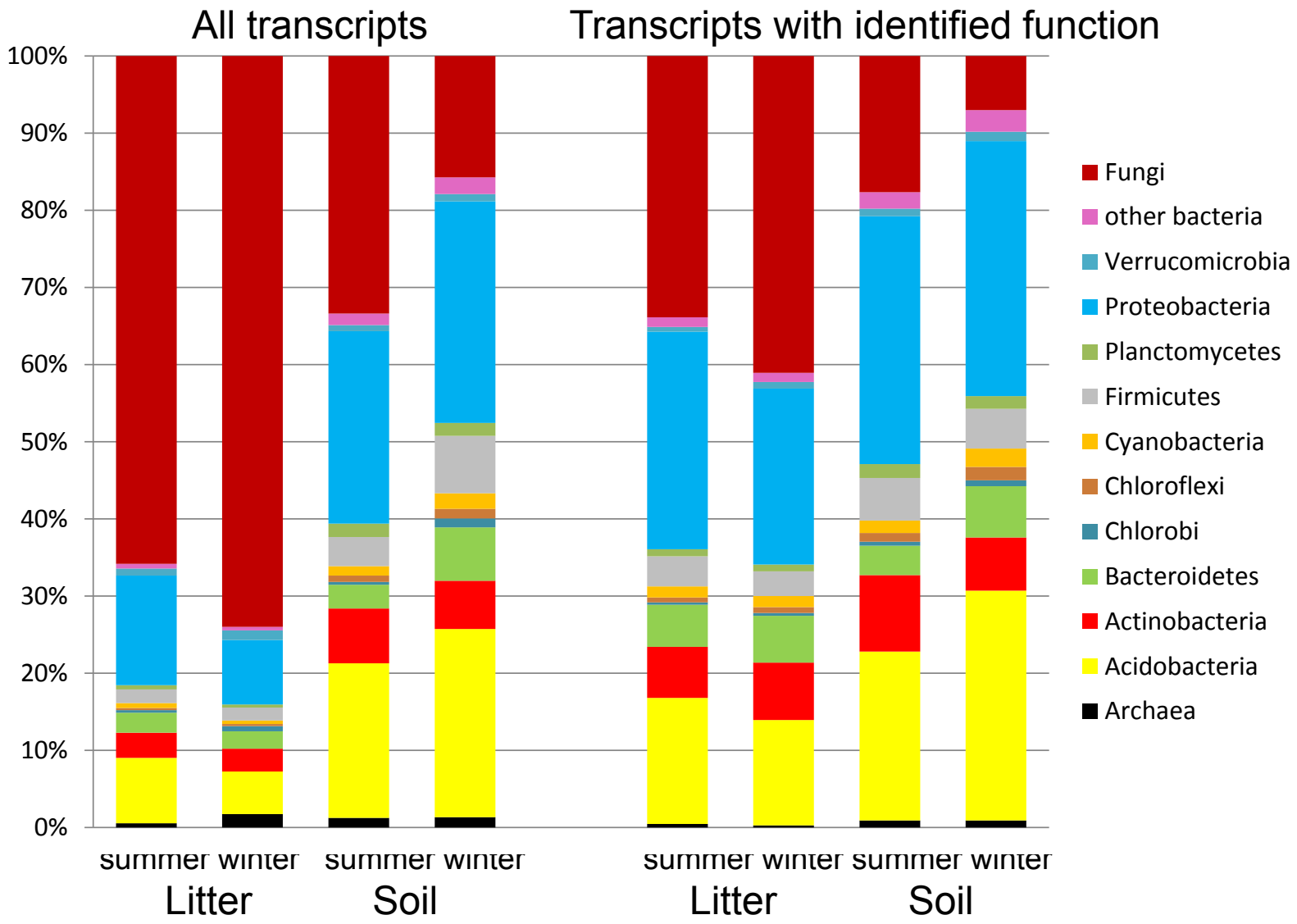
Exploring microbial activity: combining taxonomy and function



Bacterial (but not fungal) reads can be reliably identified on the level of phyla.

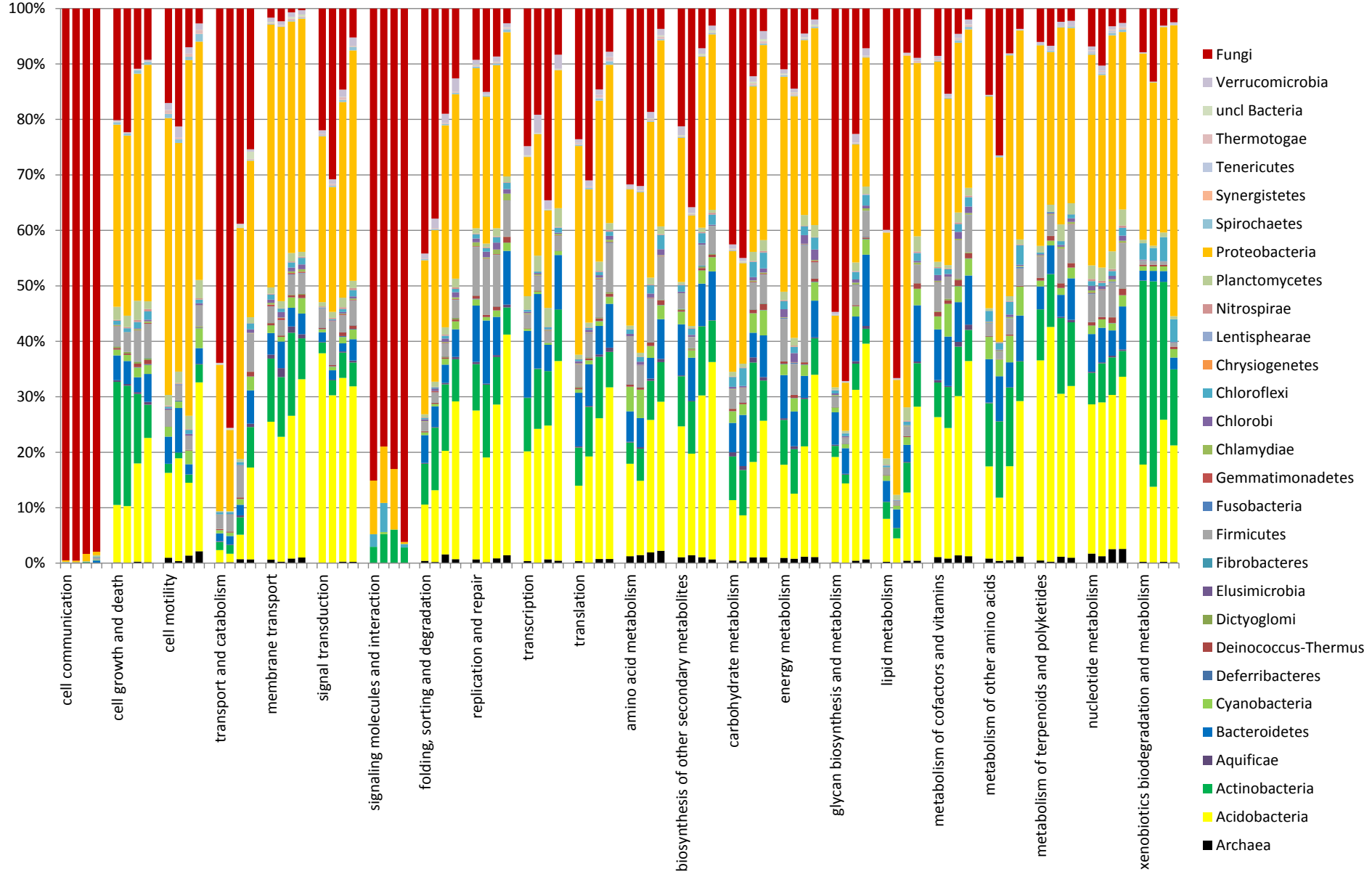
NMDS shows that profiles of functions of various microbial taxa differ

Seasonal contribution of microbial taxa to mRNA production



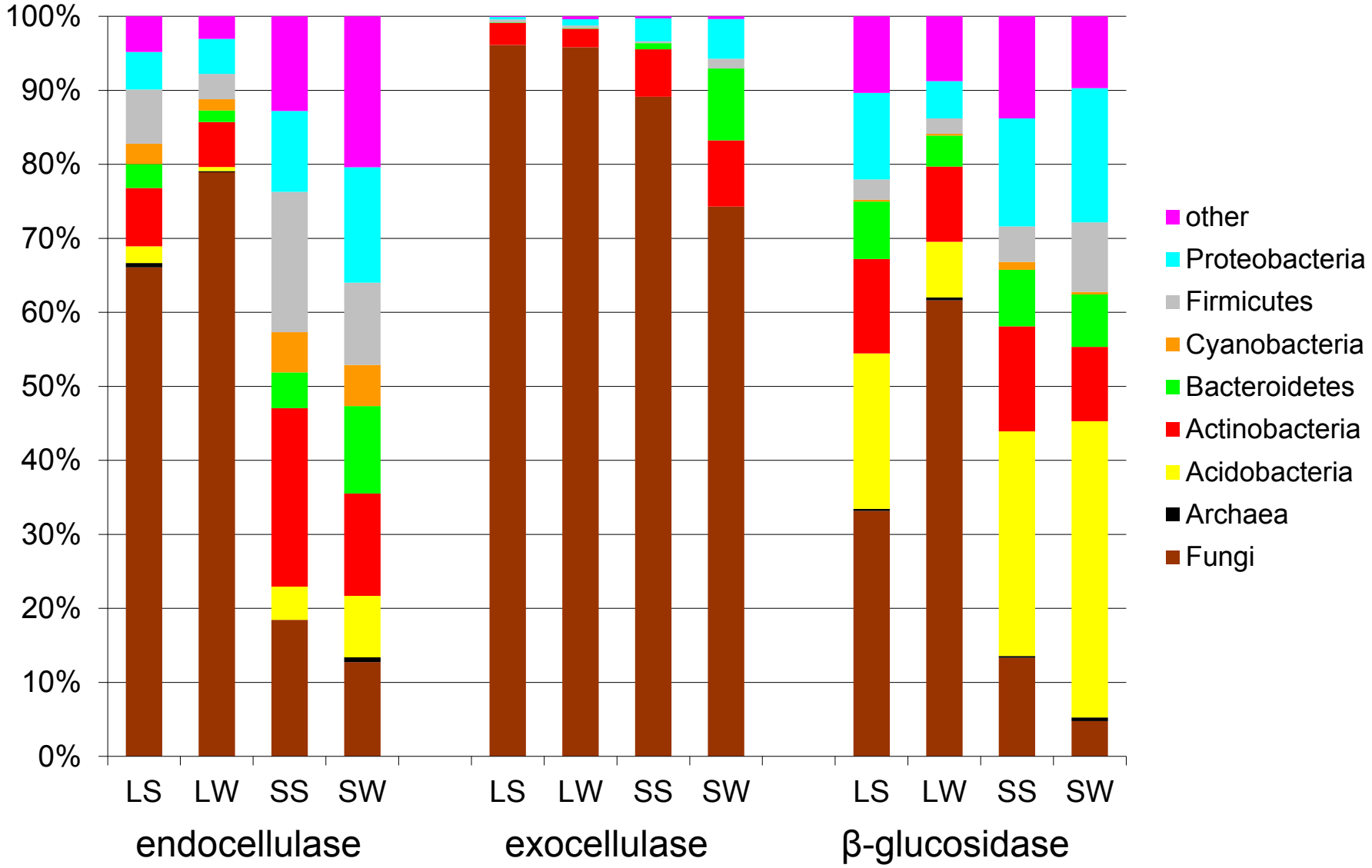
The share of fungal transcripts is higher in litter than in soil.
Many fungal reads are unidentified „hypothetical“ proteins.
In soil, fungal share of fungal transcripts dramatically decreases in winter.

Involvement of microbial taxa in soil processes



Contributions to activity in litter/summer, litter/winter, soil/summer and soil/winter

Involvement of taxa in the decomposition of cellulose



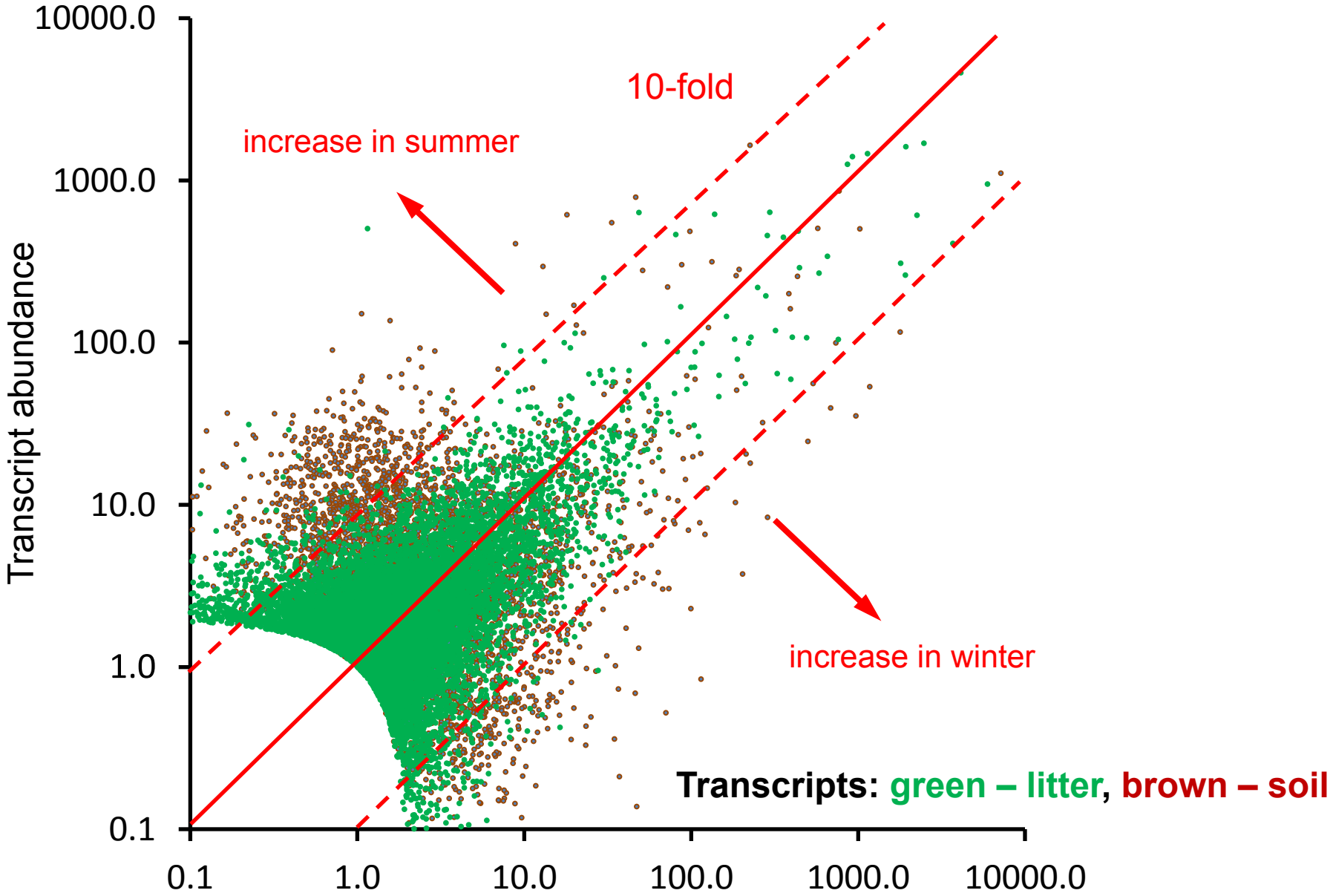
Fungi are dominant producers of cellulolytic enzymes in litter and important (but not dominant) producers in soil.

Functional biodiversity: high redundance of functions (starch and sucrose metabolism as an example)

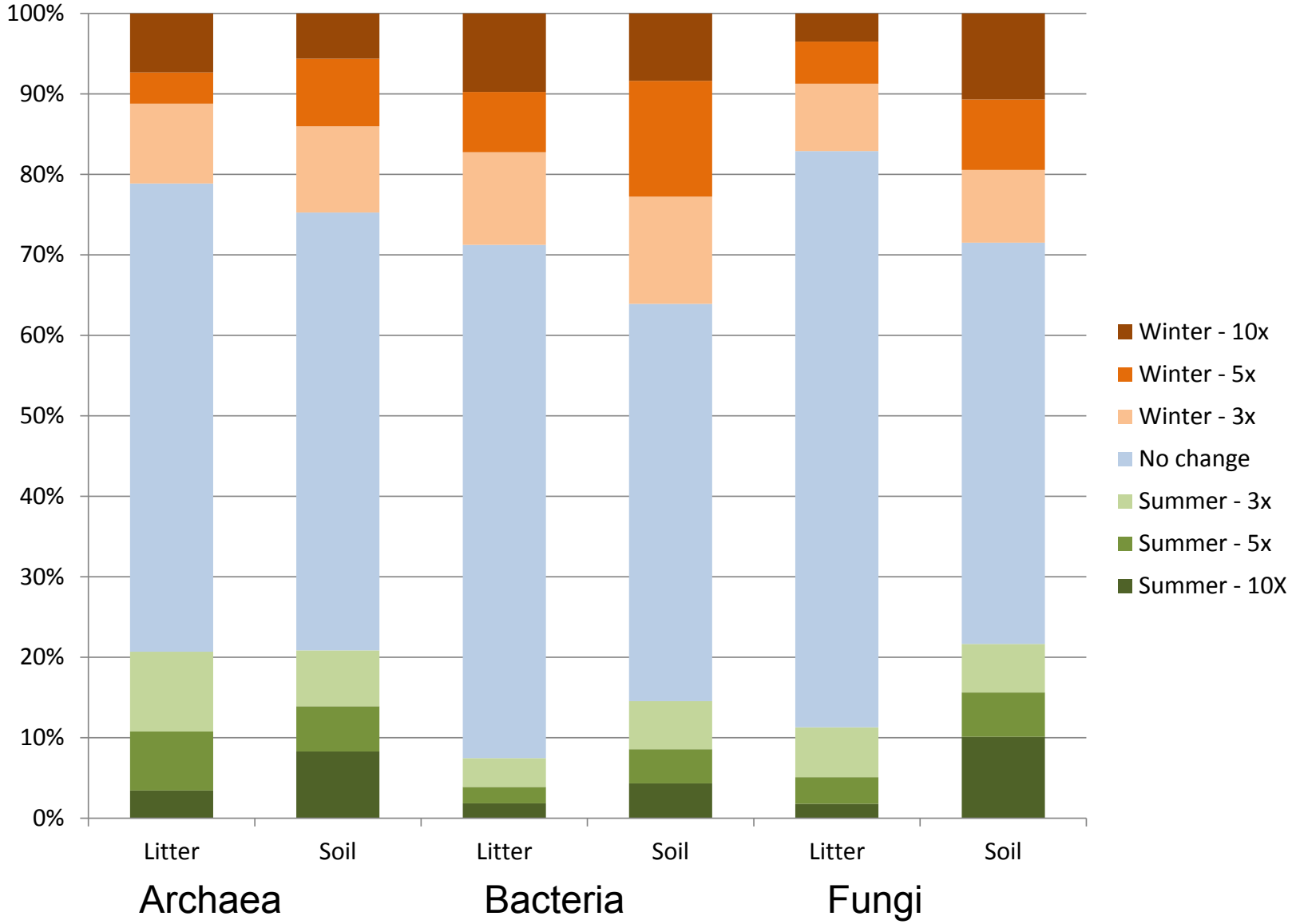
		fungi	bacteria
AGL	glycogen debranching enzyme	117	0
bcsA	cellulose synthase (UDP-forming)	0	145
E2.4.1.1	starch phosphorylase	191	1023
E2.4.1.20	cellobiose phosphorylase	0	11
E2.4.1.21	starch synthase	0	284
E2.4.1.34	1,3-beta-glucan synthase	463	0
E3.2.1.4	endoglucanase	136	204
glgB	1,4-alpha-glucan branching enzyme	132	666
glgC	glucose-1-phosphate adenylyltransferase	0	355
malQ	4-alpha-glucanotransferase	0	396
otsA	trehalose 6-phosphate synthase	213	511
rfbF	glucose-1-phosphate cytidylyltransferase	0	317
treS	maltose alpha-D-glucosyltransferase/ alpha-amylase	0	664

Numbers indicate transcript counts for each function.
(one species may produce one or more transcripts)

Seasonal changes in fungal expression are more intensive in soil



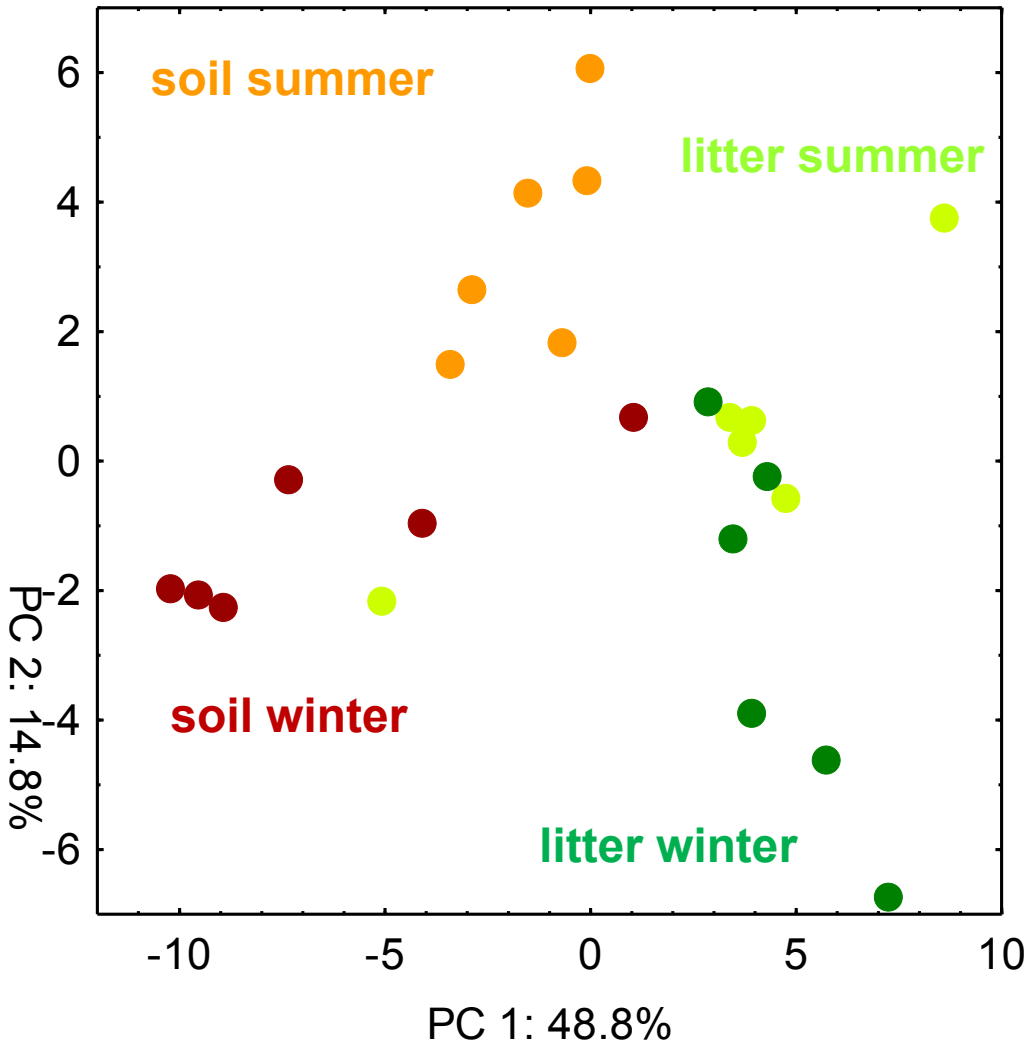
Seasonal changes in expression



29-51% of dominant transcripts show seasonal changes in relative abundance

Žifčáková et al. Environ. Microbiol. 2016

Fungal transcription profiles



Fungal transcription in soil

Increased in summer:

Spliceosome components
Ribosome components
Phenylalanine metabolism

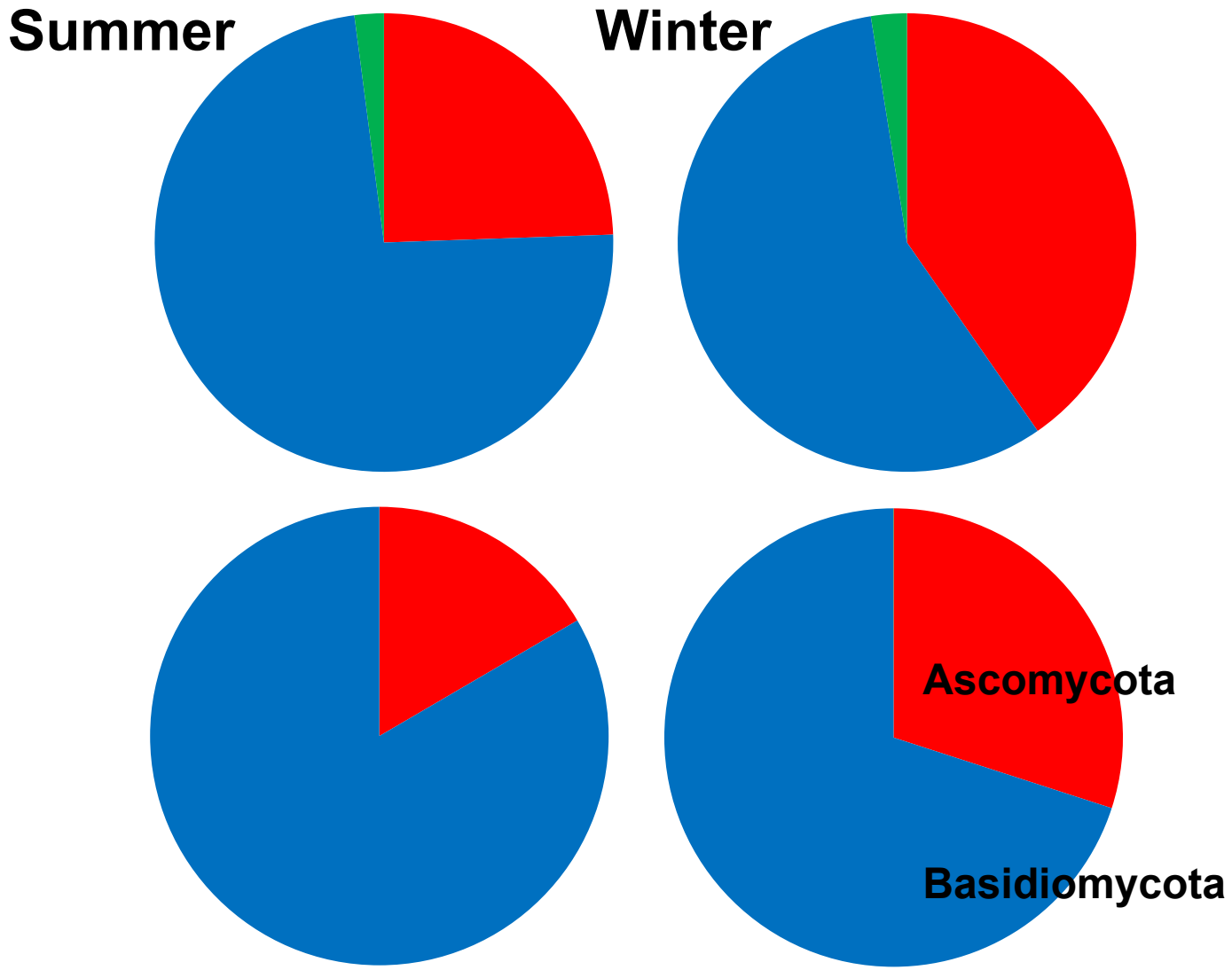
Increased in winter:

Fatty acid biosynthesis
DNA repair
RNA degradation
Amino acid metabolism
Yeast-specific cell processes

Transcription profiles differ among litter and soil.

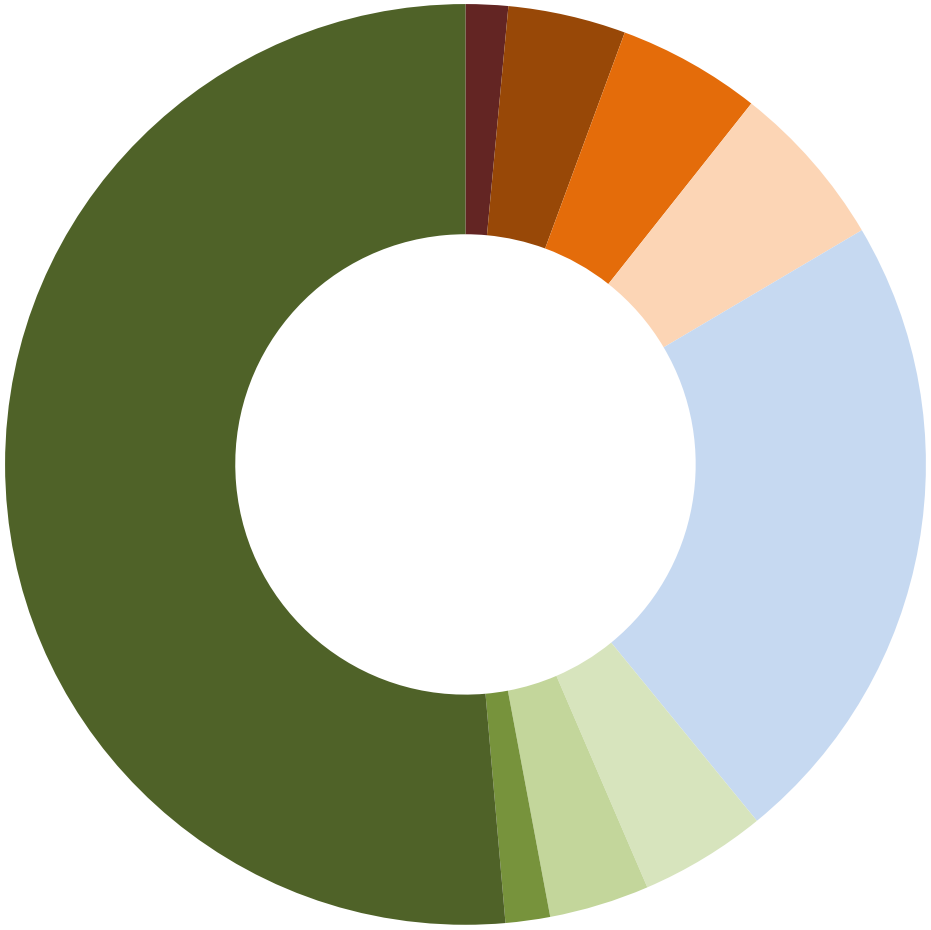
Seasonal pattern of transcription is more apparent in soil than in litter.

Seasonal activity of fungal divisions: transcription of beta-tubulin



Beta tubulin can be reasonably assigned to fungal phyla.
Relative share of transcripts of Basidiomycota (most ectomycorrhizal fungi) decreases in winter.

Fungal genes involved in mycorrhizal symbiosis



Transcripts with high similarity (tblastx, E < 10^{-50})to Lacccaria laccata *genes involved in mycorrhizal symbiosis are more frequently transcribed in summer.*

Over 50% of them are exclusively transcribed in summer.

Exclusively transcribed in summer in winter
Transcription increased in summer in winter
No seasonal difference in trancription

Functional biodiversity: high redundance of functions (starch and sucrose metabolism as an example)

		fungi	bacteria
AGL	glycogen debranching enzyme	117	0
bcsA	cellulose synthase (UDP-forming)	0	145
E2.4.1.1	starch phosphorylase	191	1023
E2.4.1.20	cellobiose phosphorylase	0	11
E2.4.1.21	starch synthase	0	284
E2.4.1.34	1,3-beta-glucan synthase	463	0
E3.2.1.4	endoglucanase	136	204
glgB	1,4-alpha-glucan branching enzyme	132	666
glgC	glucose-1-phosphate adenylyltransferase	0	355
malQ	4-alpha-glucanotransferase	0	396
otsA	trehalose 6-phosphate synthase	213	511
rfbF	glucose-1-phosphate cytidylyltransferase	0	317
treS	maltose alpha-D-glucosyltransferase/ alpha-amylase	0	664

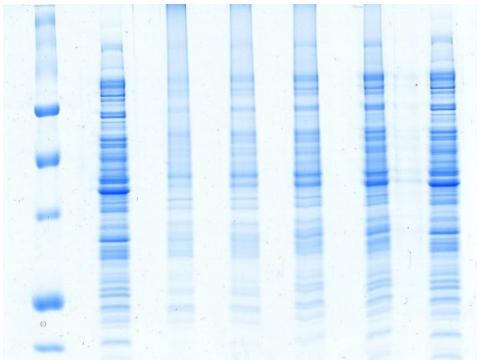
Numbers indicate transcript counts for each function.
(one species may produce one or more transcripts)

Metaproteomics

Analysis of microbial proteins by metaproteomics

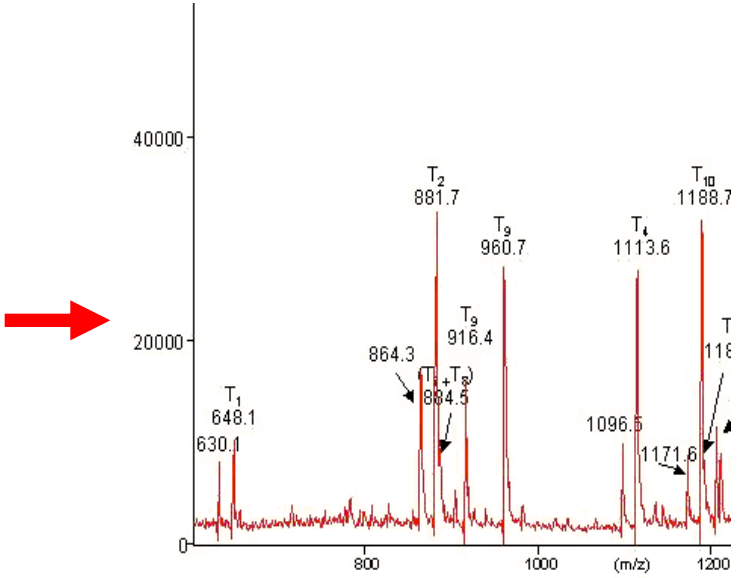
Extraction of proteins

Separation and fragmentation

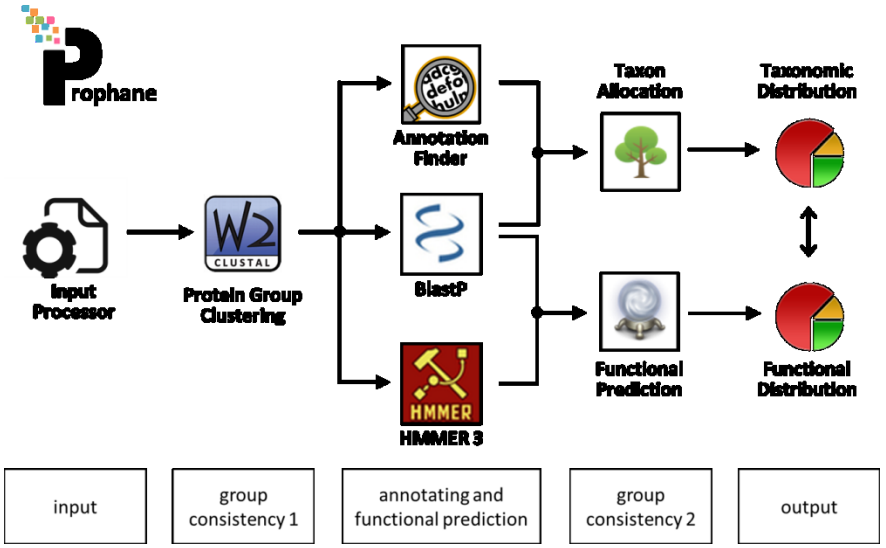


Mass spectrometry

Identification of peptide fragments



Functional and taxonomic annotation



Illustrationsby Katharina Riedel and Stephan Fuchs, Greifswald

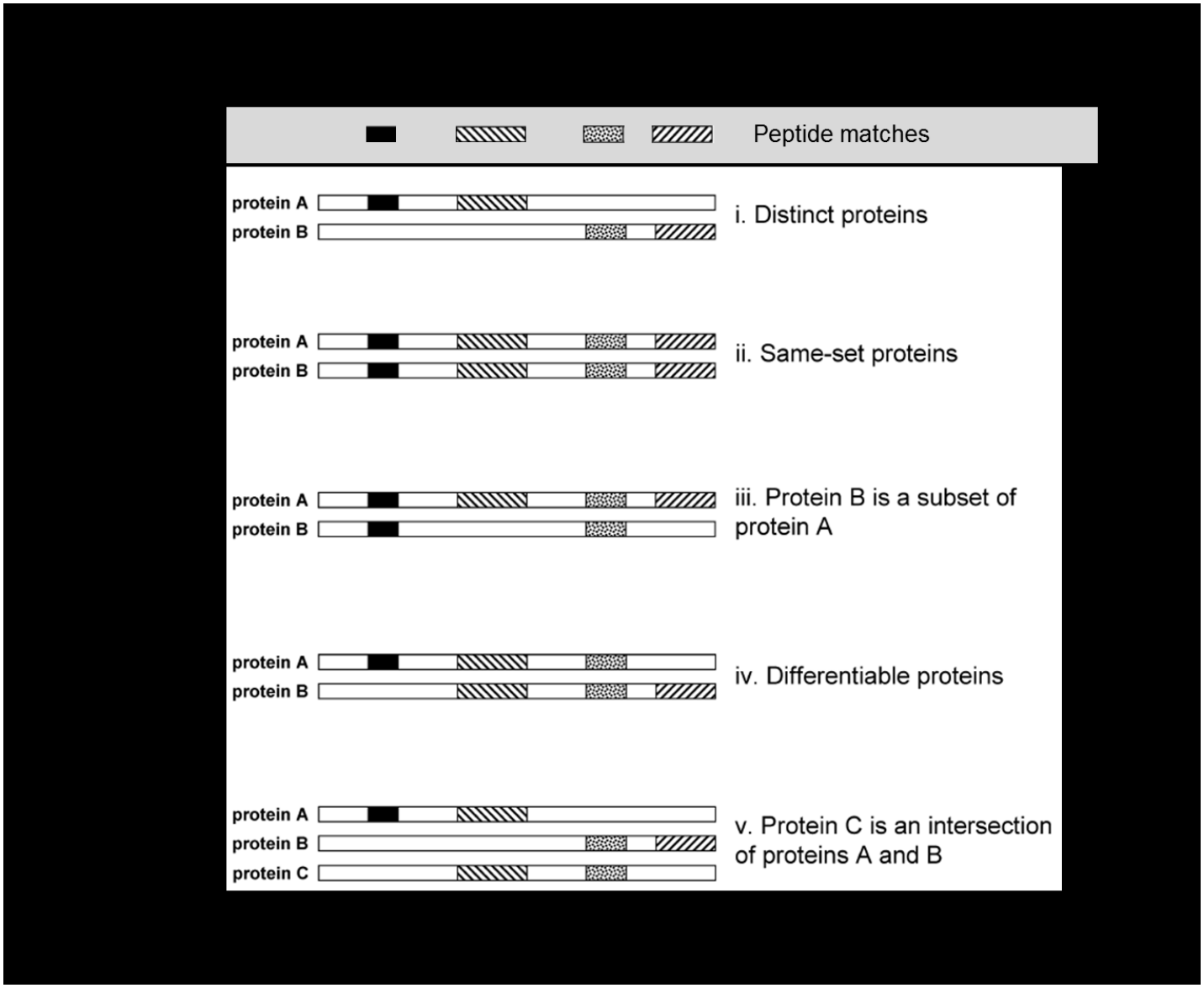
Metatproteomics - opportunities

- measures proteins produced, not genes transcribed (better proxy for function)
- proteins can be stable, so that the picture of metaproteomics represents a longer period of time (but how long?)
- protein extraction from litter or soil is no more a technical problem

Metatproteomics - limitations

- difficult annotation (complex metatranscriptomes can be only annotated when metatranscriptome or metagenome assembly is available)
- sequencing of peptides, not proteins, unclear which peptides originate from the same protein
- only a fraction of proteins is typically annotated
- annotation is extremely computationally demanding
- some proteins can be too much stable
- high costs (>1000 EUR for one metatranscriptome sample) force into shallow replication

Metaproteome annotation



Slide provided by Stephan Fuchs, Greifswald

Eukaryotic proteins are quantitatively important in soils

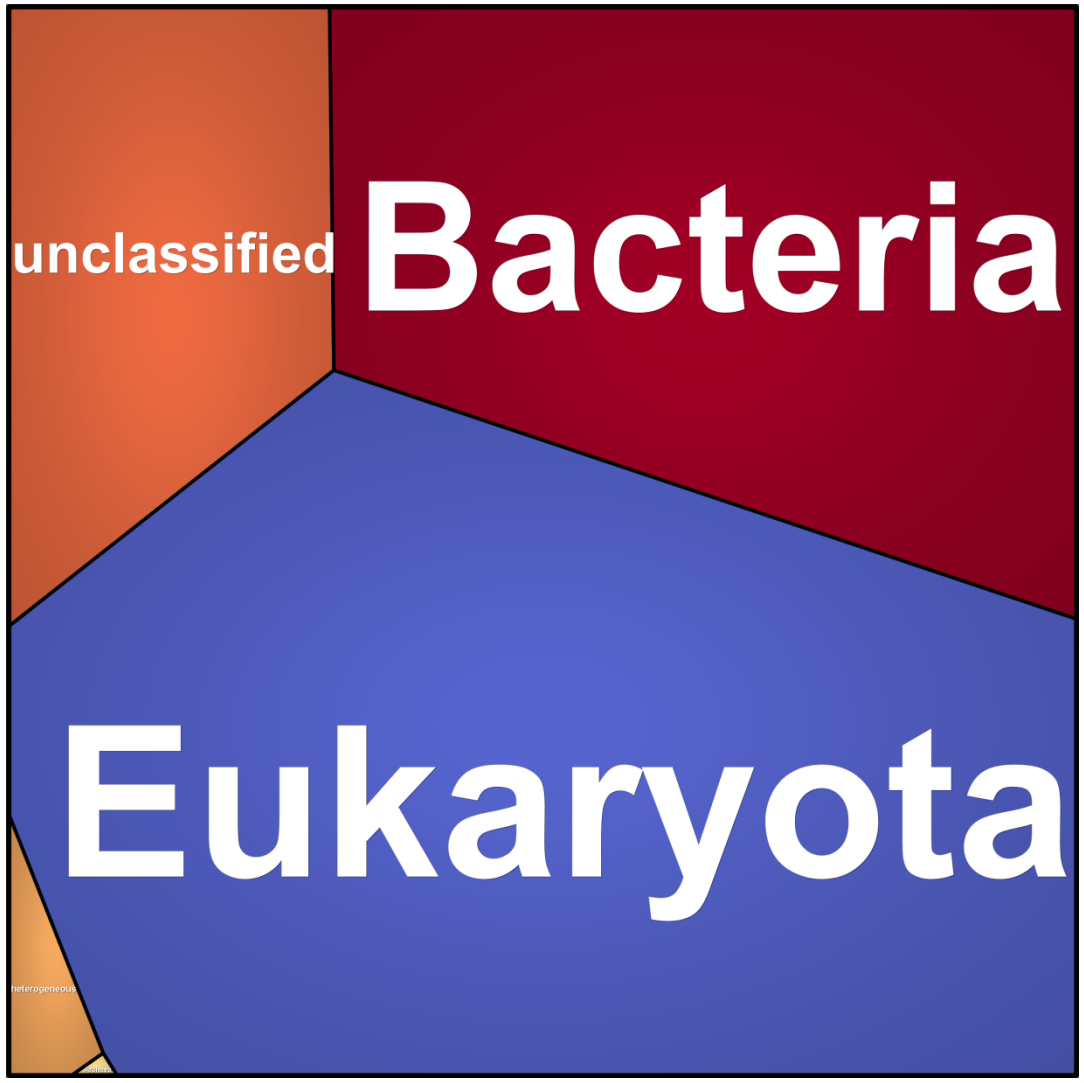


Illustration by Stephan Fuchs, Greifswald

The share of annotated fungal proteins in litter is comparable to those of bacteria

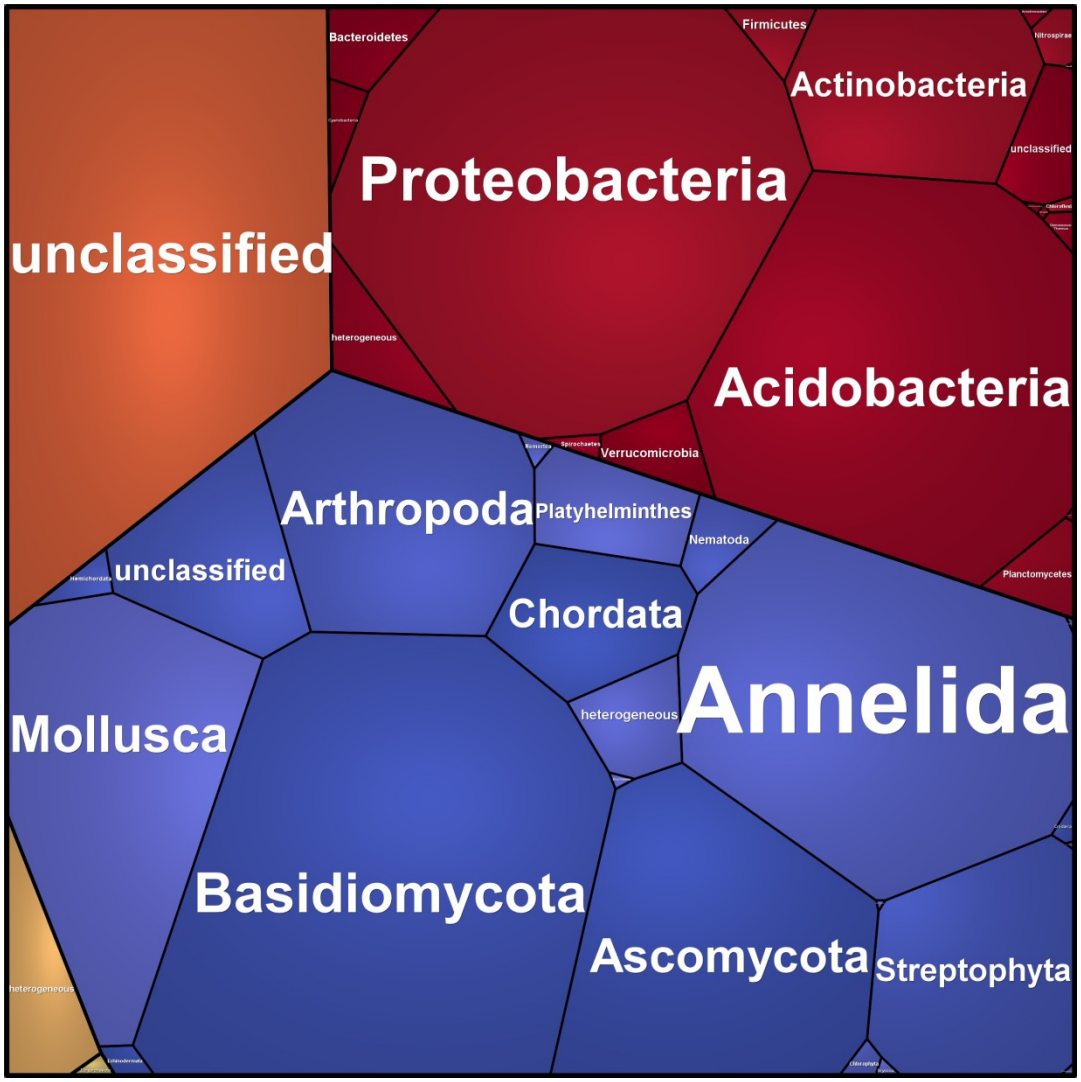
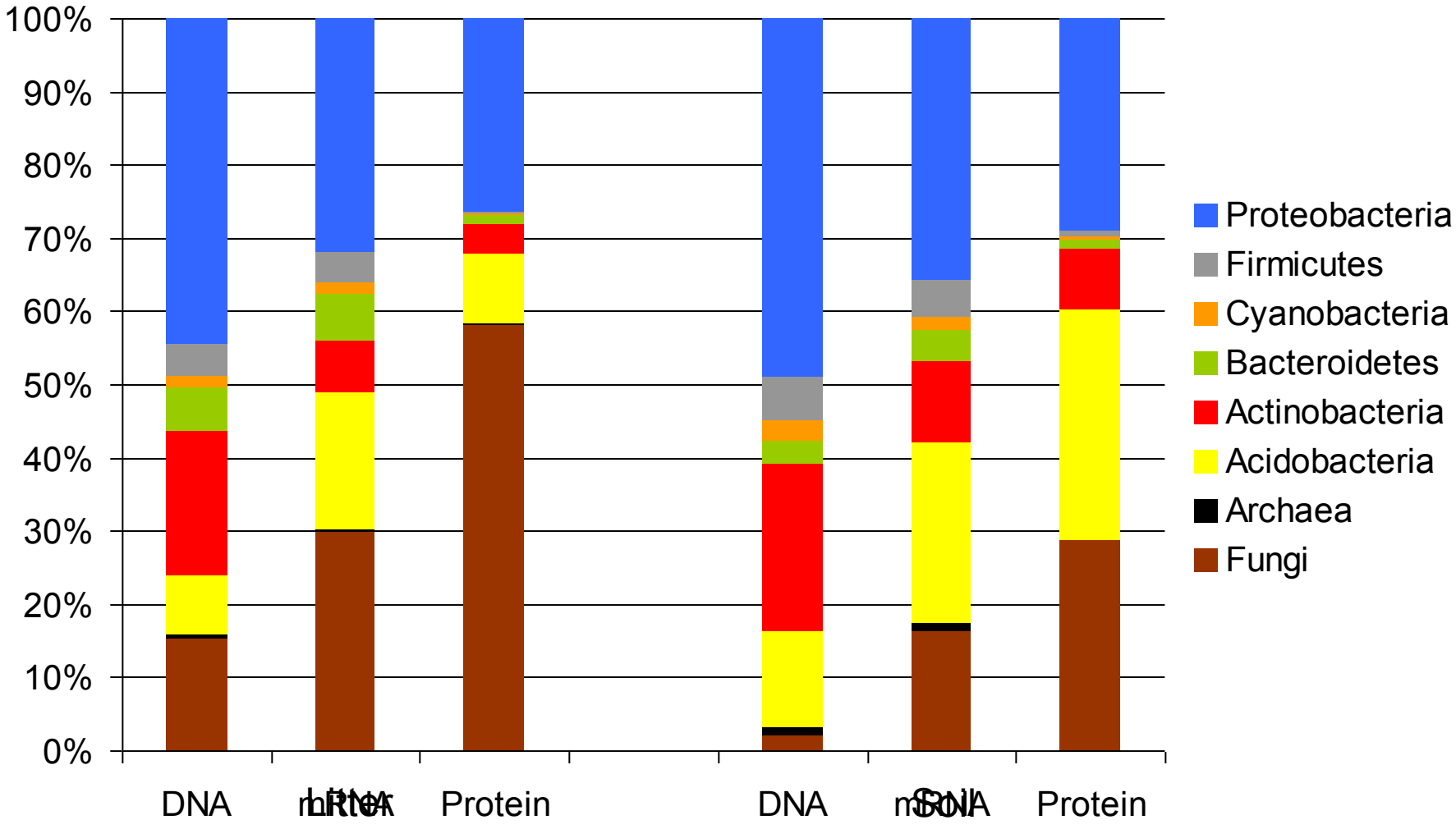


Illustration by Stephan Fuchs, Greifswald

Microbial community composition and activity as reflected in the metagenome, metatranscriptome and metaproteome (only annotated reads considered)



Methodological challenges:
annotation of eukaryotic genes in soils is challenging

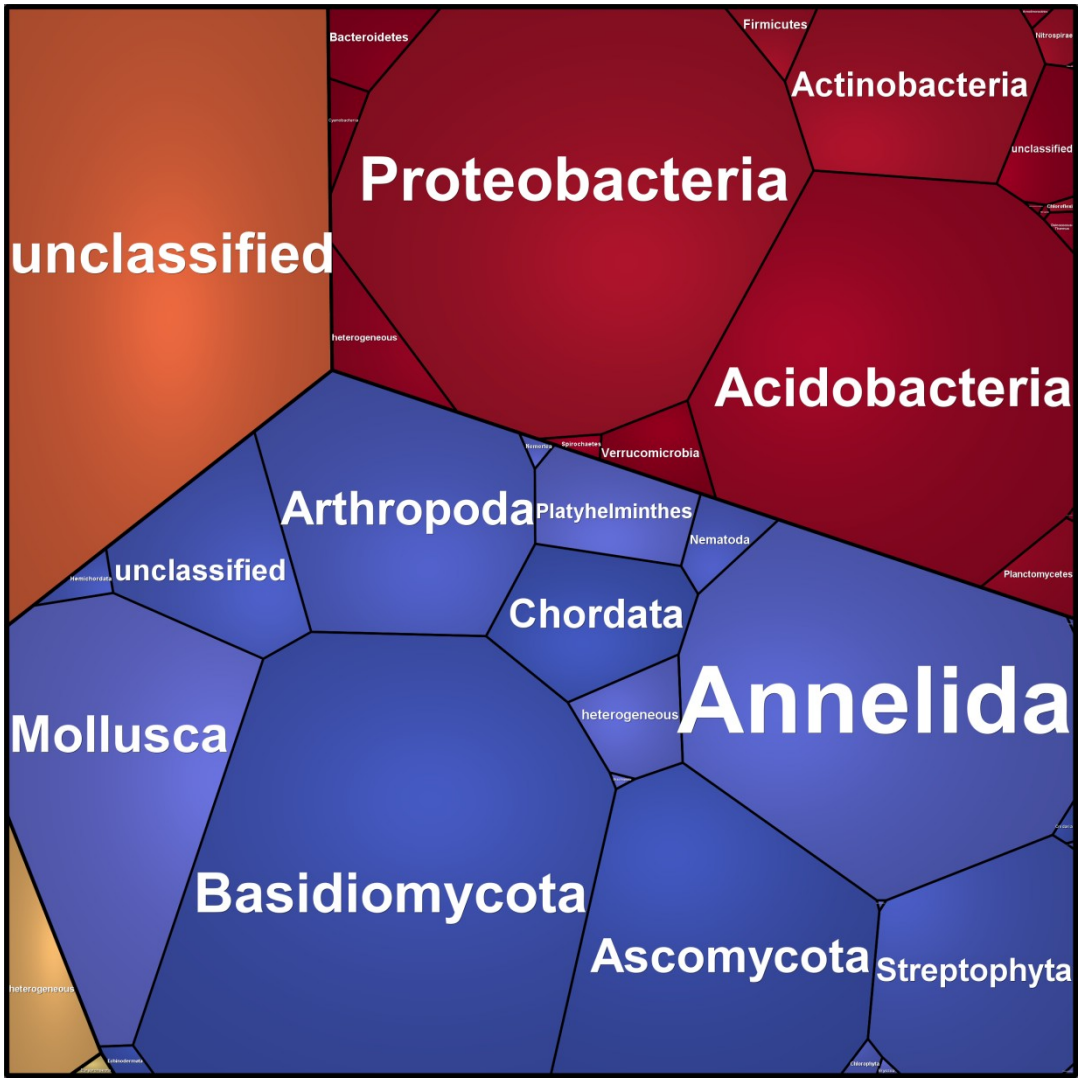


Illustration by Stephan Fuchs, Greifswald

Methodological challenges: annotation of eukaryotic genes in soils is challenging

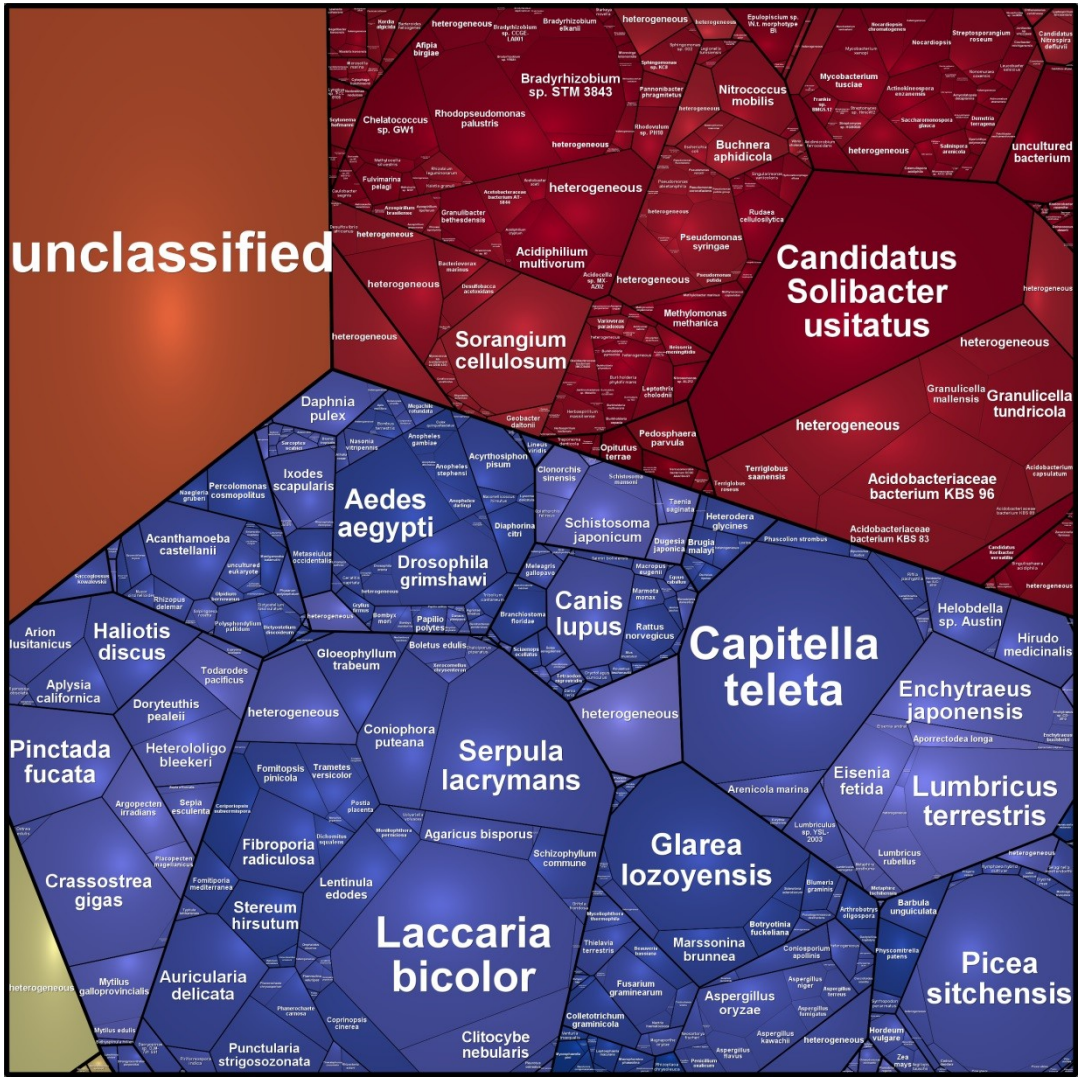


Illustration by Stephan Fuchs, Greifswald

Methodological challenges: annotation of eukaryotic genes in soils is challenging

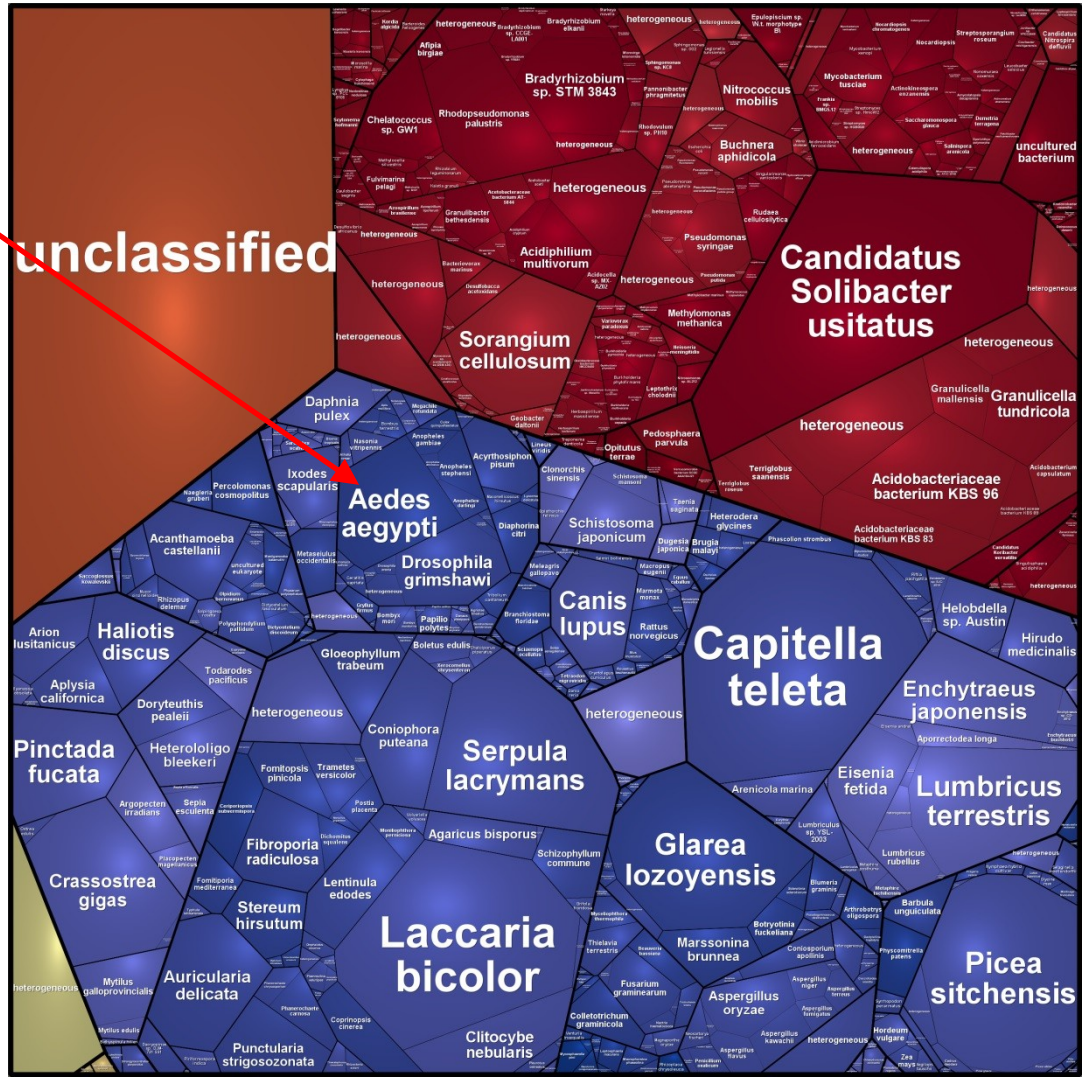


Illustration by Stephan Fuchs, Greifswald

Methodological challenges: annotation of eukaryotic genes in soils is challenging

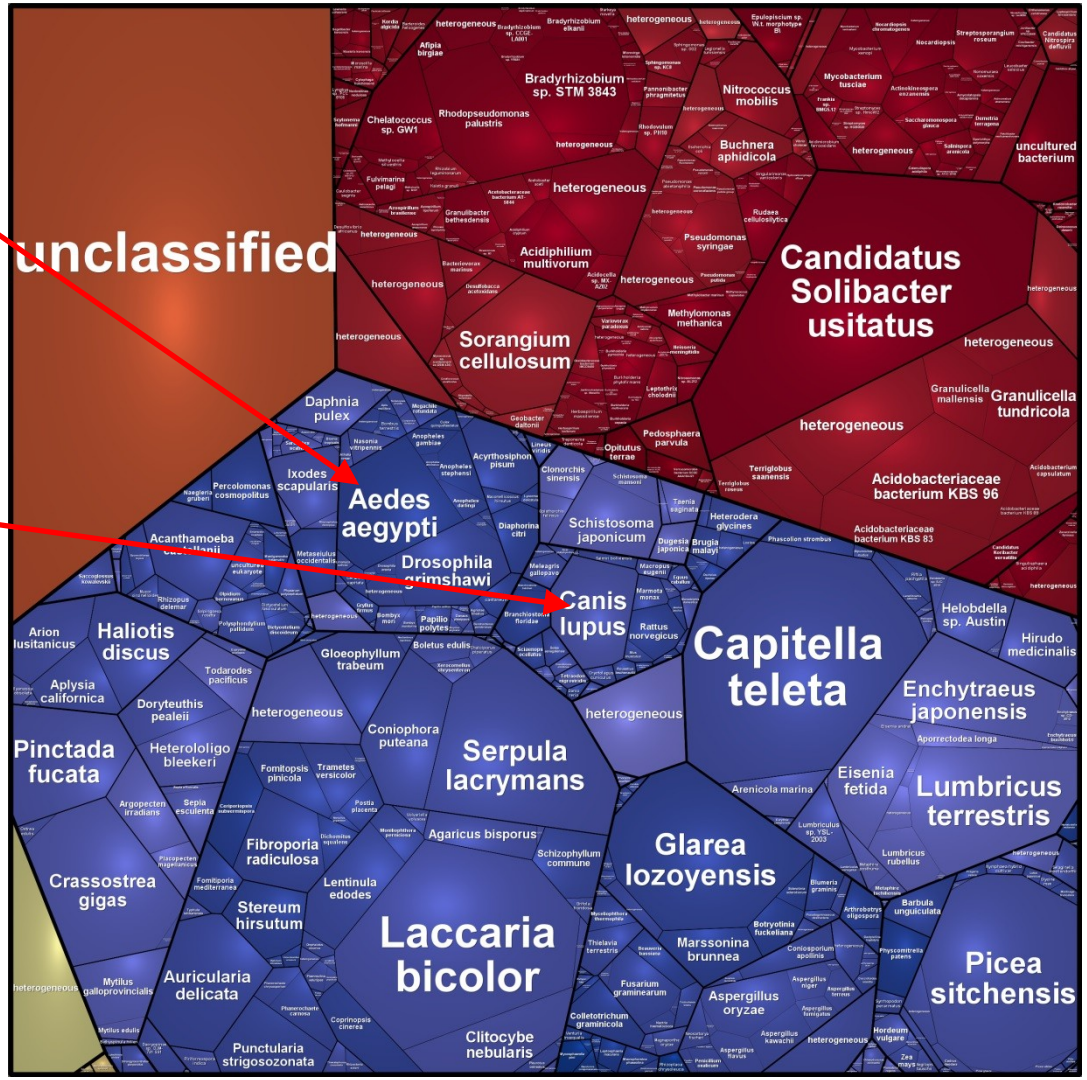


Illustration by Stephan Fuchs, Greifswald

Other methods to detect microbial activity

Stable Isotope Probing (SIP)

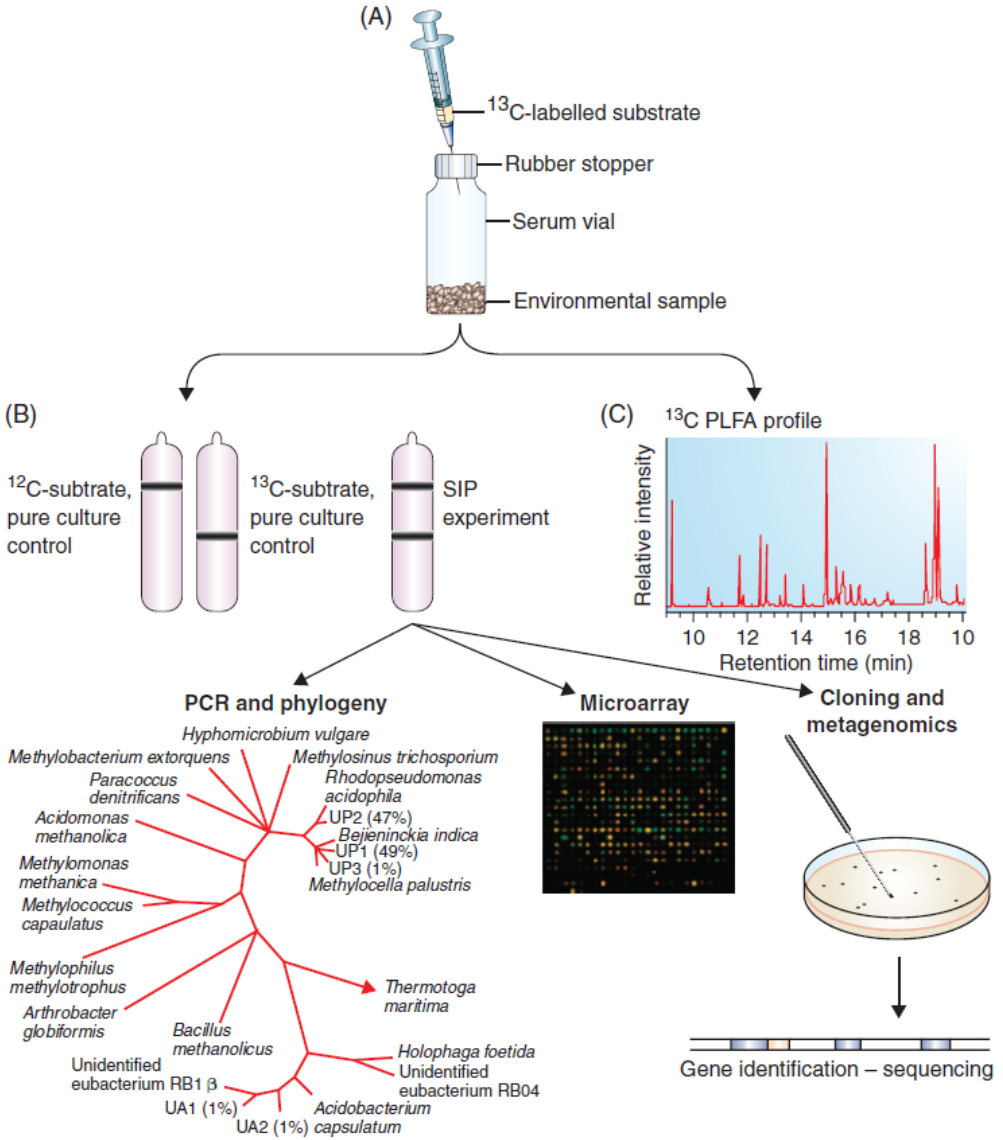


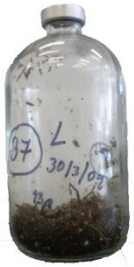
FIGURE 4.8 Stable isotope probing procedure with associated assays for assessing incorporation of isotopically labeled substrate into cellular constituents and determining the sequence(s) of bacteria that have incorporated the label (Dumont and Murrell, 2005).

Po přidání substrátu, značeného stabilním isotopem ^{13}C, se značený isotop akumuluje v biomase aktivně metabolizujících mikroorganismů. Složení „aktivního“ společenstva lze pak vyhodnotit při analýze lipidů anebo – po separaci „těžkých“ (značených) a „lehkých“ molekul DNA pomocí molekulárních metod.

Separace nukleových kyselin mezi ^{12}C a ^{13}C není obvykle v praxi úplně bezproblémová, protože část mikroorganizmů je značena pouze částečně. Proto mezi frakcemi s ^{12}C a ^{13}C obvykle bývá postupný přechod.

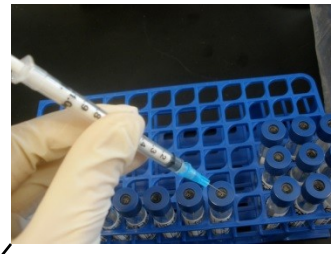
Stable Isotope Probing (SIP)

13C-značená celulóza



Mikrobiální společenstvo

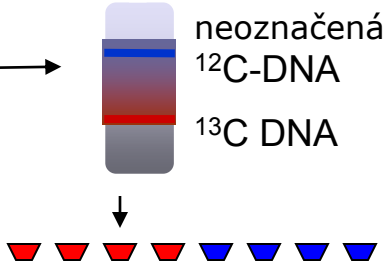
Vzorek CO_2



Podíl ^{13}C-CO_2
Celková respirace

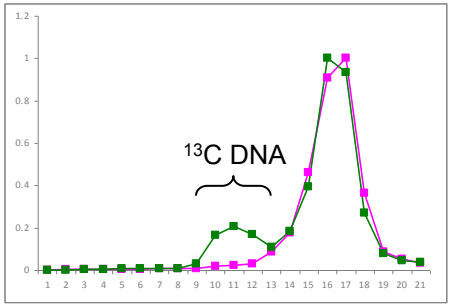
DNA extrakce

Rozdělení DNA ultracentrifugací na základě vznášivé hustoty



vysráženi DNA ve frakcích

identifikace frakcí v kterých je přítomna DNA pomocí qPCR



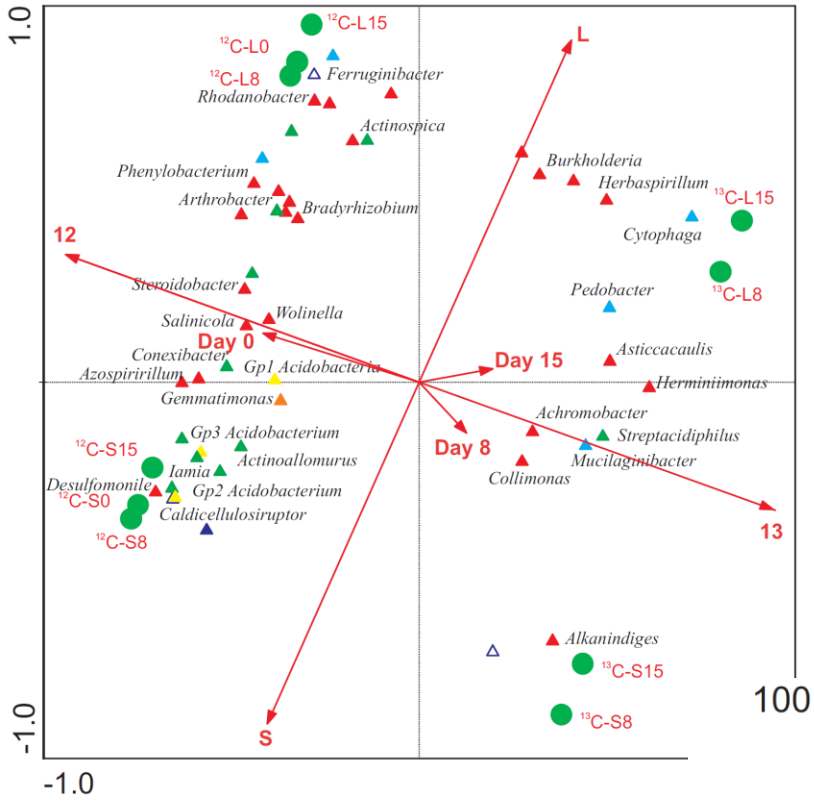
vzorky obsahující ^{13}C, nebo ^{12}C DNA

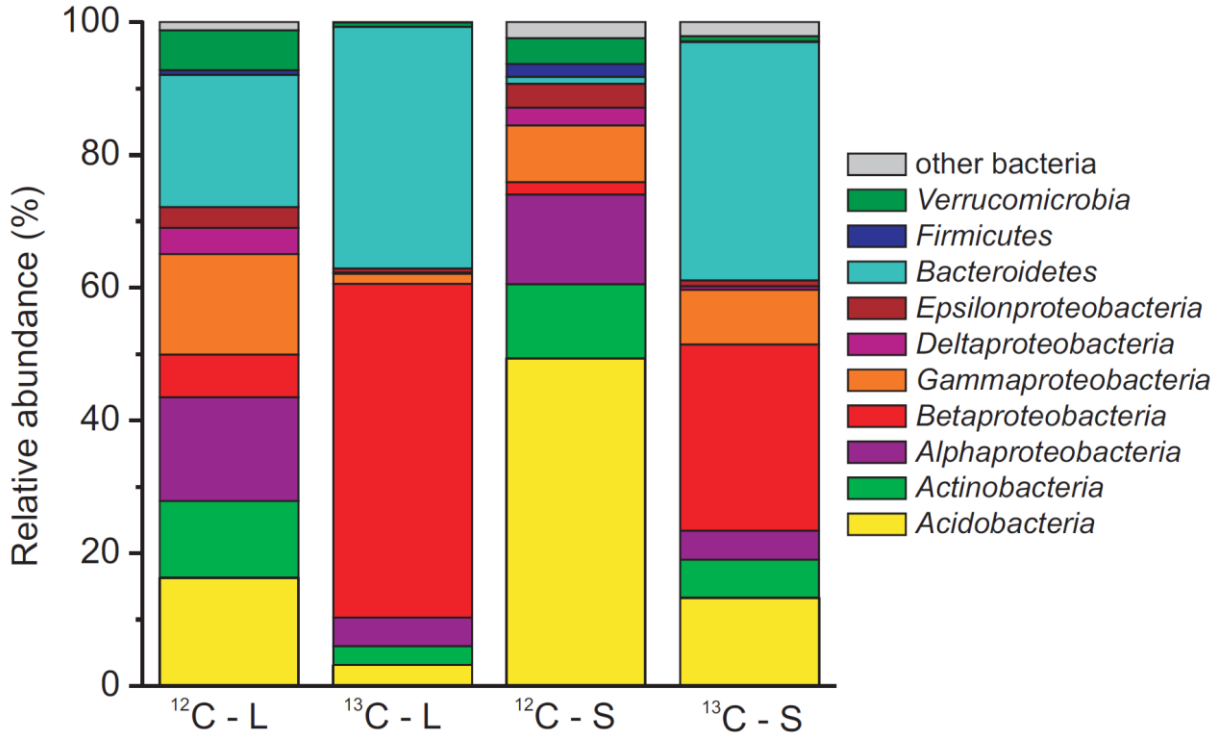
- DGGE
- T-RFLP
- Sekvenace

Příklad: Bakteriální rozkladači celulózy

Bakterie označené ^{13}C patří hlavně mezi Betaproteobacteria *a* Bacteroidetes.

Rody Mucilaginibacter *a* Herminiimonas *akumulovaly nejvíce ^{13}C uhlíku z celulózy.*





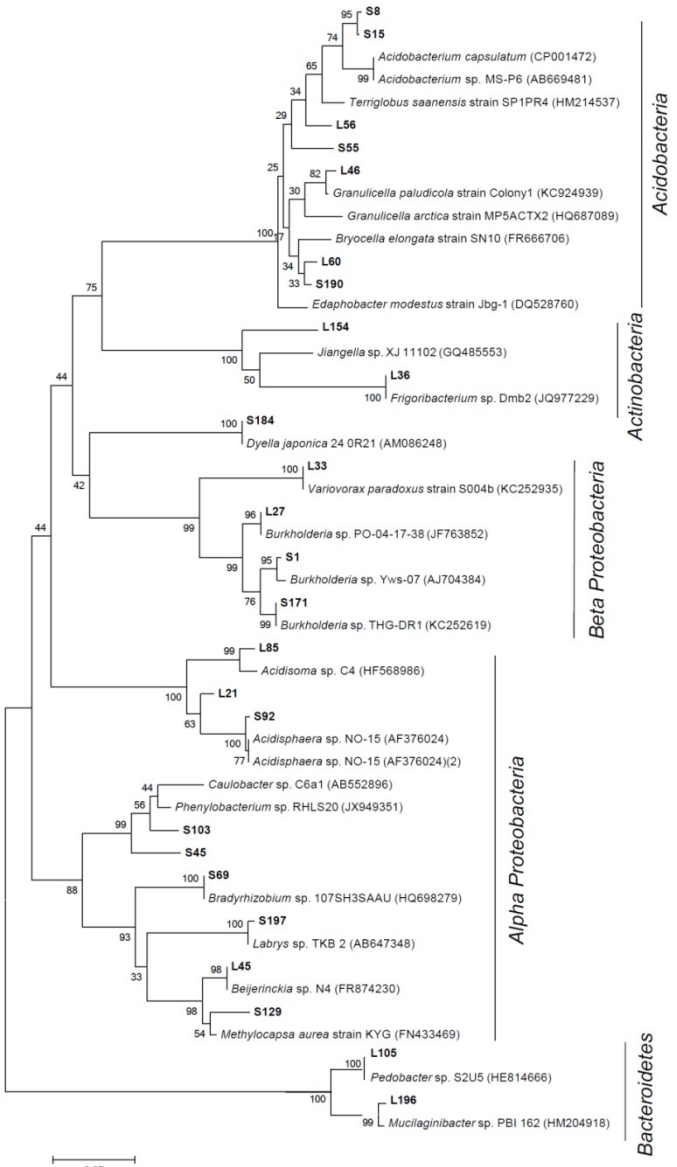
Stable Isotope Probing (SIP)

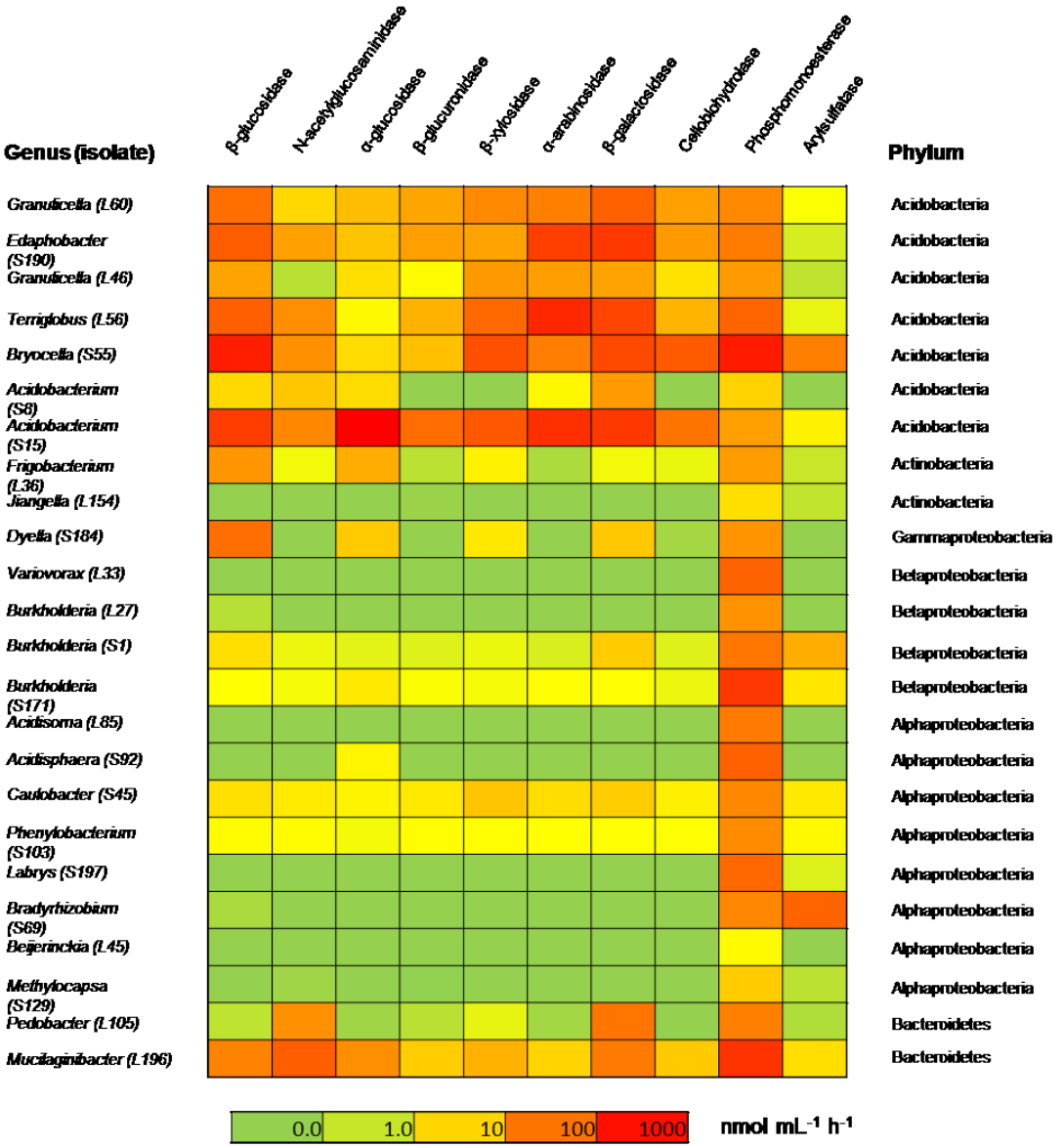
Výhody a nevýhody metody

- rozlišuje aktivní a neaktivní mikroorganismy (biomasu)
- umožňuje selektivně sledovat využití určitého (značeného) substrátu
- ve spojení s RT-PCR je extrémně citlivá a umožňuje detekci složení nerostoucího společenstva

- vyžaduje spojení s dalšími technikami (PCR nebo RT-PCR a DGGE, RFLP nebo FAME)
- nutné speciální vybavení (ultracentrufuga a nejlépe také RT-PCR)
- substráty značené stabilními isotopy jsou drahé (13C glukóza – 2.000 Kč/g, ^{13}C celulóza – 36.000 Kč/g, 13C lignin – 500.000 Kč/g) a je jich k dispozici omezený počet

Isolation and characterization of bacterial strains: enzyme activity

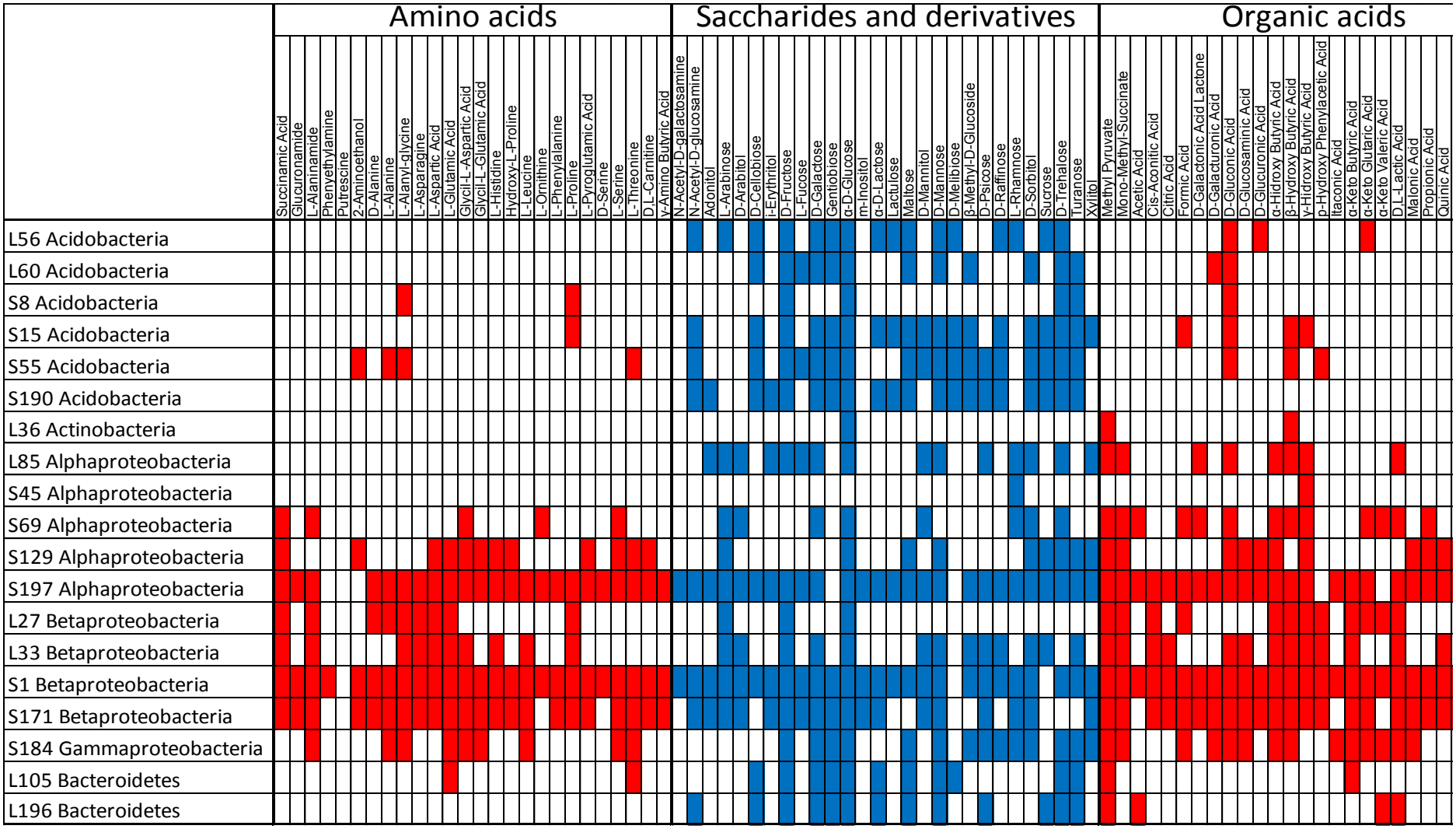




SEE MORE IN THE LECTURE OF SALVADOR LLADÓ ON WEDNESDAY

Lladó et al. Biol. Fertil. Soils in press

Isolation and characterization of bacterial strains: carbon sources



Most versatile taxa belong to Proteobacteria.
Acidobacteria and Bacteroidetes can grow on multiple mono- and oligosaccharides.

SEE MORE IN THE LECTURE OF SALVADOR LLADÓ ON WEDNESDAY

Lladó et al. Biol. Fertil. Soils in press

Dominant bacteria: Expression in forest soil and litter across seasons

Isolated major bacteria from forest litter and soil were genome sequenced

Genome sequences were used to identify transcripts belonging to the isolated taxa

Strain	Identification	Relative abundance (%)		Genes in	Expressed in	Transcription (ppm)		
		Soil	Litter	genome	metatranscriptome	Soil	Litter	Seasonality
L45	*Beijerinckia sp.* (Alphaproteobacteria)	1.2	0.5	5651	14	9	15	NO
L46	*Granulicella sp.* (Acidobacteria)	0.5	1.2	4731	253	288	406	**YES**
L56	*Terriglobus sp.* (Acidobacteria)	1.2	1.8	4062	13	7	19	**YES**
S8	*Acidobacterium sp.* (Acidobacteria)	5.6	6.7	4913	112	24	14	**YES**
S15	*Acidobacterium sp.* (Acidobacteria)	3.6	4.1	5653	278	261	128	**YES**
S55	*Bryocella sp.* (Acidobacteria)	2.0	1.6	4569	127	95	574	**YES**
S69	*Bradyrhizobium sp.* (Alphaproteobacteria)	2.4	4.4	7017	125	319	150	**YES**
S103	*Phenylobacterium sp.* (Alphaproteobacteria)	0.2	1.0	7707	101	103	36	**YES**
S129	*Methylocapsa sp.* (Alphaproteobacteria)	6.2	7.4	6957	15	212	119	**YES**
S190	*Edaphobacter sp.* (Acidobacteria)	0.5	1.3	4440	208	32	61	**YES**

Between 10 and 300 genes were identified as transcripts in litter and soil

Bacteria expressed similar genes in litter and soil

Spectra of expressed genes differed between summer and winted

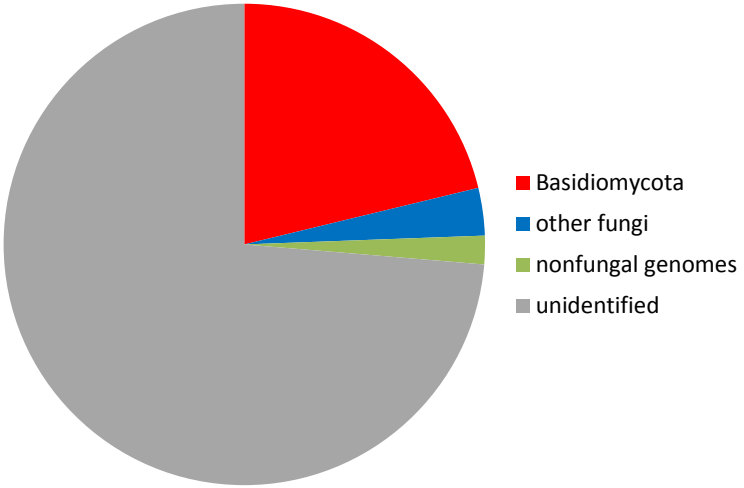
SEE MORE IN THE LECTURE OF SALVADOR LLADÓ ON WEDNESDAY

Activity of the dominant fungus, mycorrhizal *Russula ochroleuca*

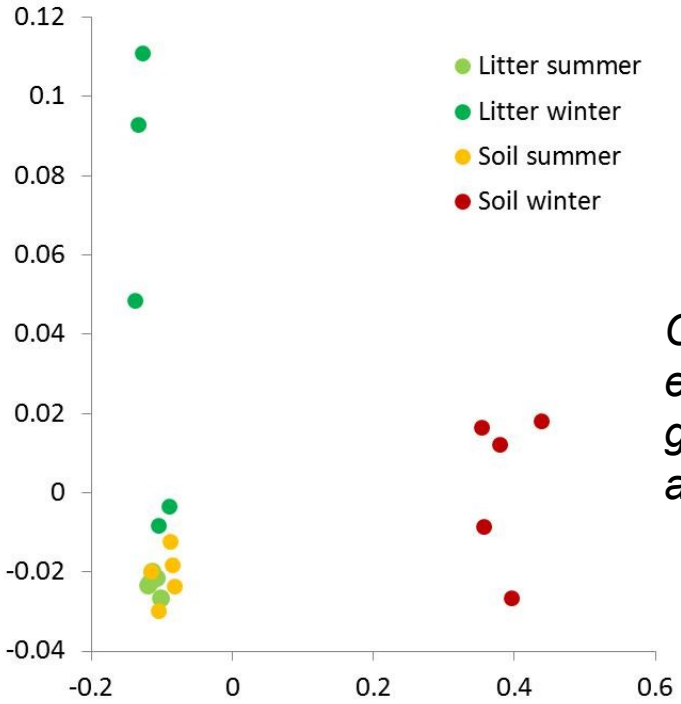
Russula ochroleuca represents the second most abundant fungal taxon in the studied forest soil

As mycorrhizal symbiont, it does not grow in culture

Genome was obtained by shotgun sequencing DNA, isolated from a fresh fruitbody



The vast majority of predicted genes does not have clear relatives for which producer and function are known



Composition of expressed genes differs among seasons

Metatranscriptomics – summary

- higher percentage of reads receive taxonomy annotation than functional annotation (due to hypothetical proteins known from sequenced genomes)

- functional annotation is more reliable than taxonomic annotation

- in fungi, most transcripts can not be reliably assigned to either Ascomycota or Basidiomycota

- current resources for annotation (GenBank, MG Rast) contain still little fungal sequences (and genomes), in reality, many more genomes are sequenced, but it is difficult to use them for annotation

- most microbial transcripts/functions represent basic metabolism which can be of limited value for the exploration of environmental processes

Soil metatranscriptomics is currently technically feasible and can deliver interesting data.