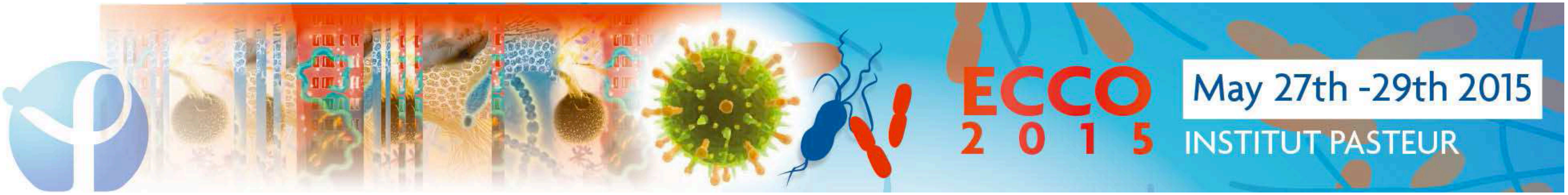


Contribution of metagenomics (barcoding and metabarcoding) in the study of fungal diversity

Marc Buée
INRA Nancy, France

UMR 1136 Interactions Arbres - Microorganismes, INRA, Université de Lorraine

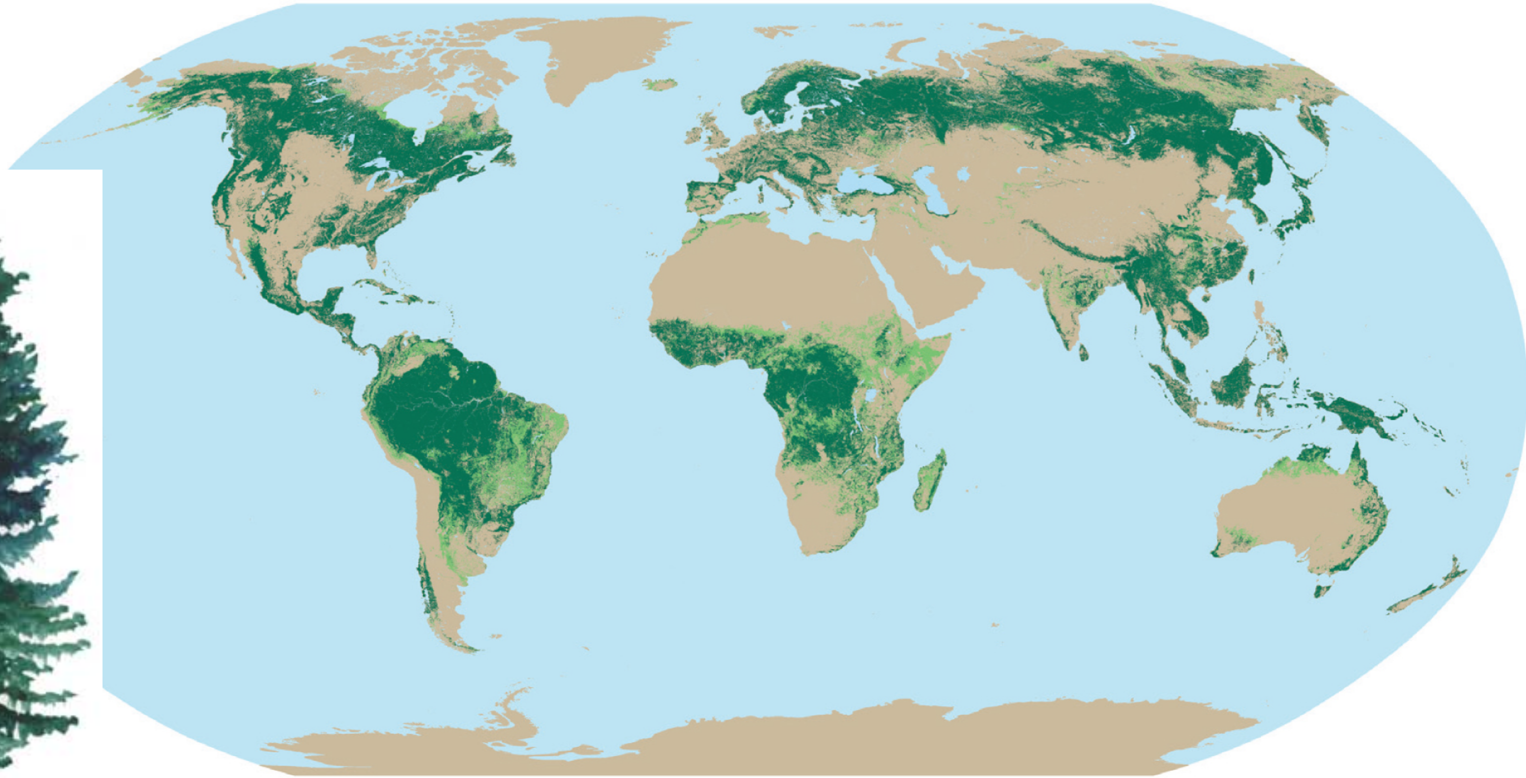
INTERACTIONS ARBRES-MICROORGANISMES

UNIVERSITÉ DE LORRAINE

INRA SCIENCE & IMPACT

arbre

The world's forests

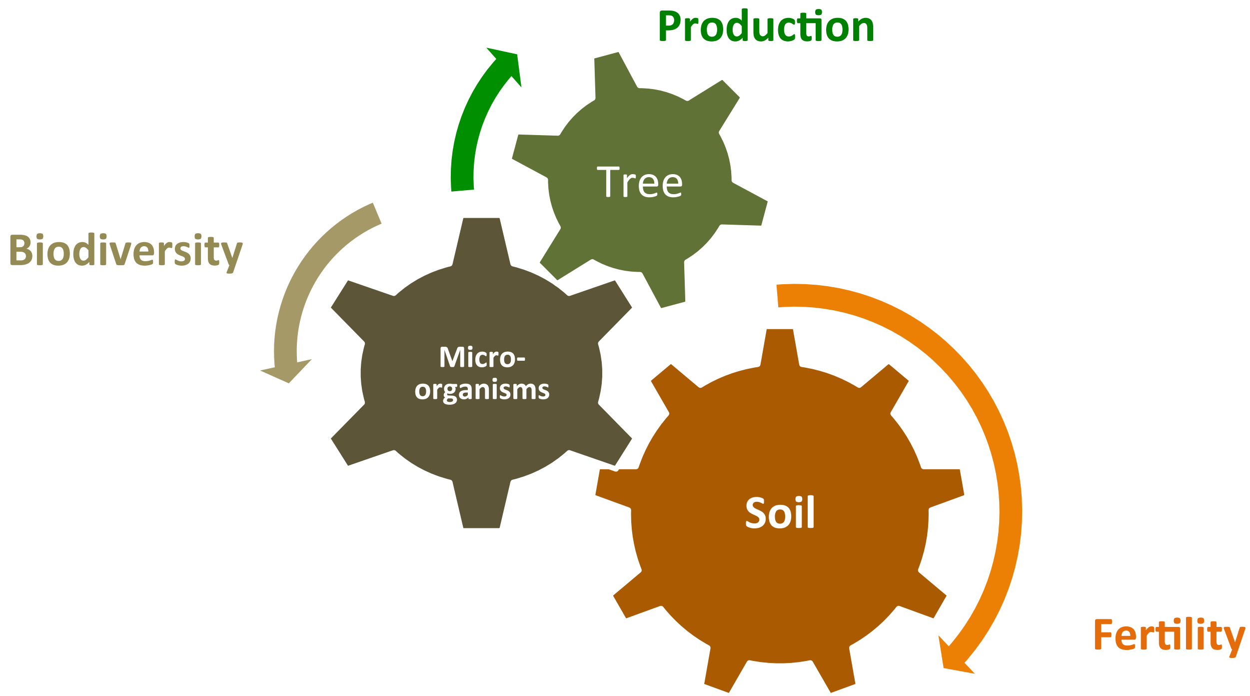


Forest
Other wooded land
Other land
Water

In 2006, the world's total forest area was estimated at 3 952 million hectares, which corresponds to about 30% of the land area. Important role in nutrient cycles and C sequestration.

Forest ecosystem functioning

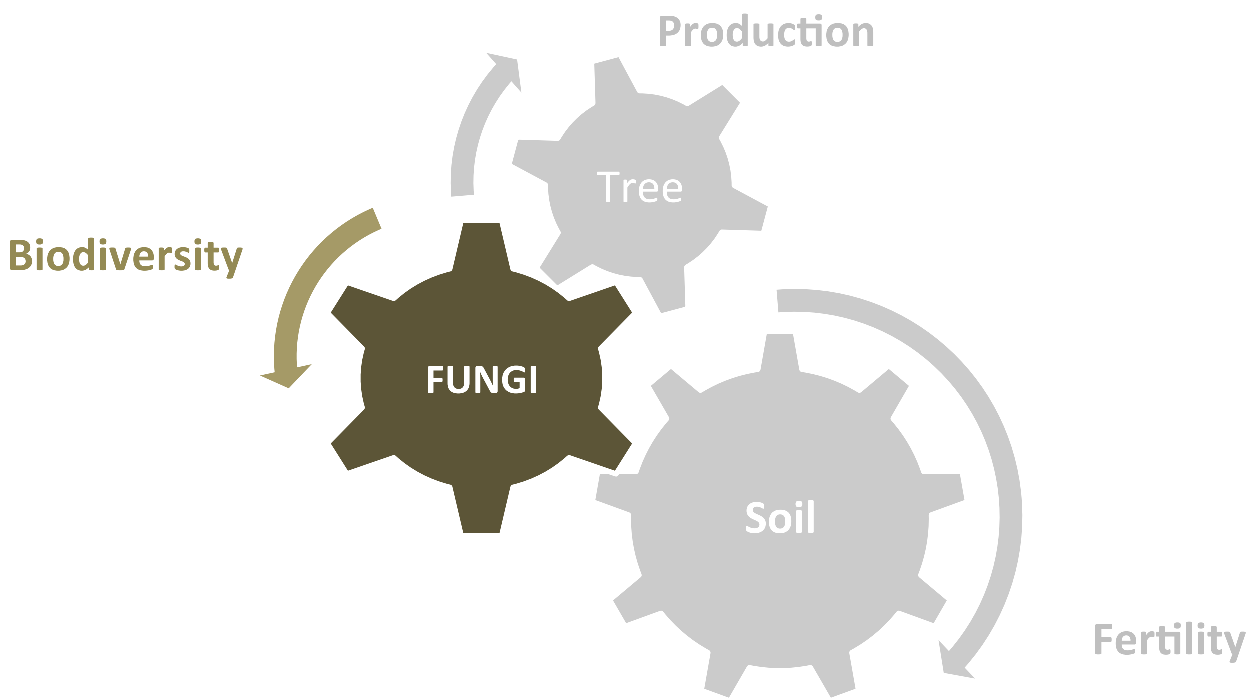
Diversity of microbial communities in the forest ecosystems



Between the biomass production and the forest soil, the microbes are involved in fixation and mineralization of elements, degradation of the organic matter, mobilization of nutriments and plant nutrition.

Forest ecosystem functioning

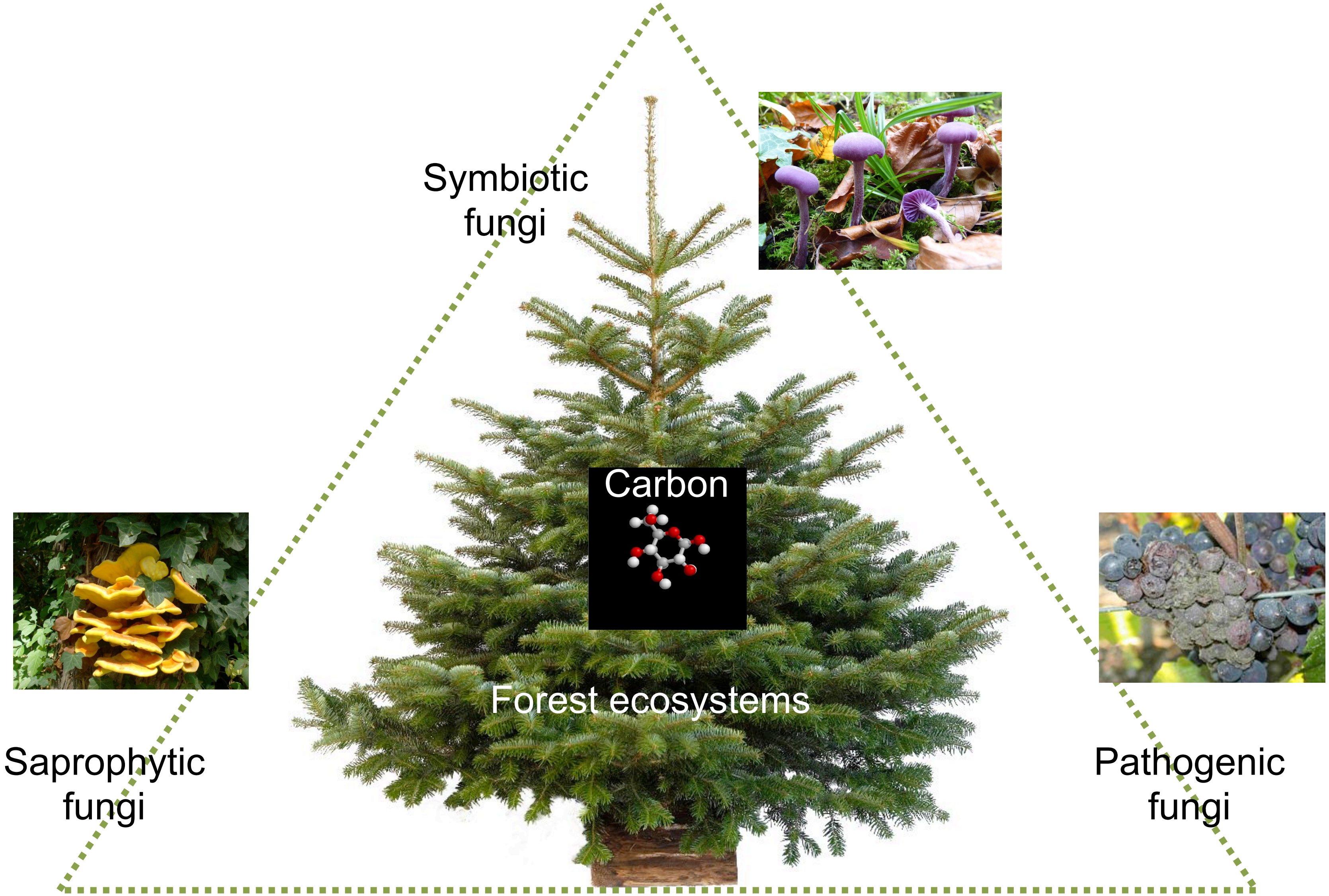
Diversity of fungal communities in the forest ecosystems



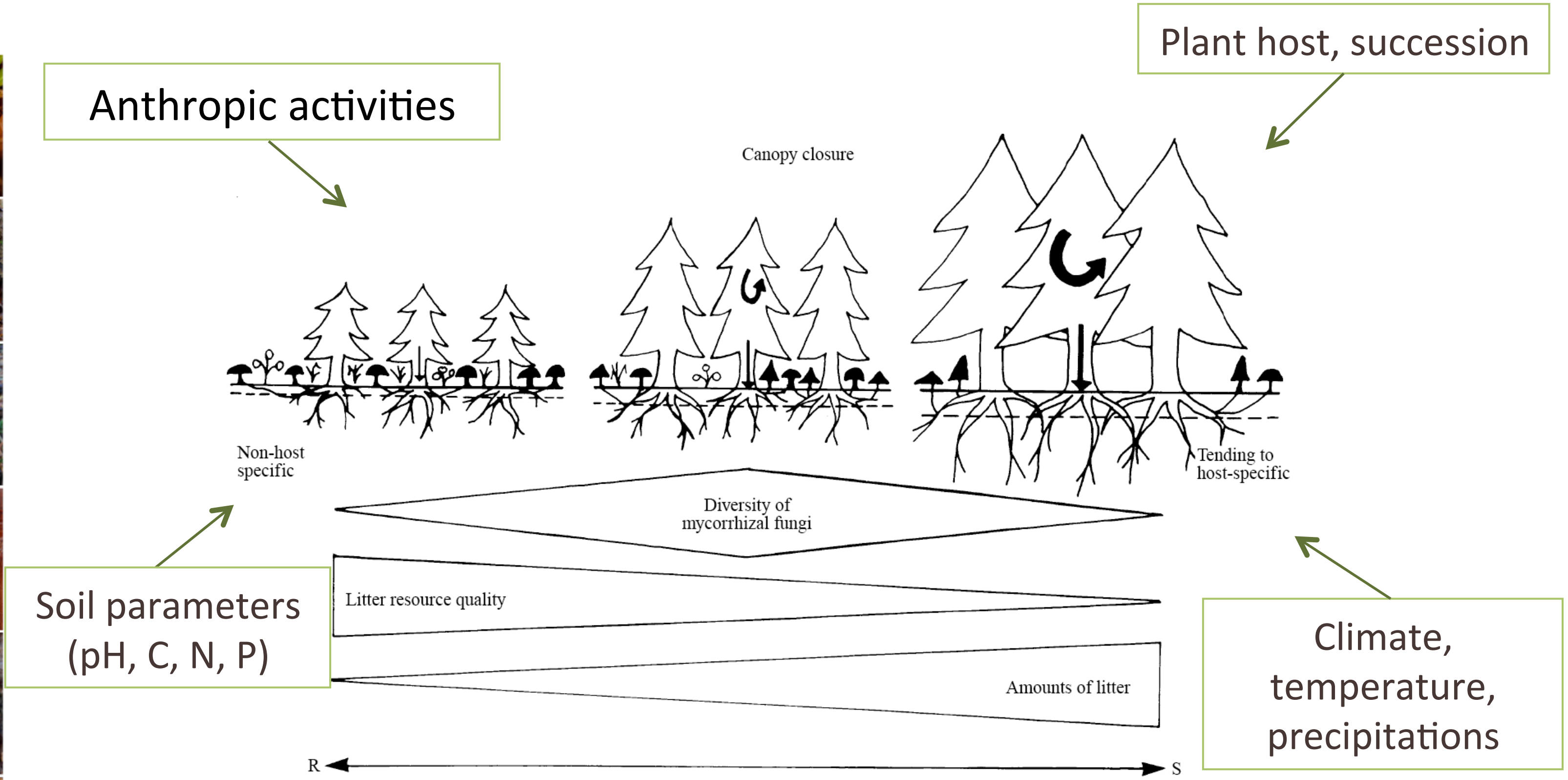
« … About 100,000 fungal species have been described so far, but it has been estimated that there may be from 1.5 to 5.1 million extant fungal species. Over the last decade, about 1200 new species of Fungi have been described in each year. At that rate, it may take up to 4000 y to describe all species of Fungi using current specimen-based approaches. »

Hibbett D. 2011. Fungal Biology Reviews

Diversity of fungi is linked to diversity of habitats and ecological traits (mutualistic, saprophytic or pathogenic fungi)



Why this huge diversity?

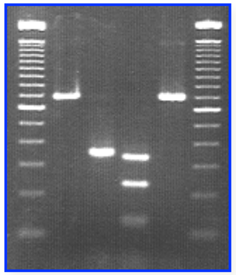


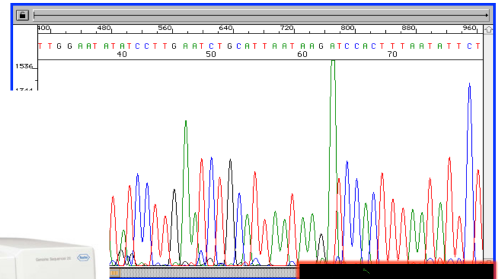
Adapted from Dighton & Masson, 1985

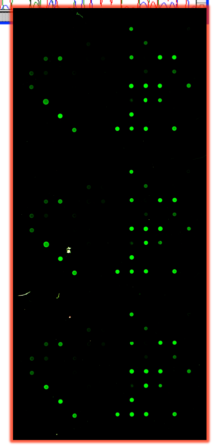
Soil fungi occur in remarkably species-rich assemblages. One of the prevailing hypotheses to explain this diversity is niche differentiation; by occupying distinct ecological niches within a site, multiple fungal species are able to co-occur.

How investigate this huge diversity ?

From sporocarp inventory to molecular approaches(complementary approaches)







Molecular Ecology ~~(2001)~~ 2015 **10**, 1855–1871

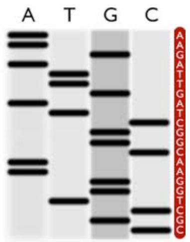
How investigate this huge diversity ?

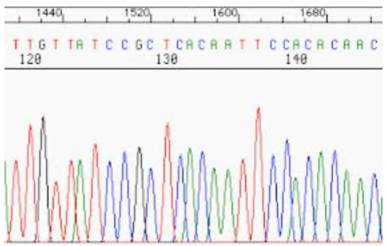
REVIEW ARTICLE

The second molecular revolution in ectomycorrhizal ecology: peeking into the black-box at the global scale

The recent improvements in sequencing techniques and bio-informatics make the mark of the second molecular revolution in fungal ecology (meta-barcoding).

Abstract





1977
Sanger sequencing

1990
Sequencing with fluorescence measurement

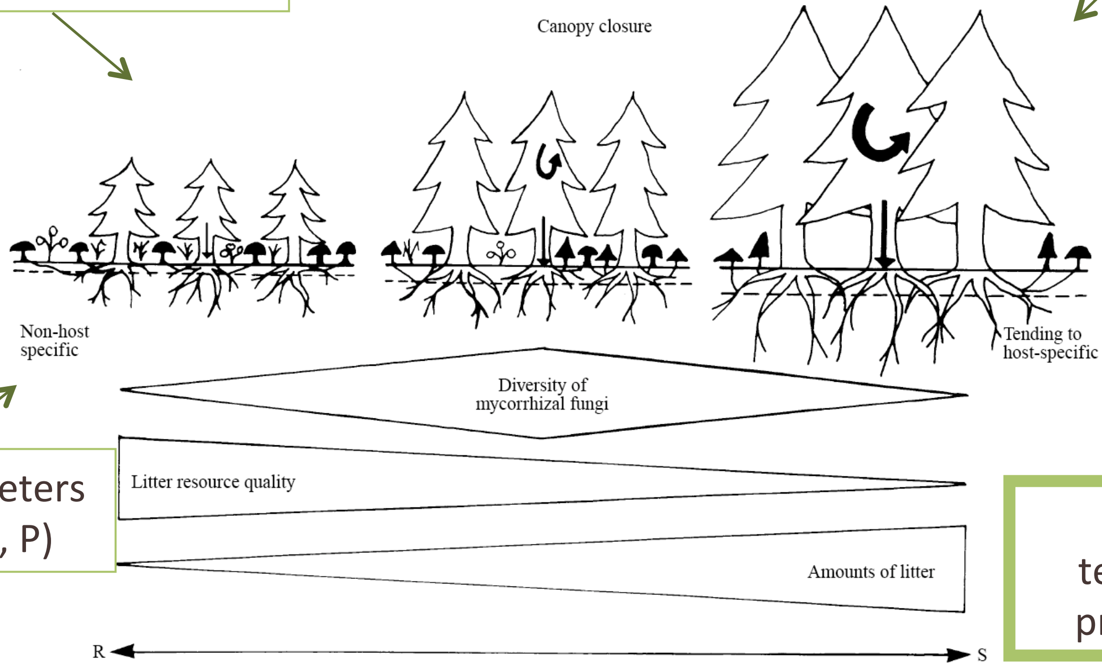
1999
Sequencing with capillary electrophoresis

2006-2014
New Generation Sequencing (NGS)

Why this huge diversity?

Anthropic activities

Plant host, succession



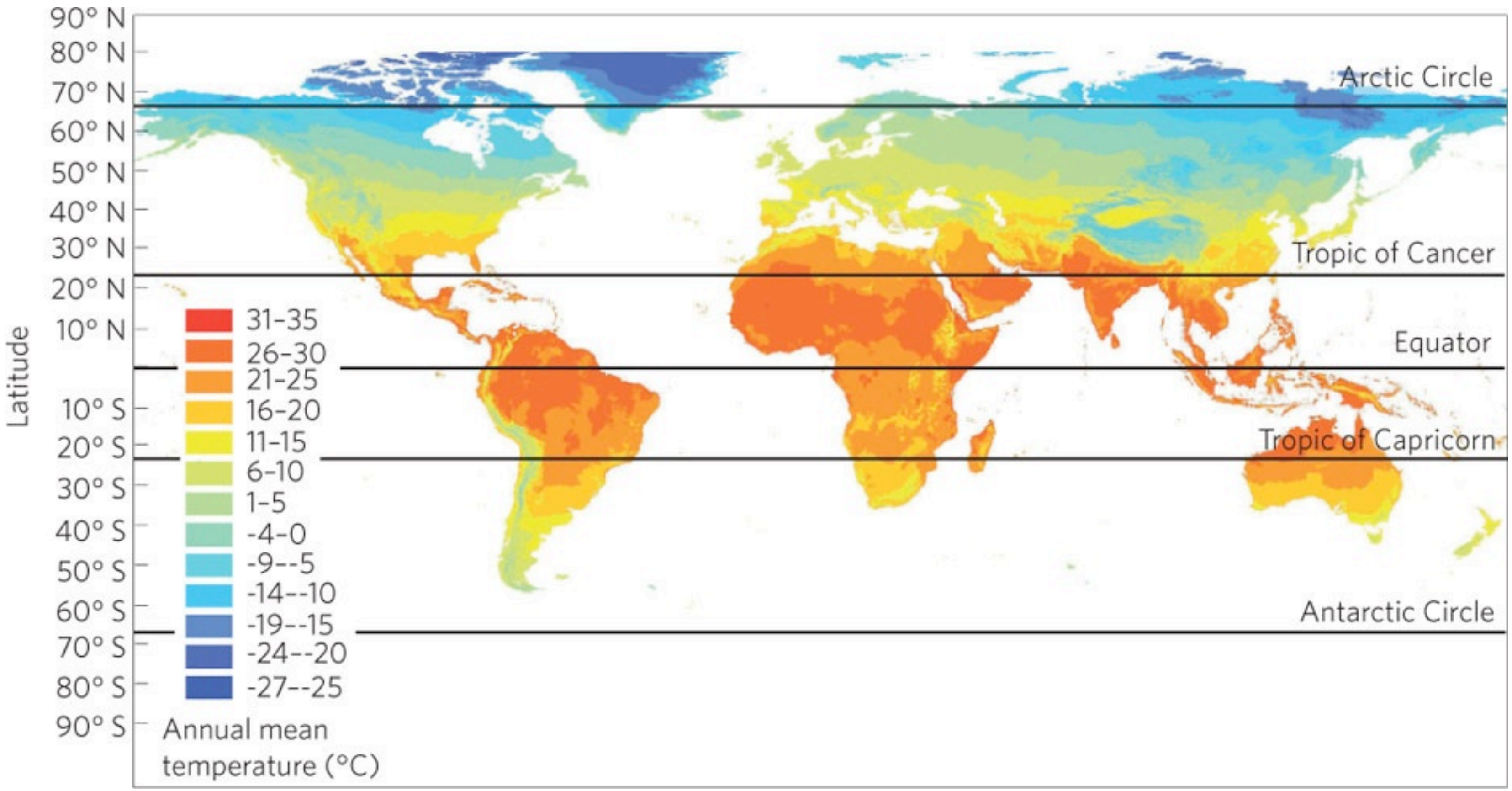
Soil parameters (pH, C, N, P)

Climate, temperature, precipitations

Adapted from Dighton & Masson, 1985

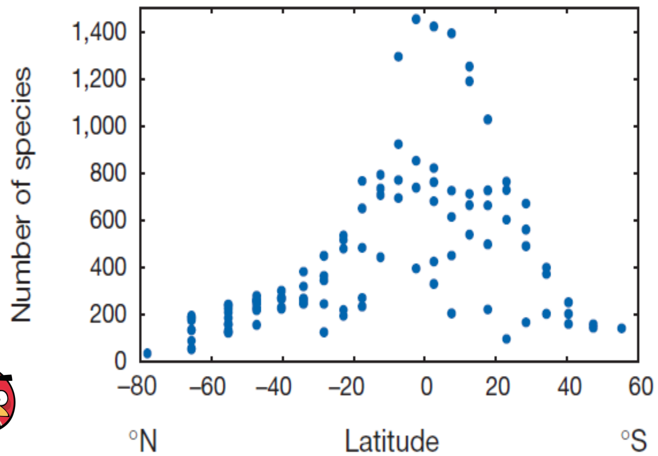
Advantage of New Generation Sequencing (NGS) to investigate unprecedented scale of sampling in fungal ecology… => Use natural climatic gradients to investigate the impact of temperatures and precipitations on fungal diversity and community structure

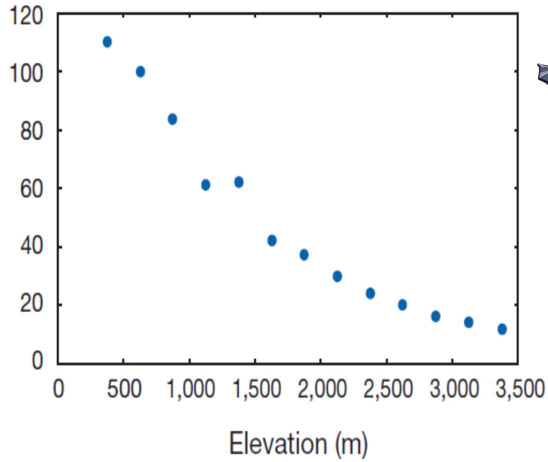
Are fungi follow the same global distribution pattern than macro-organisms?



<u>Richness and temperature gradients</u>: a biogeographic approach

A common pattern in ecology is the latitudinal (and altitudinal) gradient of diversity. These examples with birds and bats show these common patterns : the trend is a lower species richness when moving away from the equator (or when the altitude increases). Positive correlation between energic (thermic) gradient and richness.

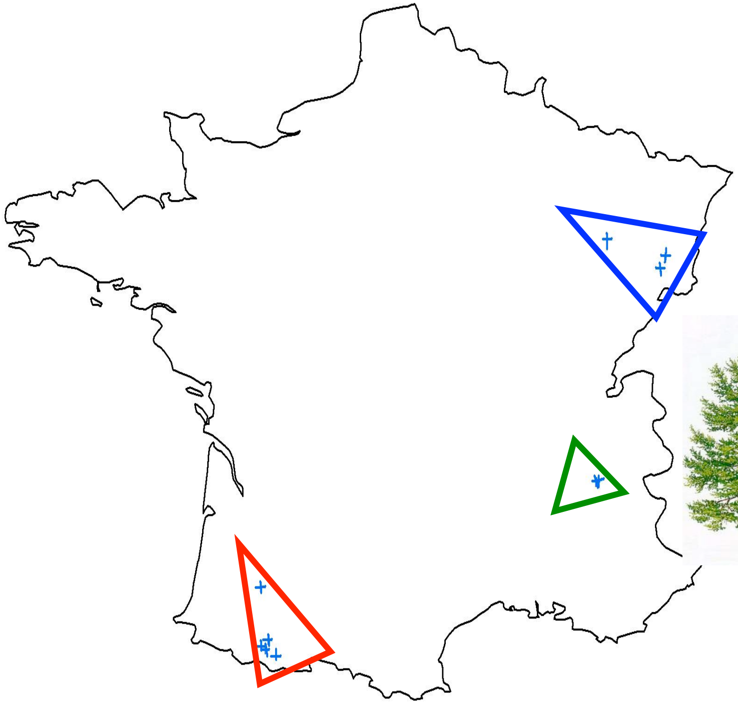




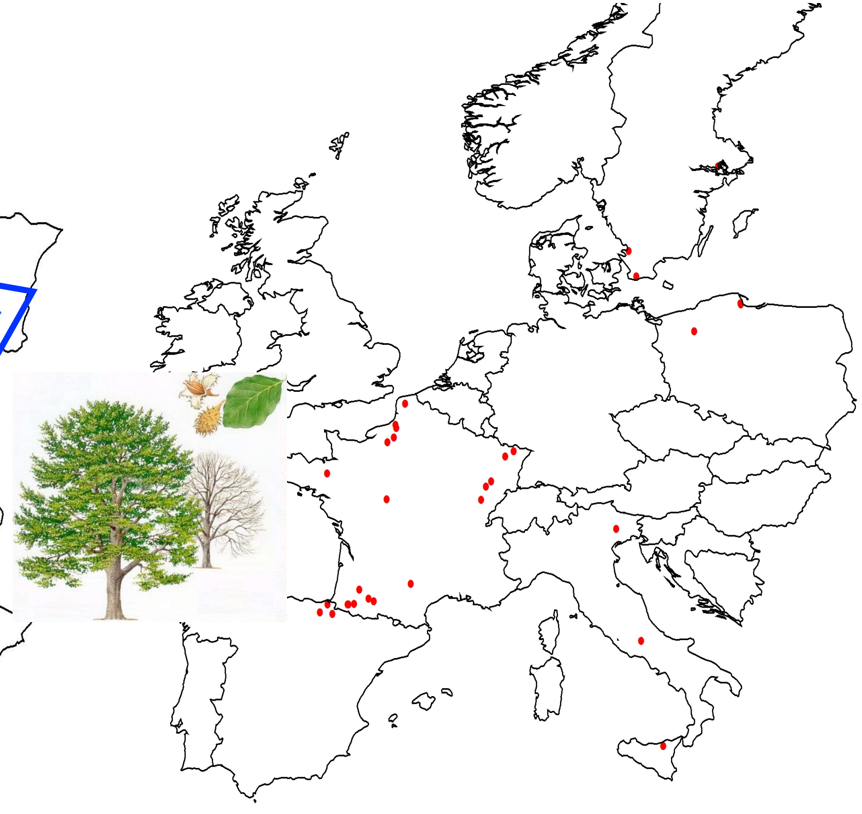
Gaston, 2000 Nature

For fungal communities: numerous controversies (Arnold and Ludzoni, 2007; Amend et al. 2010; Fierer et al. 2011; Bahram et al. 2012; Miyamoto et al. 2014; Tedersoo et al. 2014…)

Ecology of fungal communities in beech forests:
Use of NGS (metabarcoding) in a biogeographical approach



Altitudinal gradients in France

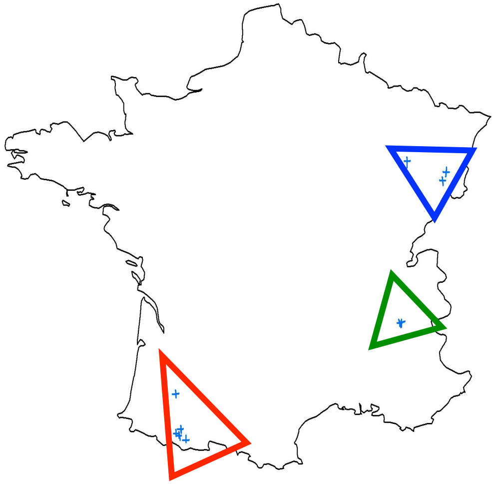


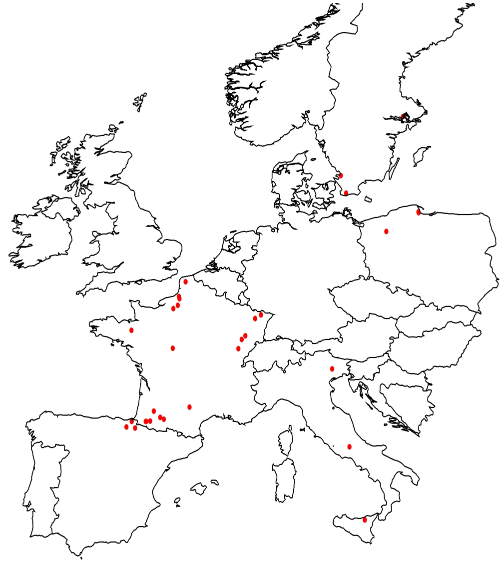
Latitudinal gradient in Europe

Aurore Coince PhD Thesis (UE Baccara project)

BACCARA

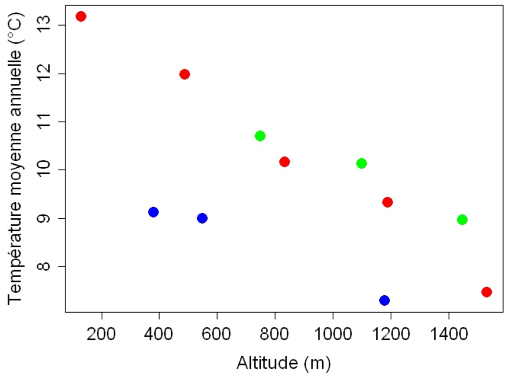
Ecology of fungal communities in beech forests: Use of NGS (metabarcoding) in a biogeographical approach

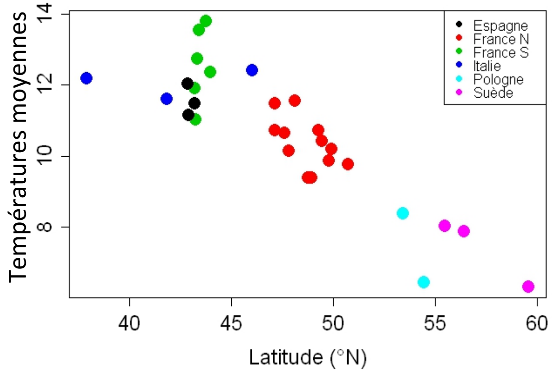




Altitudinal gradients in France

Latitudinal gradient in Europe

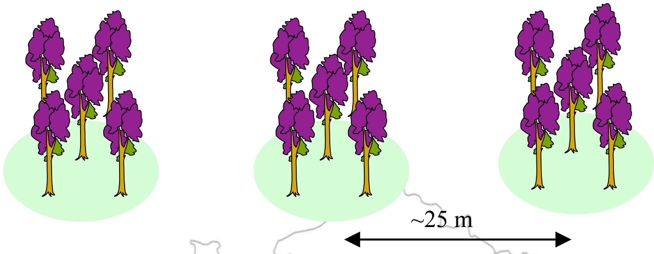




BACCARA

Coince et al. 2013 Fungal Ecology, Coince et al. 2014 PlosOne , Coince et al. (in preparation)

Methods

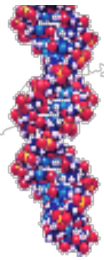


– Sampling strategy –

495 soil cores for Altitudinal gradients
465 soil cores for Latitudinal gradient (31 sites)

– Roots handling (1 sample = 1 tree = pool of 3 cores)–

– DNA extraction –



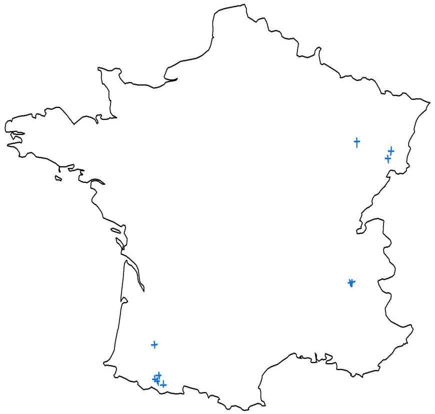
– PCR amplification (ITS1 region) –
ITS1F-ITS2 primer pairs

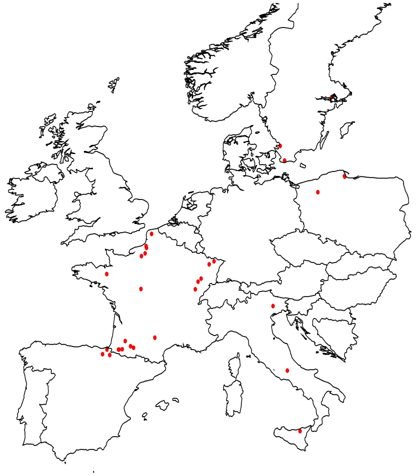
– Amplicons libraries (with molecular tag) and Pyrosequencing –

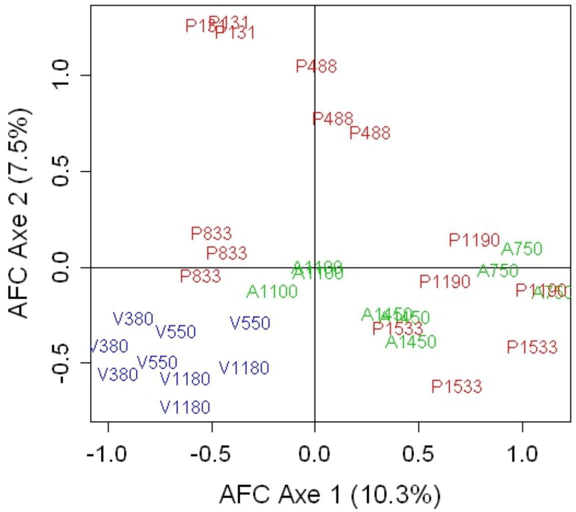
– Bioinformatic analyses and taxonomic assignment –

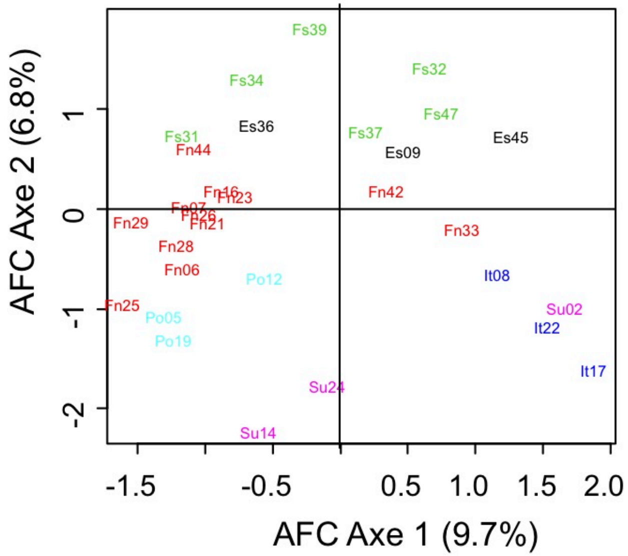
Composition (species assemblage) of fungal communities: a biogeographical approach

BACCARA



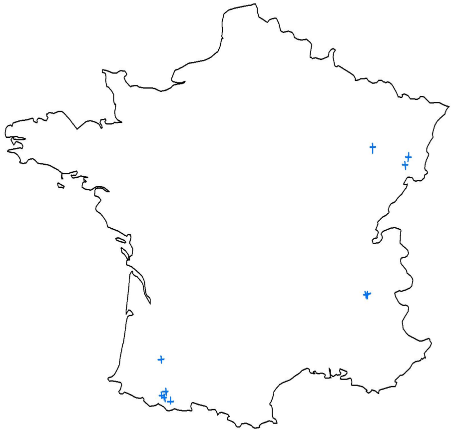


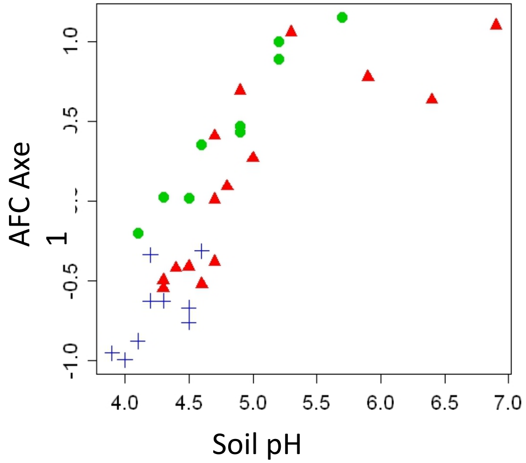


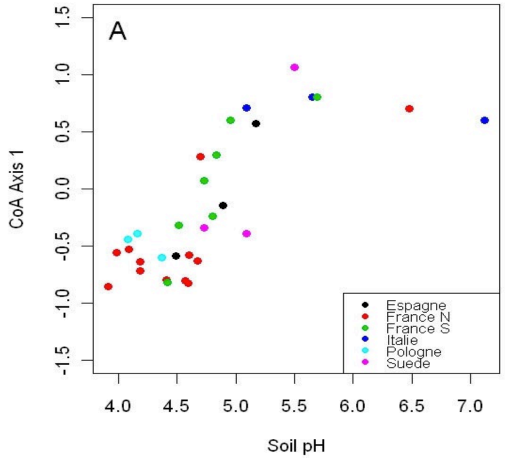


Correspondence analysis (two first axes) – links with environmental parameters?

Composition (species assemblage) of fungal communities: a biogeographical approach

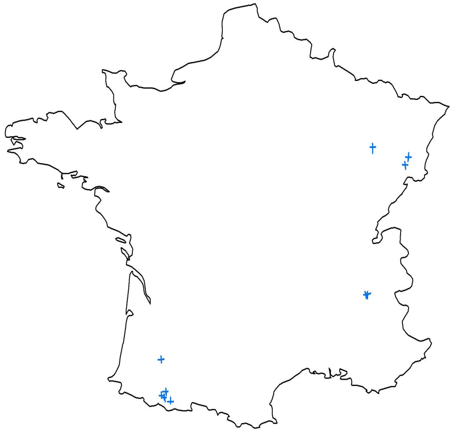


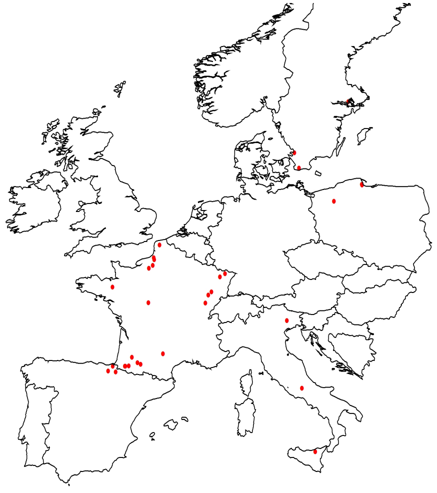


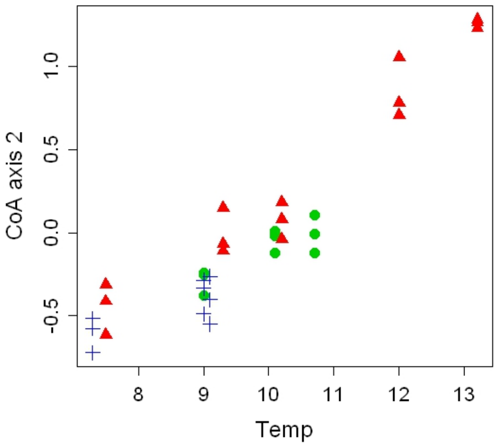


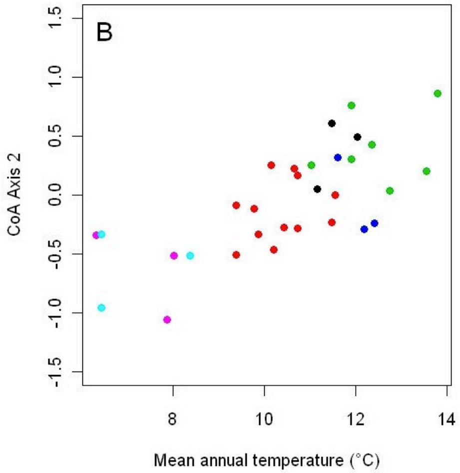
Axe 1 = soil pH in both studies

Composition (species assemblage) of fungal communities: a biogeographical approach





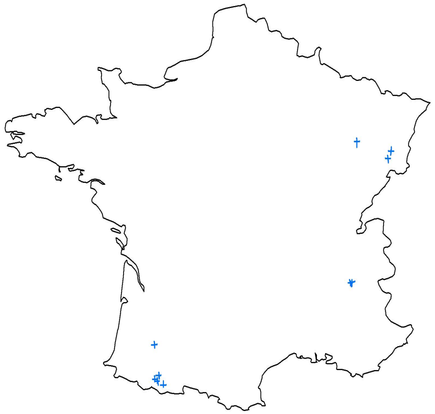




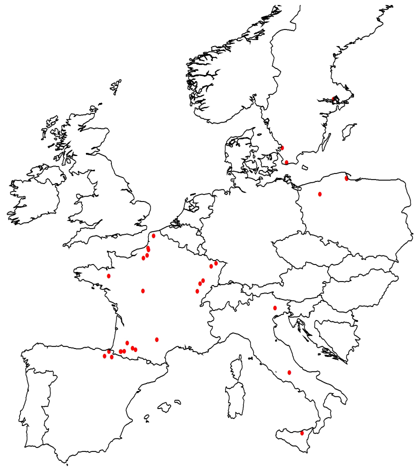
Axe 2 = mean temperatures in both studies

BACCARA

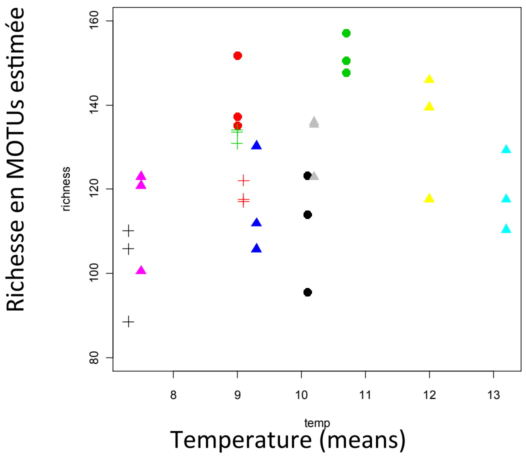
Richness estimation (proxy) in fungal communities along geographical gradients

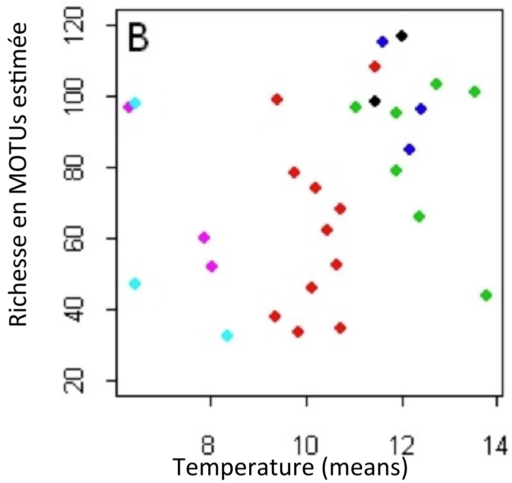


Elevation Diversity Gradient (EDG)
Mid Domain Effect model?

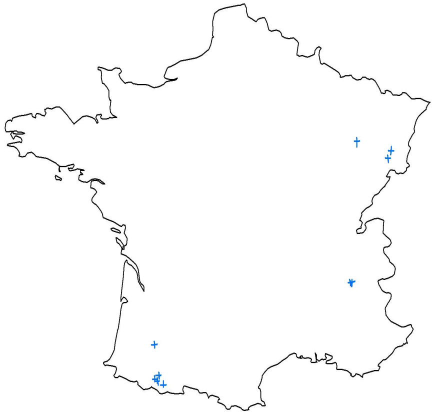


Latitudinal Diversity Gradient (LDG)
Monotonic relationship species /temperature ?

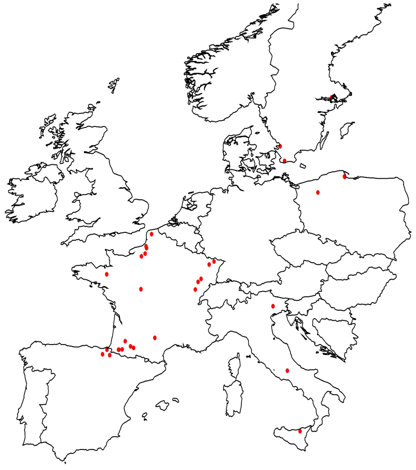




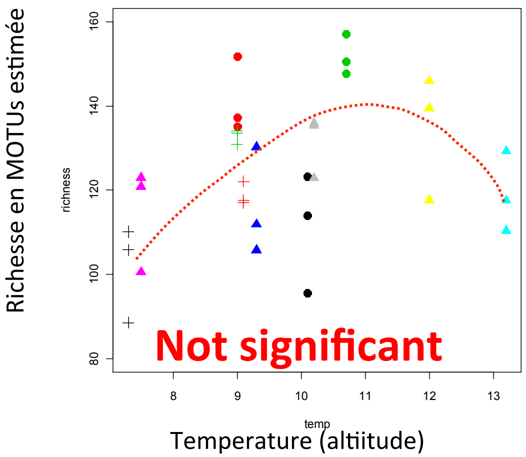
Richness estimation (proxy) in fungal communities along geographical gradients

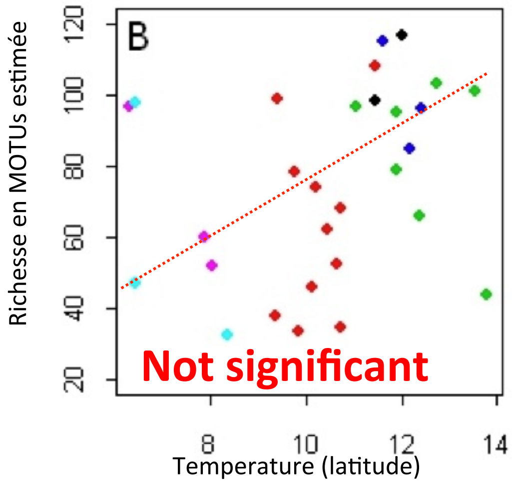


Elevation Diversity Gradient (EDG)
Mid Domain Effect model? No!



Latitudinal Diversity Gradient (LDG)
Monotonic relationship species /temperature ? No!





Richness of fungi communities (altitudinal gradient) in *Pinus sylvestris* forests

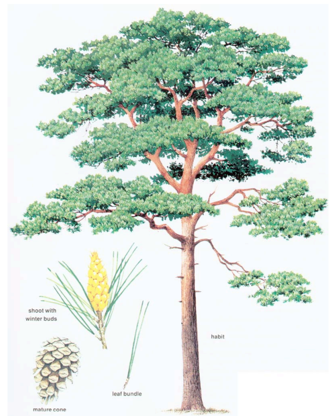






	ALL FUNGI[§]			MYCORRHIZAL FUNGI[††]						OTHERS[Ω]		
	BS			in RT			in BS			in BS		
	Estimate ± SE	F	p-value	Estimate ± SE	F	p-value	Estimate ± SE	F	p-value	Estimate ± SE	F	p-value
All MOTUs												
Temperature	-0.06 ± 0.02	0.2	ns	-0.02 ± 0.04	0.2	ns	-0.01± 0.04	0.0	ns	-0.04 ± 0.02	2.9	ns
pH	0.23 ± 0.06	13.4	**0.002**	0.02 ± 0.08	0.7	ns	0.08 ± 0.07	1.4	ns	0.26 ± 0.06	22.7	**0.0003**
C/N	-0.01 ± 0.00	1.4	ns	0.01 ± 0.00	0.2	ns	-0.01 ± 0.00	1.8	ns	0.003 ± 0.00	1.6	ns
roots	--	--	--	-0.61 ± 0.70	0.7	ns	--	--	--	--	--	--
Precipitation	0.001 ± 0.00	0.5	ns	0.01± 0.00	0.3	ns	-0.01± 0.00	6.0	**0.027**	0.000 ± 0.00	0.43	ns
pH	0.19 ± 0.05	9.6	**0.007**	0.01 ± 0.08	0.7	ns	0.06 ± 0.04	2.3	ns	0.222 ± 0.04	30.4	**0.0001**
C/N	-0.01 ± 0.00	2.3	ns	0.01 ± 0.00	0.2	ns	-0.05 ± 0.00	2.4	ns	0.002 ± 0.00	0.88	ns
roots	--	--	--	-0.50 ± 0.66	0.6	ns	--	--	--	--	--	--

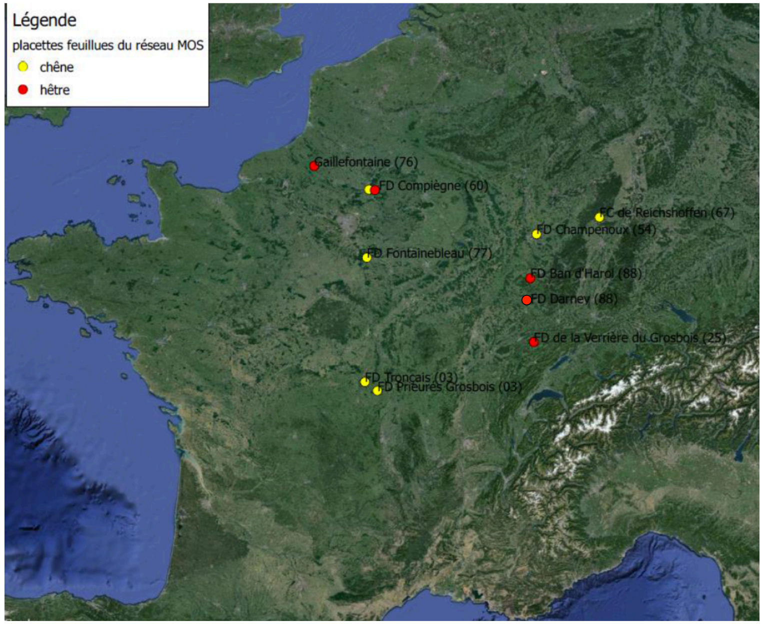
Generalized Linear Mixed Models testing the response of soil and mycorrhizal fungal communities to climatic (independent models for temperature and precipitation) and edaphic variables, considering the estimated richness as fixed effect.

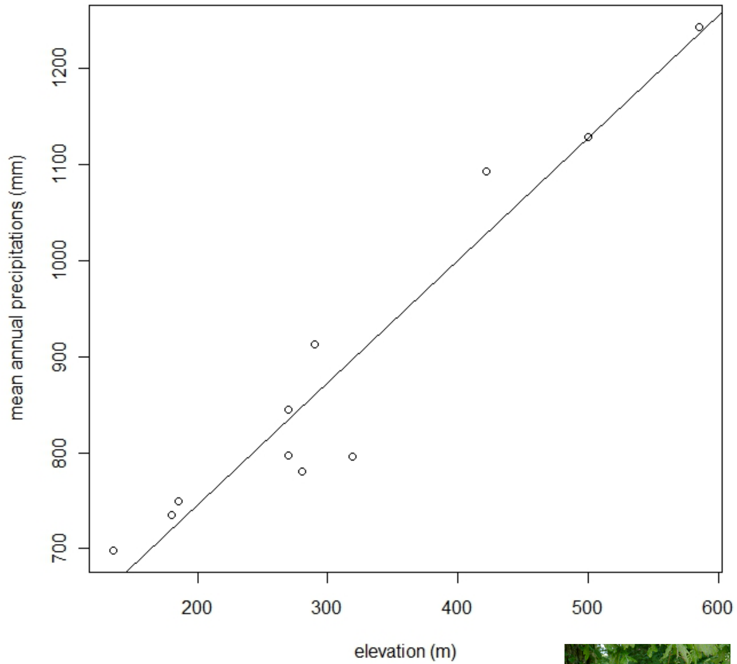


Strong and significant positive correaltion between fungal richness and soil pH (not for EcM fungi), but correlation between EcM fungi and annual precipitations (mm)
The host tree may modulate the impact of edaphic parameters on ECM fungal richness (?)

Rincon et al. 2015, Env. Microbiol.

<u>Fungal richness</u> and climatic parameters: Do ectomycorrhizal and non-symbiotic fungi response differently to precipitations?

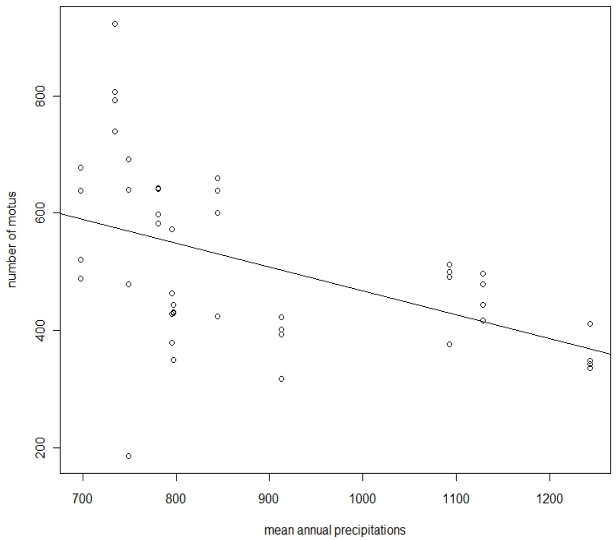




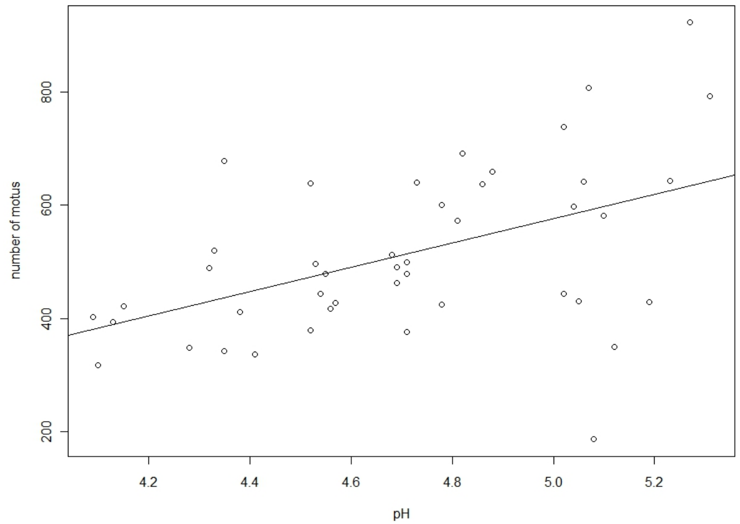
We focused especially on ECM fungi associated with Fagaceae *Fagus sylvatica* and *Quercus petraea,* which are particularly sensitive to climate changes using an East-West rainfall gradient in temperate French lowland forests (n=88 soil cores).

Akroume et al. (in preparation)

Fungal richness and climatic parameters: Do ectomycorrhizal and non-symbiotic fungi response differently to precipitations?



Linear regression between fungal **richness** and mean **annual precipitations** (mm)

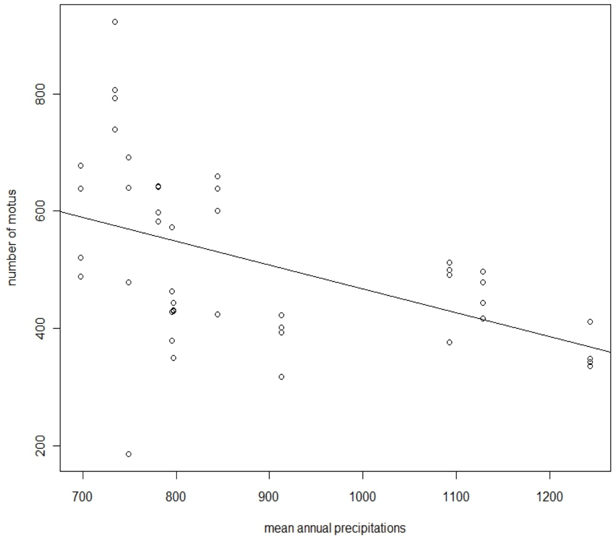


Linear regression between fungal richness and **soil pH**

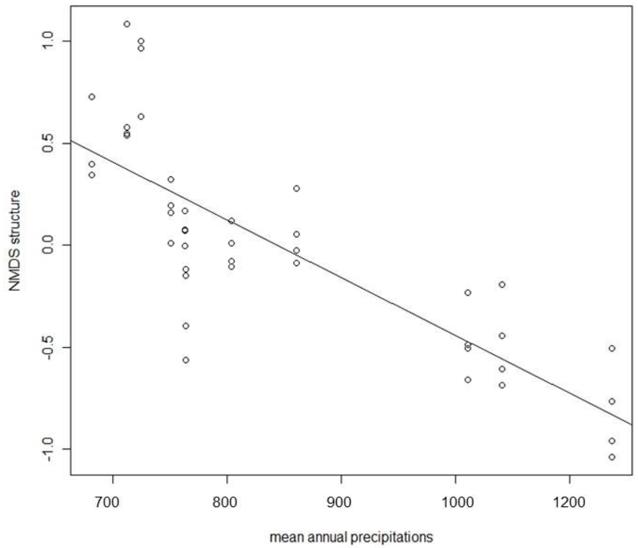
We confirm the relationship between the soil pH and fungal richness
+ significant link between fungal richness and precipitation (negative correlation).
By contrast, ECM richness is not linked with this climatic factor

For fungal communities: numerous controversies (Arnold and Ludzoni, 2007; Amend et al. 2010; Fierer et al. 2011; Bahram et al. 2012; Miyamoto et al. 2014; Tedersoo et al. 2014…)

Composition (species assemblage) of fungal communities in beech / oak forests



Linear regression between fungal **richness** and mean **annual precipitations** (mm)



Correlation between NMDS **structure** (axis1+axis2) and mean **annual precipitations** (mm)

Ectomycorrhizal fungi are more sensitive to mean annual precipitations and less structured by soil characteristics like pH or Ca concentrations than saprobes. The host tree modulates the impact of edaphic parameters on ECM fungal richness. In return EcM fungi may be more sensitive to climatic parameters, as their hosts…

Akroume et al. (in preparation)

In the “Omics era”, how fungal culture collections (and voucher individuals) can be a research tools in ecology?

Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*

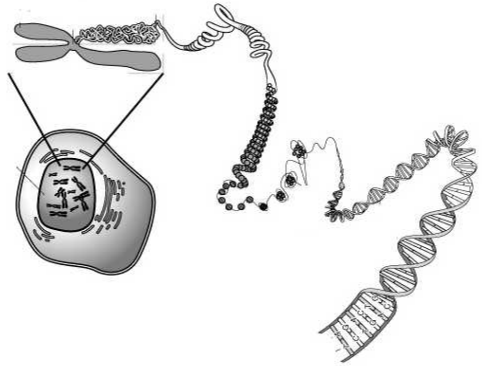
Conrad L. Schoch[a,1], Keith A. Seifert[b,1], Sabine Huhndorf[c], Vincent Robert[d], John L. Spouge[a], C. André Levesque[b], Wen Chen[b], and Fungal Barcoding Consortium[a,2]

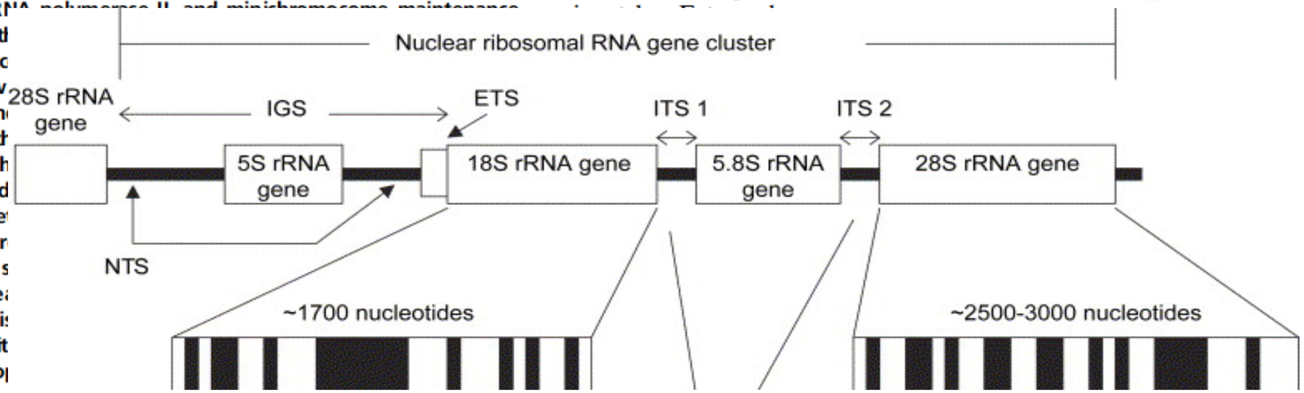
[a]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Healt
and Microbiology), Agriculture and Agri-Food Canada, Ottawa, ON, Canada K1A 0C6; [c]Department of Bota
[d]Centraalbureau voor Schimmelcultures Fungal Biodiversity Centre (CBS-KNAW), 3508 AD, Utrecht, The Ne

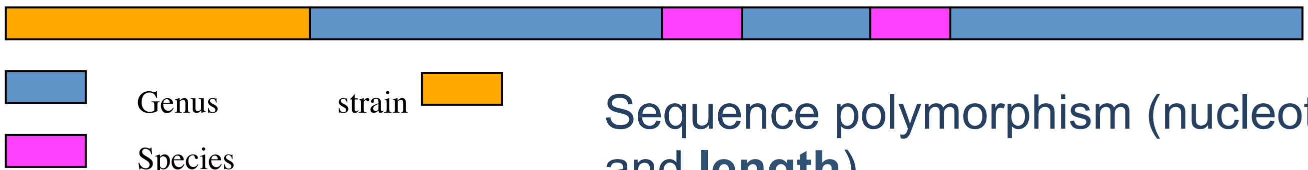
Edited* by Daniel H. Janzen, University of Pennsylvania, Philadelphia, PA, and approved February 24, 2012

Six DNA regions were evaluated as potential DNA barcodes for *Fungi*, the second largest kingdom of eukaryotic life, by a multinational, multilaboratory consortium. The region of the mitochondrial cytochrome *c* oxidase subunit 1 used as the animal barcode was excluded as a potential marker, because it is difficult to amplify in fungi, often includes large introns, and can be insufficiently variable. Three subunits from the nuclear ribosomal RNA cistron were compared together with regions of three representative protein-coding genes (largest subunit of RNA polymerase II, second largest subunit of R... protein). Alt... a higher perc... markers, low... them as can... regions of th... (ITS) region h... for the broad... code gap be... ribosomal lar... groups, had s... such as the e... was otherwi... small subunit... formally pro...

the intron of the *trn*K... reconsidering *CO1* as t... *CO1* functions reaso... genera, such as *Penicil*... species resolution (67... results in the few othe... consistent, and cloning... primers applicable to ... sess, because amplific...





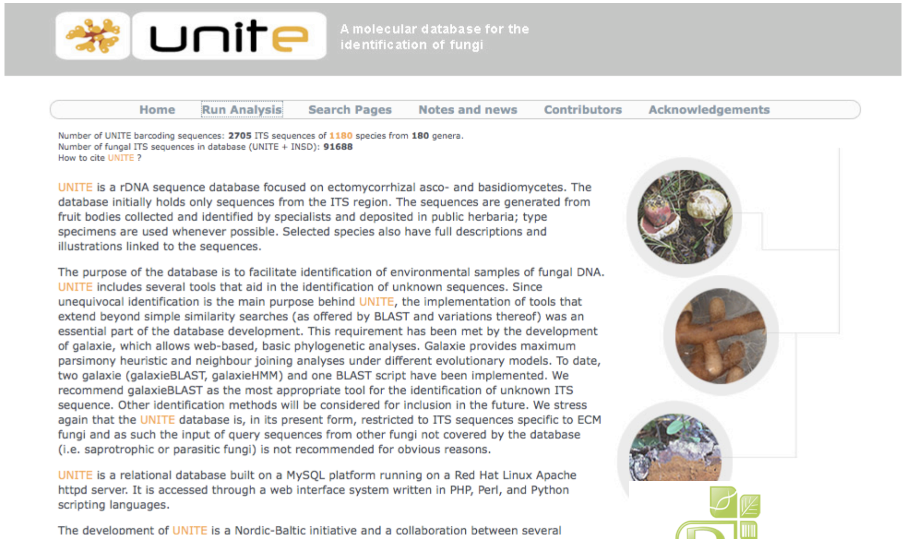


Sequence polymorphism (nucleotide and **length**)

Schoch et al. 2012

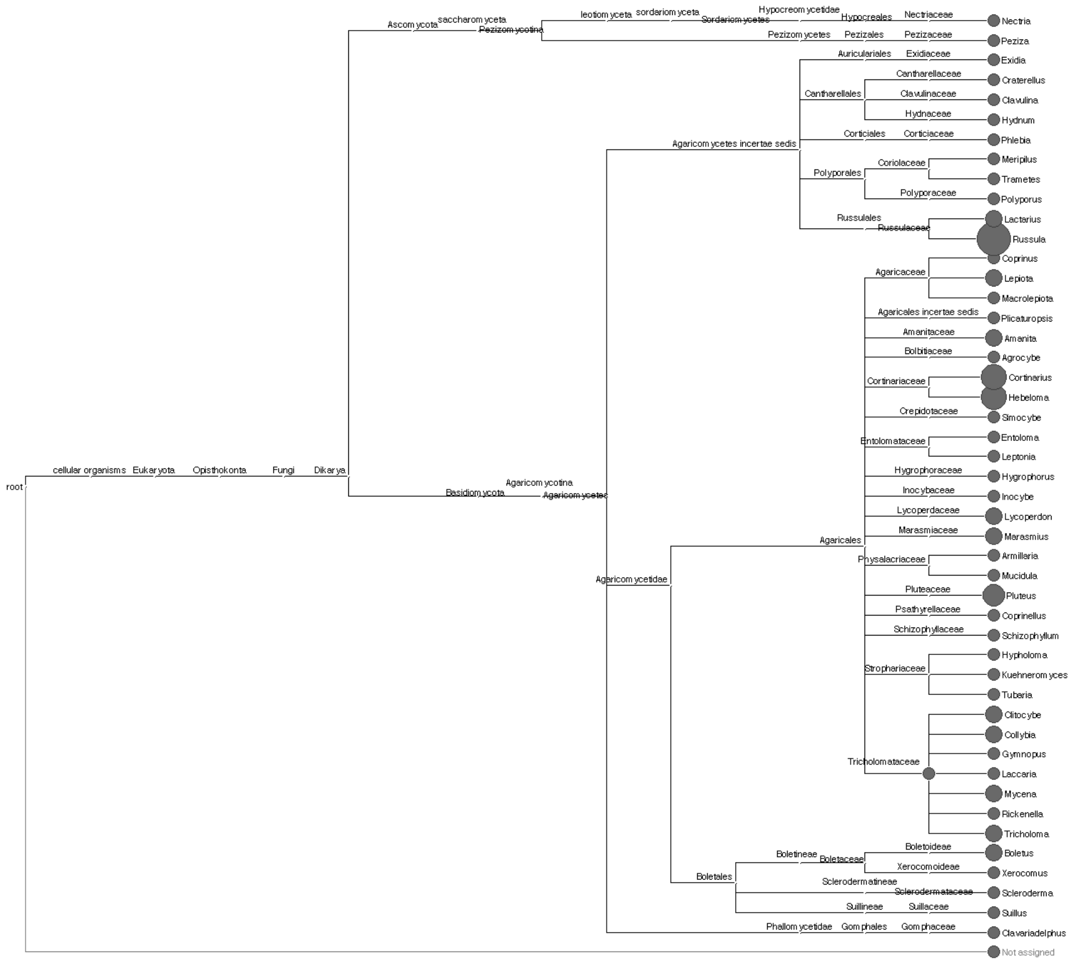
Implementation of high quality Fungal Database (ITS regions)

(annotated fruiting bodies, voucher specimens)



Production of Mock communities (positive controls in NGS process)

(annotated fruiting bodies, voucher specimens, fungal culture collections)

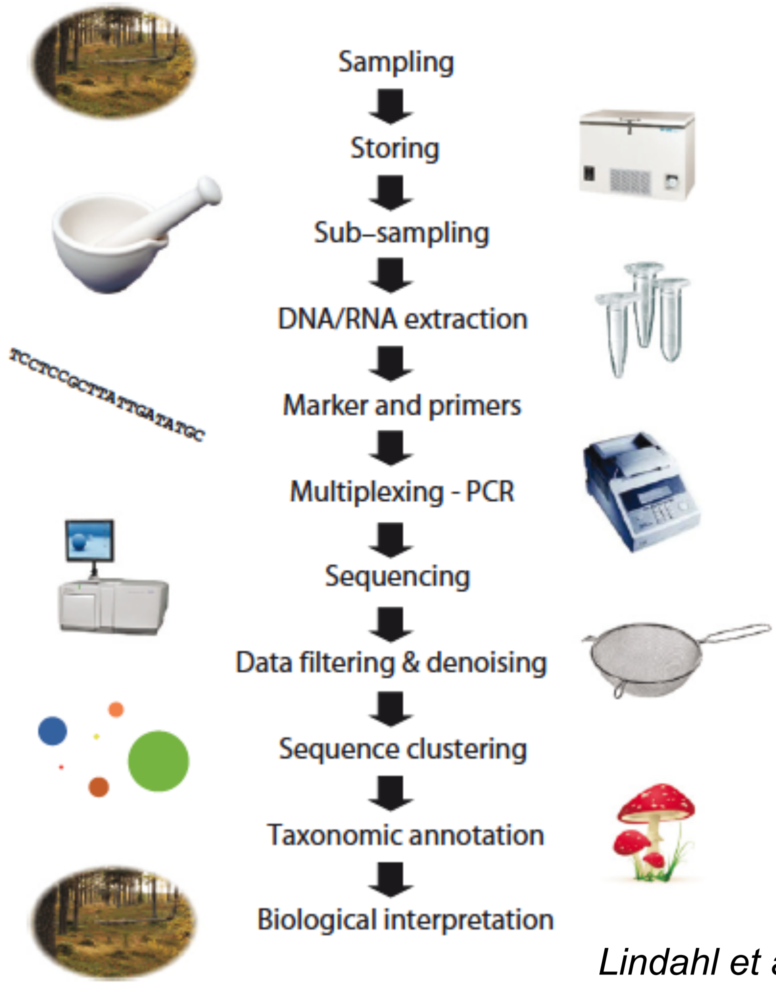


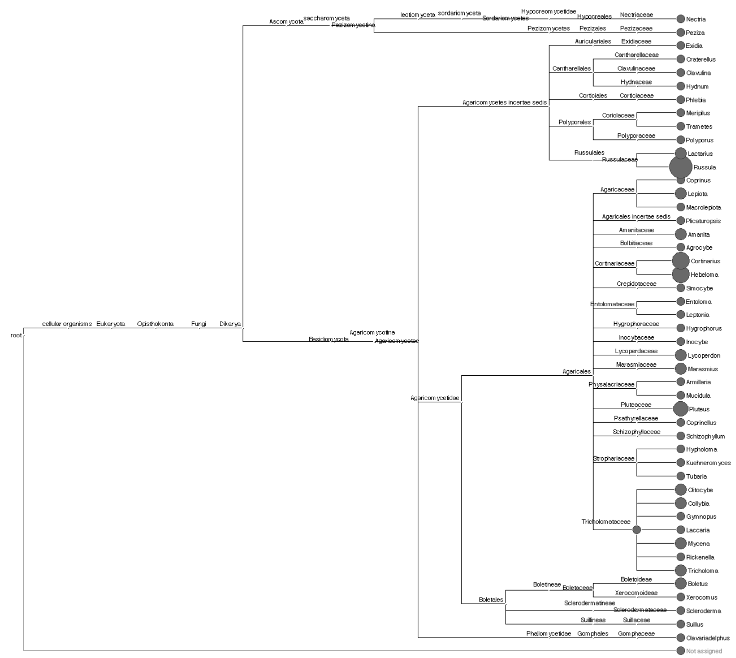
A mock community will also serve to inform the quality of the sequencing run (i.e. helps address run-to-run variation) and the processing steps necessary to retain the most data (i.e. addresses sequencing data quality).

This work was based on cultures from the INRA Nancy Culture Collection and dry voucher specimens (Mock community = 77 different species and 49 different genera)

Production of Mock communities (positive controls in NGS process)
(annotated fruiting bodies, voucher specimens, fungal culture collections)

A mock community will also serve to inform the quality of the sequencing run (i.e. helps address run-to-run variation) and the processing steps necessary to retain the most data (i.e. addresses sequencing data quality).



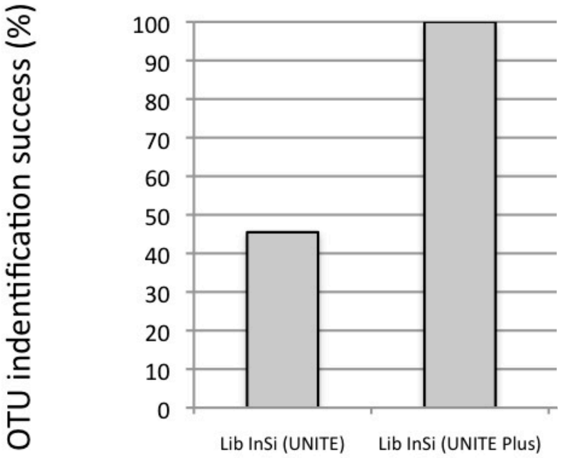


Lindahl et al. 2013

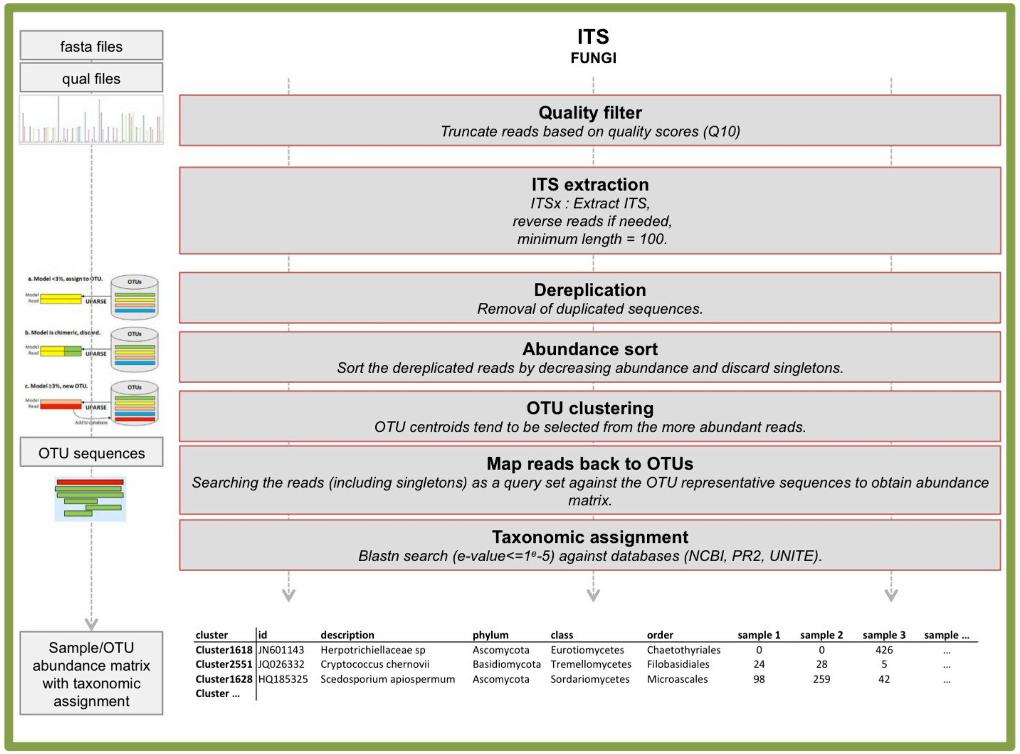
Overview of the steps involved in high-throughput sequencing of fungal communities.

Production of Mock communities

(annotated fruiting bodies, voucher specimens, fungal culture collections)



Bioinformatic pipeline v2 (UPARSE)

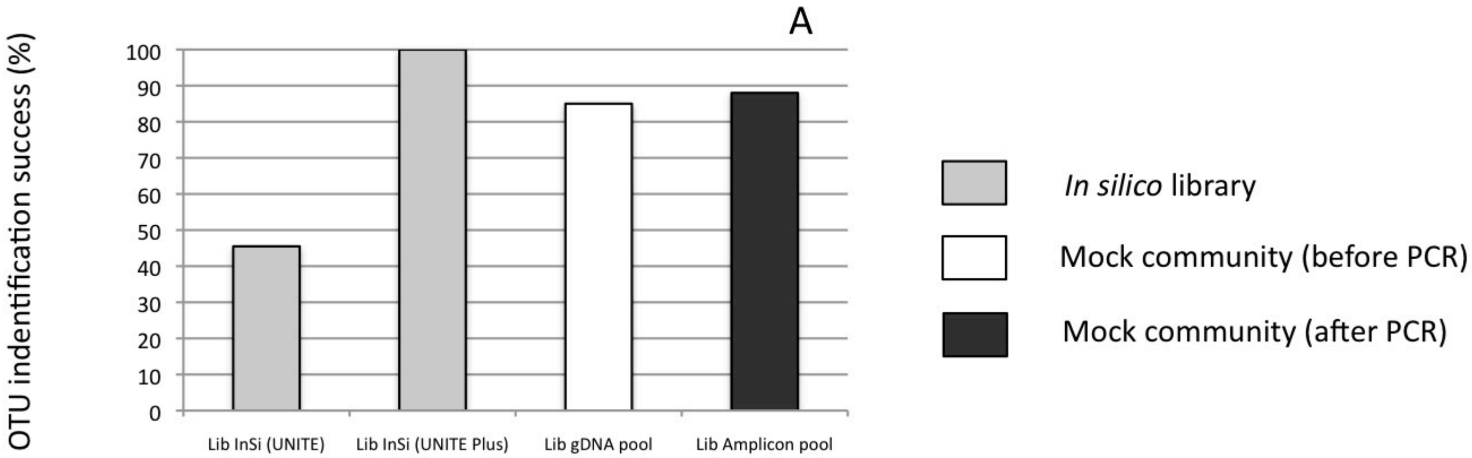


Bioinformatics analysis of a mock community from 77 fungal species using UPARSE (Edgar, 2013).
The rate of successful operational taxonomic unit (OTU) identification. Lib InSi corresponded to the in silico library of 77 Sanger sequences of ITS1 region.

Full "UNITE+INSD" dataset:
This FASTA file ("UNITE+INSDC") comprises "all" fungal ITS sequences of the UNITE and INSDC databases, updated and released some four times a year. Locked UNITE sequences and low quality (and overly short) INSDC sequences are however excluded. UNITE follows the Index Fungorum classification in nearly all regards.

Buée et al. in preparation

Production of Mock communities
(annotated fruiting bodies, voucher specimens, fungal culture collections)



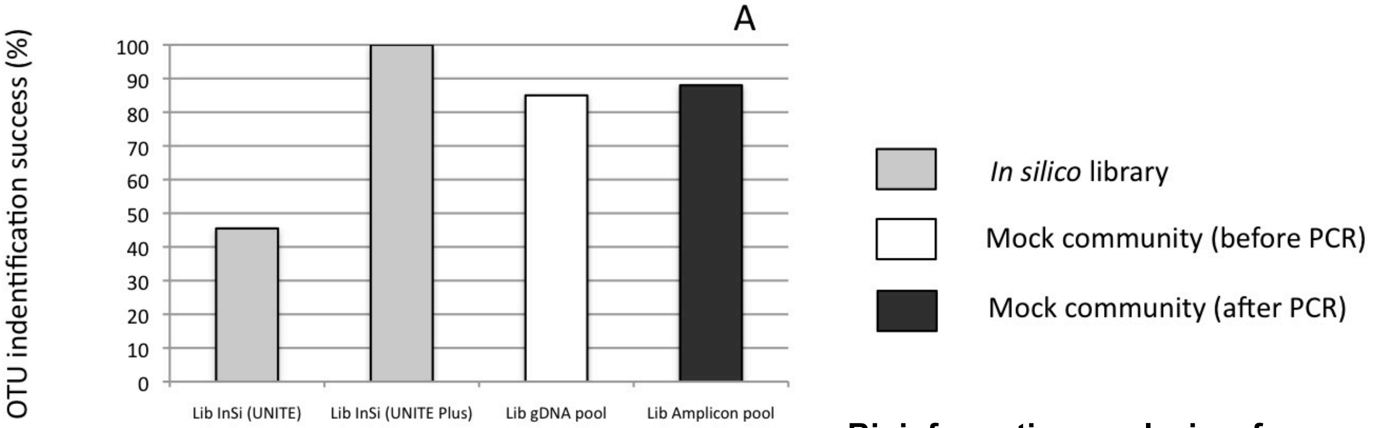
Bioinformatics analysis of a mock community from 77 fungal species using UPARSE (Edgar, 2013).

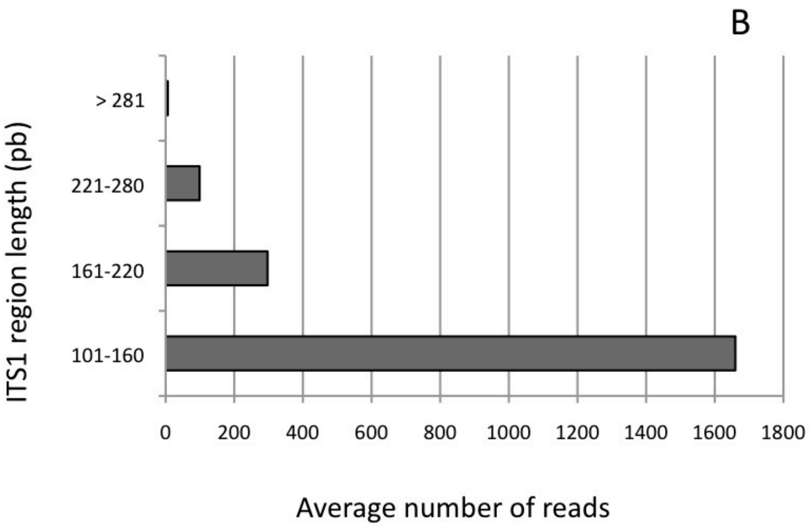
A. The rate of successful operational taxonomic unit (OTU) identification. Lib InSi corresponded to the in silico library of 77 Sanger sequences of ITS1 region. Lib gDNA pool was built by mixing equimolar amounts of genomic DNA from the 77 corresponding species, and Lib Amplicon pool by mixing equimolar amounts of independent PCR products from the same 77 species. "Lib gDNA" and "Lib Amplicon pool" were processed in the same way, and all the datasets were treated identically with the bioinformatics quality filtering and OTU clustering using UPARSE. The sequences were aligned using BLAST on "UNITE Plus" database (UNITE database [Kõljalg et al., 2013] together with the Sanger sequences of the mock community absent in the initial UNITE database). Only 45.5% of sequences of the mock community was present in the UNITE database.

Buée et al. in preparation

Production of Mock communities

(annotated fruiting bodies, voucher specimens, fungal culture collections)





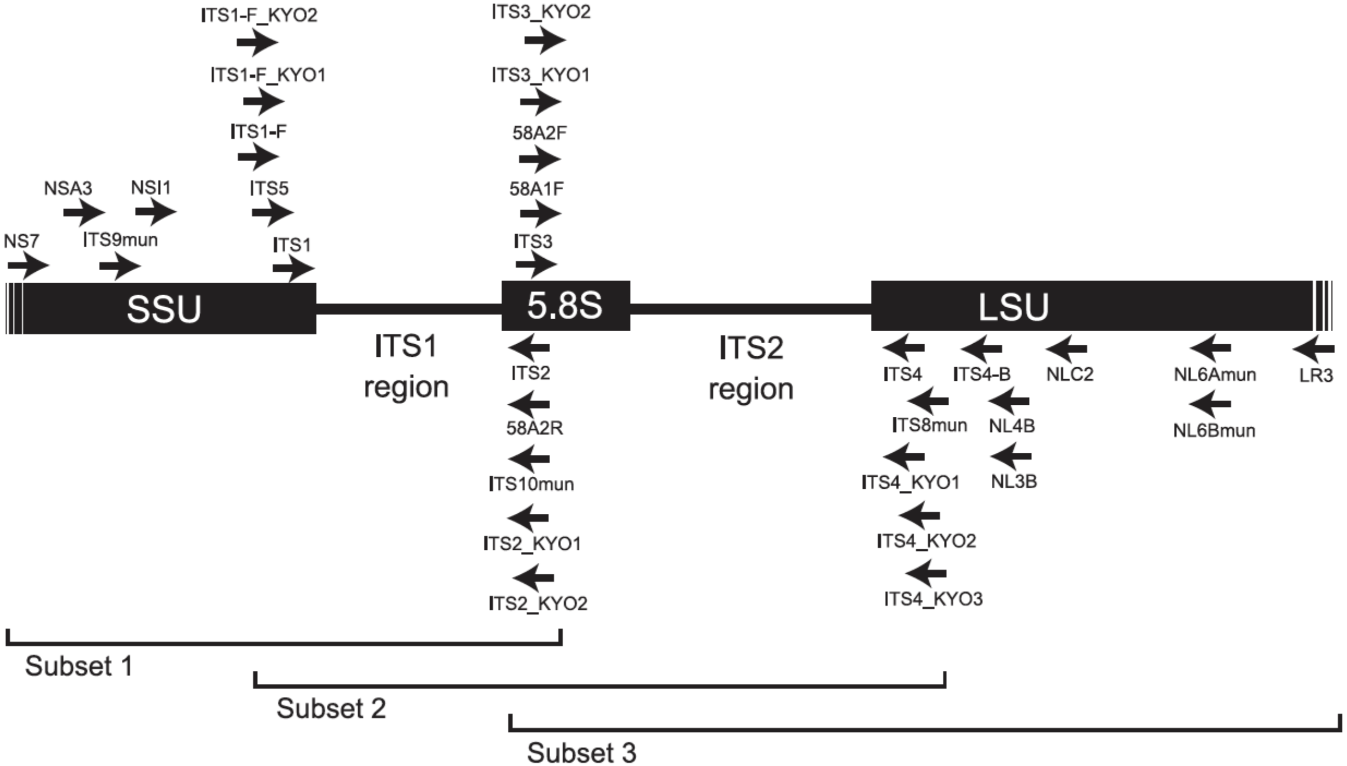
Bioinformatics analysis of a mock community from 77 fungal species using UPARSE (Edgar, 2013).

A. The rate of successful operational taxonomic unit (OTU) identification. Lib InSi corresponded to the in silico library of 77 Sanger sequences of ITS1 region. Lib gDNA pool was built by mixing equimolar amounts of genomic DNA from the 77 corresponding species, and Lib Amplicon pool by mixing equimolar amounts of independent PCR products from the same 77 species. "Lib gDNA" and "Lib Amplicon pool" were processed in the same way, and all the datasets were treated identically with the bioinformatics quality filtering and OTU clustering using UPARSE. The sequences were aligned using BLAST on "UNITE Plus" database (UNITE database [Kõljalg et al., 2013] together with the Sanger sequences of the mock community absent in the initial UNITE database). Only 45.5% of sequences of the mock community was present in the UNITE database.

B. Relationship between the number of reads generated by the Illunima MiSeq technology and the length of the fungal ITS1 fragment.

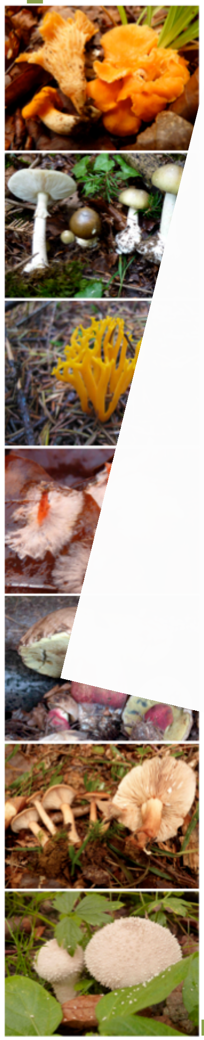
Buée et al. in preparation

Perspectives: validation of best primers and best regions (ITS2 vs ITS1 or SSU & LSU) for Miseq Illumina…



Map of nuclear ribosomal RNA genes and their ITS regions

*Perspectives: validation of best primers and best regions (**ITS2** vs ITS1 or SSU & LSU) for Miseq Illumina…*



ITS3_KYO2

MycoKeys 10: 1–43 (2015)
doi: 10.3897/mycokeys.10.4852
http://mycokeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
MycoKeys
Launched to accelerate biodiversity research

Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi

Leho Tedersoo[1], Sten Anslan[2], Mohammad Bahram[2,3], Sergei Põlme[1,2], Taavi Riit[2], Ingrid Liiv[2], Urmas Kõljalg[3], Veljo Kisand[4], R. Henrik Nilsson[5], Falk Hildebrand[6], Peer Bork[6], Kessy Abarenkov[1]

Map of nuclear ribo…

Metagenomics (or environmental genomics) is the study of genetic material recovered directly from environmental samples.

Who? Metabarcoding approach :
targeted genes (taxonomic barcodes or functional markers)

Active organisms: Metatranscriptomics: Living and active organisms and functional interactions: environmental RNA

Fungal barcode marker sequencing to produce a profile of diversity from natural samples

Focus on RNA, in particular fungal transcripts (only on the expressed genes)

Metagenomics and NGS

Putative roles: DNA metagenomics (“shotgun”)
"Shotgun" high-throughput sequencing (e.g. 454 pyrosequencing or Illumina) to get a maximum of genes from all organisms of the sampled communities

iNTERACTIONS
ARBRES-MICROORGANISMES

Metagenomics (or environmental genomics) is the study of genetic material recovered directly from environmental samples.

Who? Metabarcoding approach : targeted genes (taxonomic barcodes or functional markers)

Fungal barcode marker sequencing to produce a profile of diversity from natural samples

Active organisms: Metatranscriptomics: Living and active organisms and functional interactions: environmental RNA

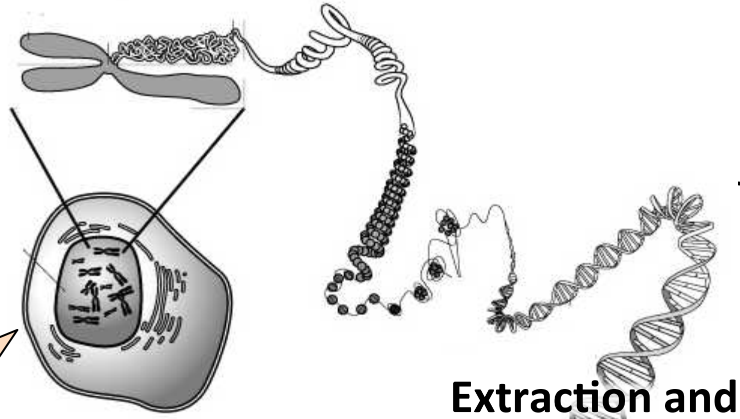
Focus on RNA, in particular fungal transcripts (only on the expressed genes)

Metagenomics and NGS

Putative roles: DNA metagenomics ("shotgun")
"Shotgun" high-throughput sequencing (e.g. 454 pyrosequencing or Illumina) to get a maximum of genes from all organisms of the sampled communities

INTERACTIONS
ARBRES-MICROORGANISMES

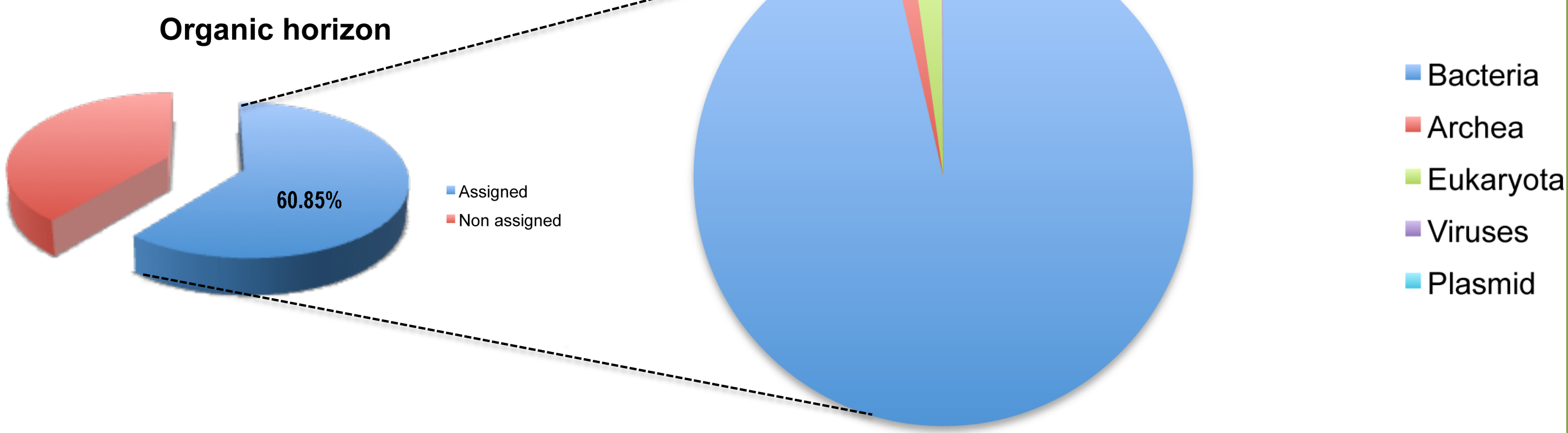
Forest soil metagenomics



Extraction and sequencing of genomes from all organisms (metagenome), which are present in the soil sample

Forest soil metagenomics

Shotgun Environmental 454 Sequencing



	Protein based	16s
Archaea	1.23% (4579)	0.00% (0)
Bacteria	91.32% (338916)	80.40% (201)
Eukaryota	1.58% (5851)	0.00% (0)
Virus	0.00% (0)	0.00% (0)
Other	5.87% (21783)	19.60% (49)

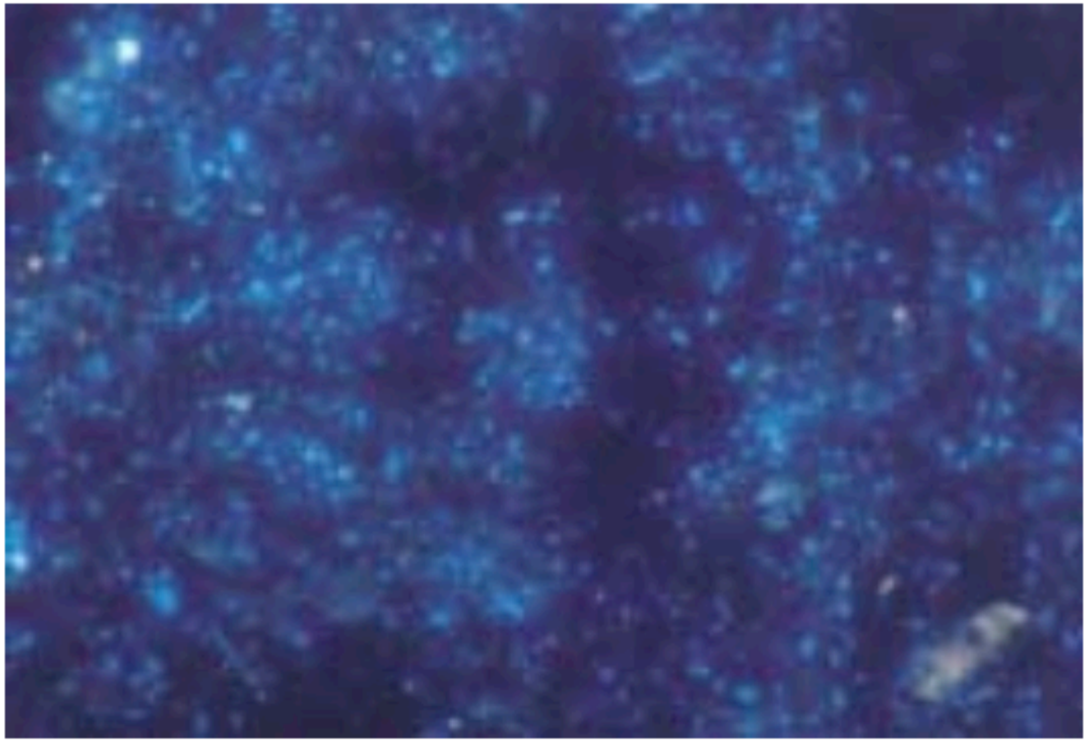
→ **Fungi :0,4% of total reads**

Most of the reads assigned belong to bacteria [Bacteria/Fungi ratio confirmed by qPCR]
Large % of orphan sequences

Uroz et al. 2013 Plos One

iNTERACTIONS ARBRES-MICROORGANISMES

The soil… A bacterial world! How study the fungal functions?



Current Opinion in Microbiology

Epifluorescence micrography of soil microorganisms stained with 4′,6-diamidino-2-phenylindole (DAPI). The total bacterial count was 4.2×10^{10} cells $gram^{-1}$ soil (dry weight) by fluorescent microscopy, and 4.2×10^{6} colony-forming units $gram^{-1}$ soil (dry weight) by plating.

Vigdis Torsvik and Lise Øvreås
Current Opinion in Microbiology 2002, 5:240–245

iNTERACTIONS ARBRES-MICROORGANISMES

<u>Metagenomics</u> (or environmental genomics) is the study of genetic material recovered directly from environmental samples.

Who? Metabarcoding approach : targeted genes (taxonomic barcodes or functional markers)

Fungal barcode marker sequencing to produce a profile of diversity from natural samples

Active organisms: Metatranscriptomics

Living and active organisms and functional interactions: environmental RNA

Focus on RNA, in particular fungal transcripts (only on the expressed genes)

Metagenomics and NGS

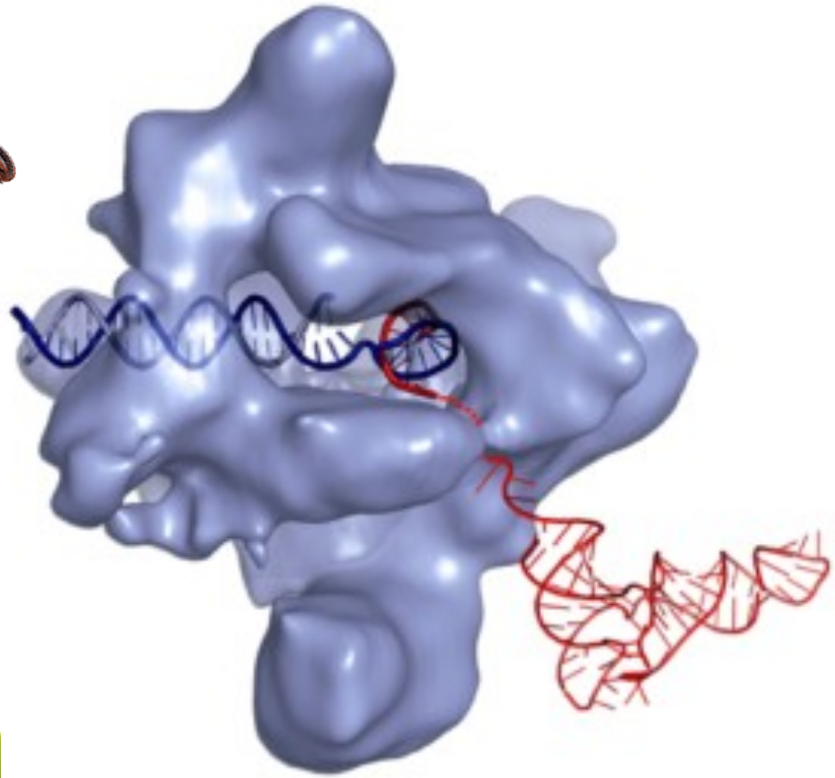
Putative roles: DNA metagenomics (“shotgun”)

"Shotgun" high-throughput sequencing (e.g. 454 pyrosequencing or Illumina) to get a maximum of genes from all organisms of the sampled communities

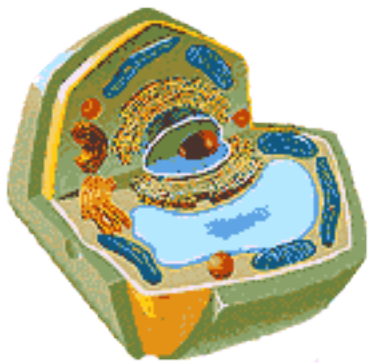
iNTERACTIONS ARBRES-MICROORGANISMES

Eukaryotic and fungal metagenomic : focus on poly A+ transcripts = "Metatranscriptomic" approaches

Use only polyA+ transcripts and "concentrate" eukaryotic transcripts (only on the expressed genes).
Investigate the functioning of the ecosystem and the role of fungal species in soil forest



©EMBL/C. Fernando-Tornero



The poly A+ transcripts include mRNA, microRNA and snoRNA generated by RNA polymerase II

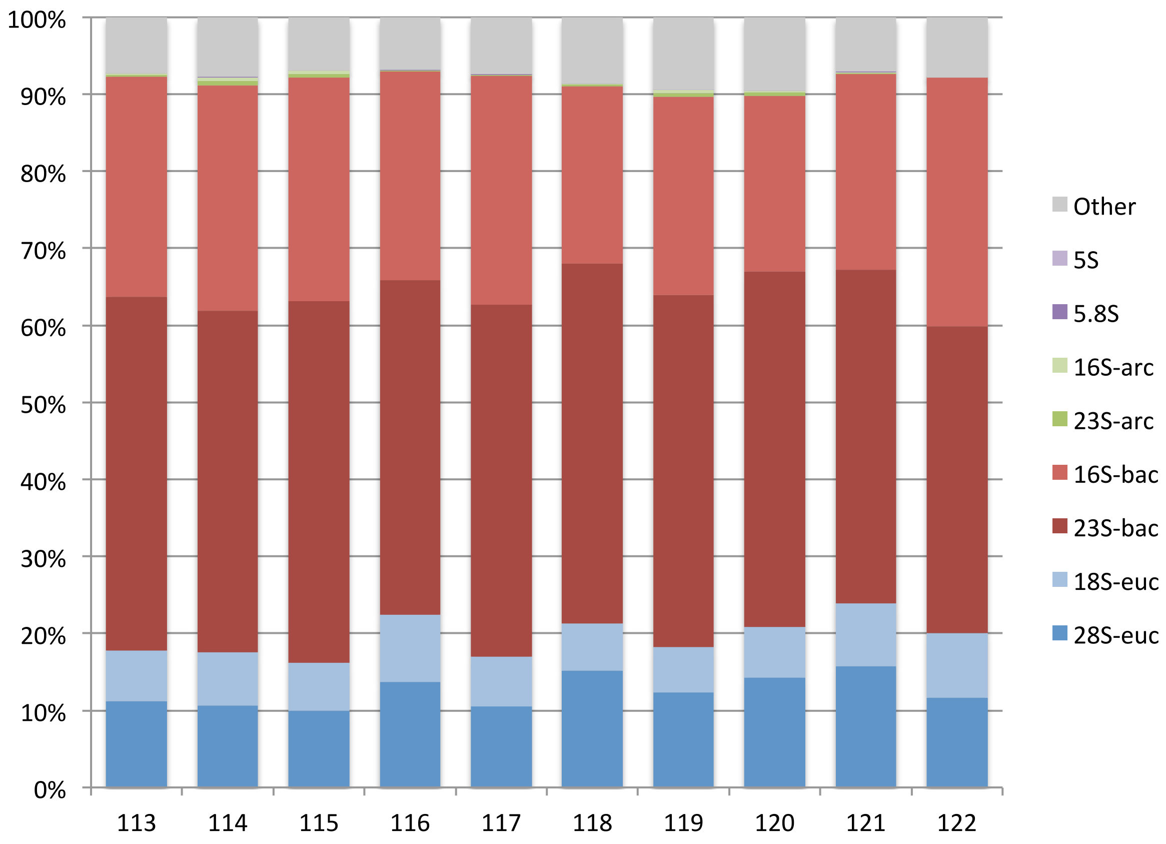
iNTERACTIONS
ARBRES-MICROORGANISMES

The Community Sequencing Program 2012

Metatranscriptomics of Forest Soil Ecosystems (cood. F. Martin)

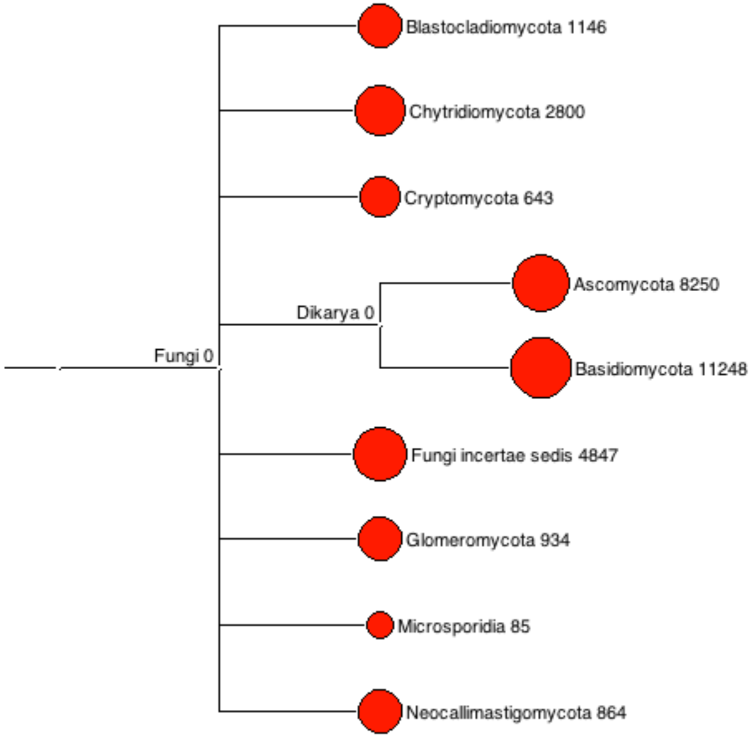
Proposal ID: CSP # 570
Eight laboratories in the world (11 sites)

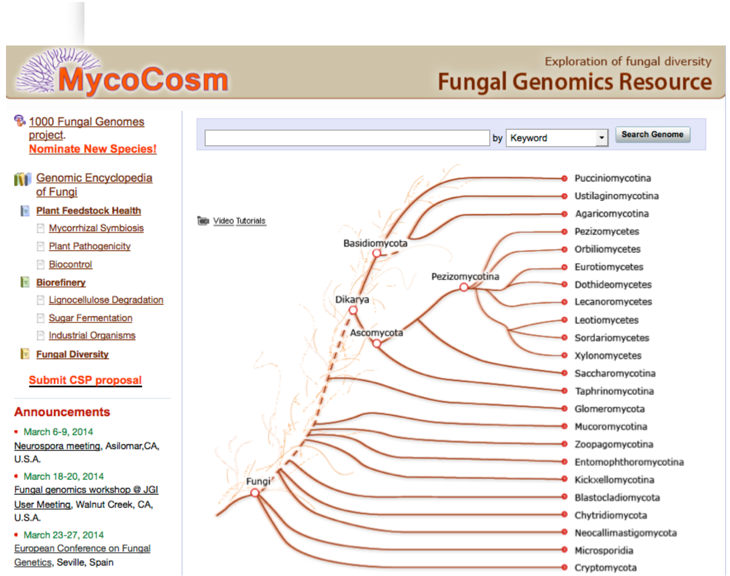
Analysis of rRNA reads (SortMeRNA) extracted from forest soils: Beech and Spruce forests (spring and autumn) - Illumina sequencing – (Pilot study: ANR Eumetasol)



About 20% of sequences corresponded to active EuK (only 1.6 % of genomic DNA)

Site Amance (54): Oak, Spring (Org) Illumina sequencing of soil mRNA (polyA)





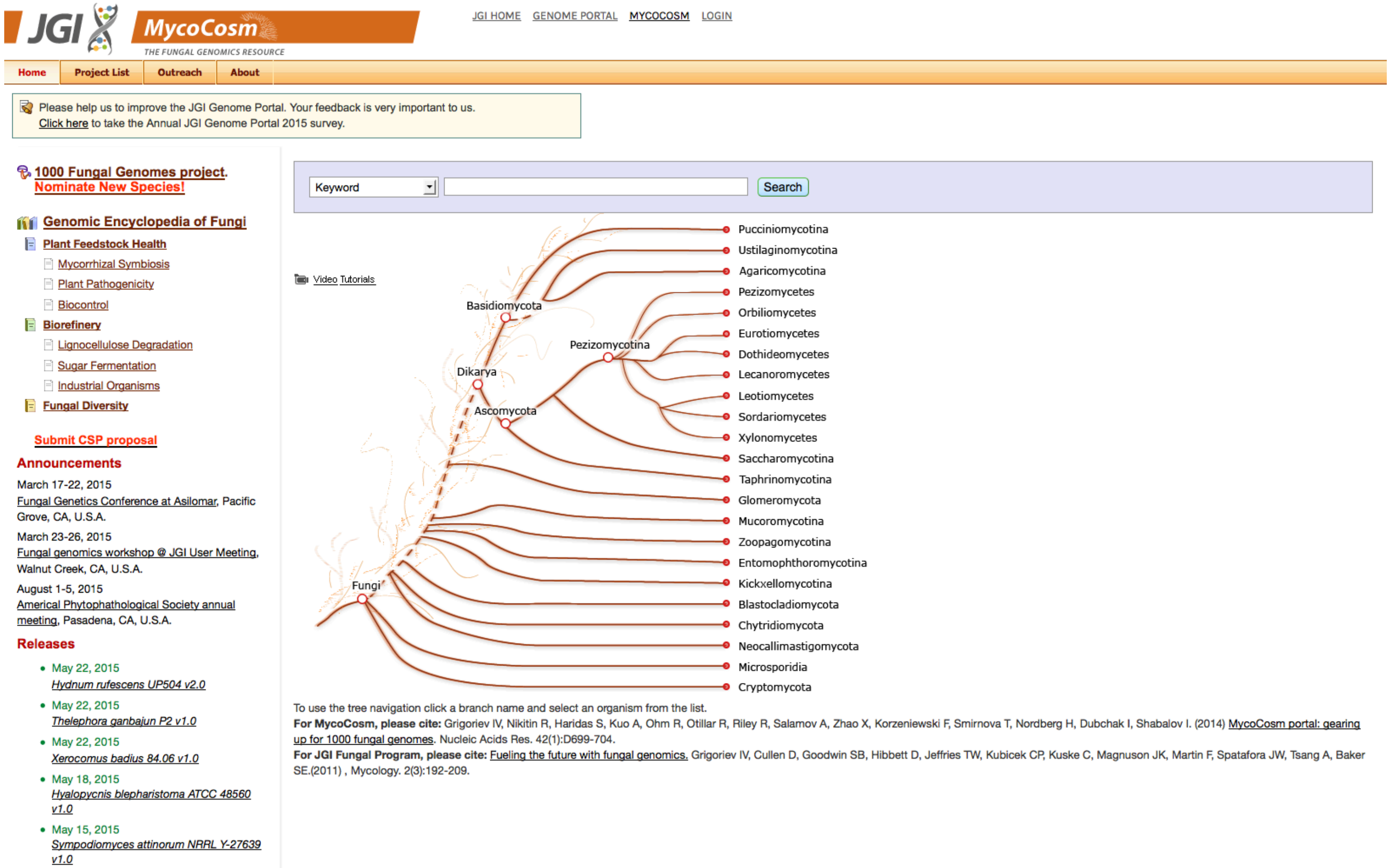
Reads = sequenced EuK mRNA
Blast on Mycocosm (JGI)

Pilot analysis (Miseq):
14M reads were assembled into 72226 contigs, median length - 238bp (**only on mRNA**)

31205 (43.2%) of contigs have hits to mycocosm proteins by blastx at Evalue threshold 1e-05

iNTERACTIONS ARBRES-MICROORGANISMES

In course: assignment of reads on fungal genomic database (JGI 1000fungalgenome project)



Analysis of Amance/Champenoux site (36 samples):
3.439.190.128 reads (Hiseq) from 36 samples, or 95.5 million reads on average per sample

Ecogenomics of tree / fungal interactions

FUNGAL COLLECTIONS

GENOMICS

(1ST models)

New Tools: NGS and bioinformatics

FUNGAL ECOLOGY

Fungal community description = environment model species (most abundant…): pan-genome

GENOMICS

e.g. 1000genomesproject

GENOMICS

e.g. Poplar, Spruce, Oak…

Catalogues of Fungal genomes

➔ « Ecogenomics »

METATRANSCRIPTOMICS (annotation)

UNIVERSITÉ DE LORRAINE

BACCARA

UE BACCARA project

INRA, Nancy
A. COINCE
B. MARCAIS
M. BUEE

CSIC, Madrid
A.RINCON

Collection of soil samples
INRA, Bordeaux France
SLU, Uppsala Sweden
CEMAGREF, Grenoble France
UNITUS, Viterbo Italy
FRI, Raszyn Poland

ANR Eumetasol project

INRA, Nancy
M. BUEE
F. MARTIN
E. SENTOSA

Univ Lyon
P. Luis
R. Marmeisse

SUPPORTED BY ANR

JGI CSP 2012

INRA, Nancy
F. MARTIN
E. SENTOSA
A. KOHLER
M. BUEE
JGI collaborators

JGI DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

IAM iNTERACTIONS ARBRES-MICROORGANISMES

INRA SCIENCE & IMPACT


THANK YOU FOR YOUR ATTENTION