The potential of the metagenomics approach

Carmen Morales Rodríguez
cmorales@unitus.it

UNIVERSITÀ DEGLI STUDI DELLA TUSCIA

UB3GUARD

Traditional Microbiology: Limitations and Challenges

Traditional microbiology has relied heavily on culturing techniques for over a century, allowing scientists to isolate and study individual microbial species. These approaches have formed the foundation of our understanding of microbial taxonomy, physiology, and genetics.

The "Great Plate Count Anomaly"

Only about 1% of environmental microorganisms can be successfully cultured in laboratory settings.

Growth Requirements

Many microorganisms have specific, often unknown, growth requirements including nutrients, oxygen levels, pH, temperature, and symbiotic relationships that are difficult to replicate in vitro.

Slow-Growing Organisms

Some microorganisms, particularly those adapted to nutrient-poor environments, grow extremely slowly and may require weeks or months to form visible colonies, making traditional approaches impractical.

The Emergence of Metagenomics

- Metagenomics emerged in the early 2000s as a response to the fundamental limitations of culture-dependent microbiology. The term "metagenomics" was coined to describe the analysis of genetic material recovered directly from environmental samples.

- This approach revolutionised our ability to study microbial communities by bypassing the need for laboratory cultivation, thereby providing access to the "microbial dark matter"—the vast majority of microorganisms that had previously remained invisible to scientific inquiry.

Key Historical Developments

- 1985: First direct cloning of environmental DNA
- 1998: First use of the term "metagenomics"
- 2004: First large-scale metagenomic study of the Sargasso Sea
- 2006: Introduction of next-generation sequencing technologies
- 2010s: Development of sophisticated bioinformatic tools for metagenomic analysis

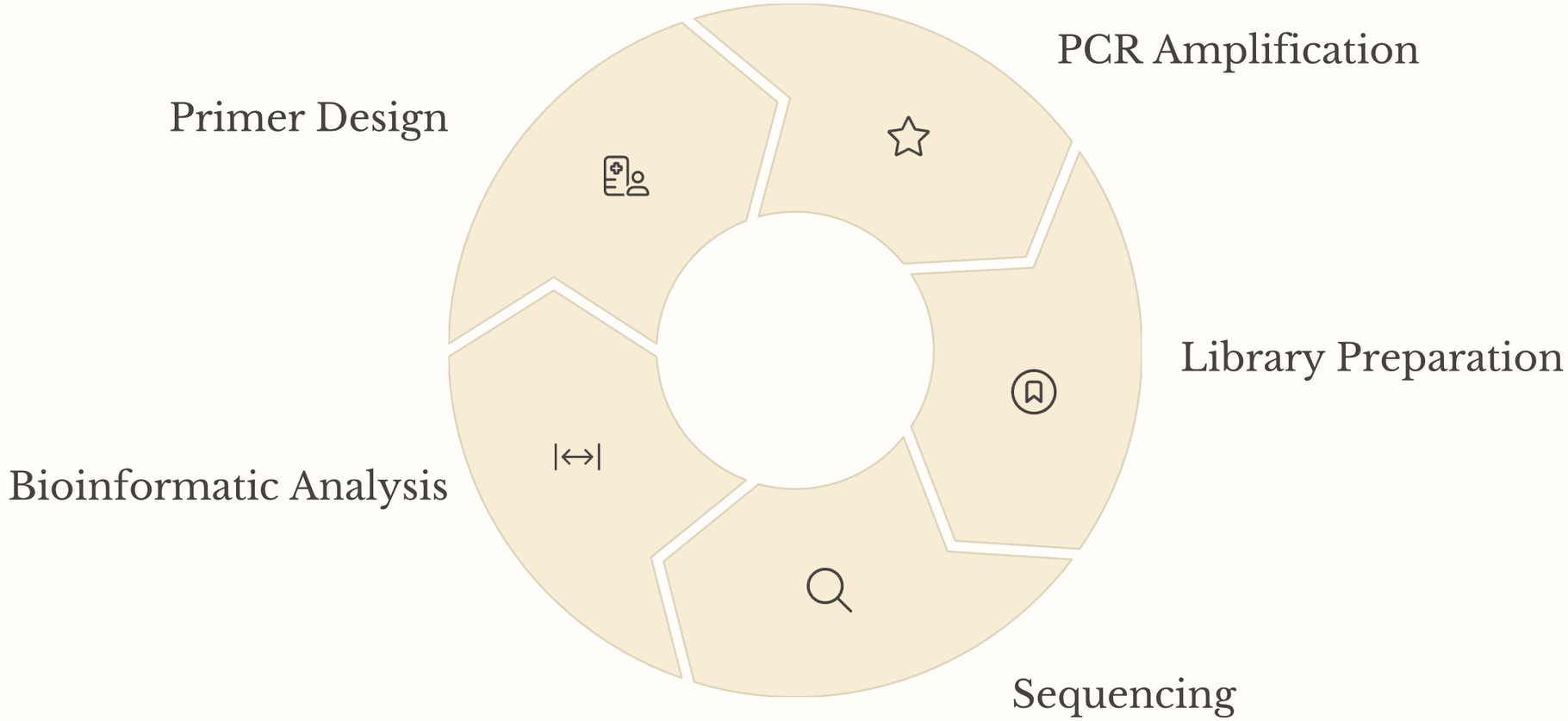
Fundamental Concepts in Metagenomics

Metagenomics fundamentally transforms how we approach microbial communities, shifting from isolating individual species to analysing entire communities simultaneously. This approach requires sophisticated molecular and computational techniques to extract meaningful biological insights.

- Sample Collection
- DNA Extraction
- Sequencing
- Bioinformatic Analysis

Amplicon-Based Metagenomics: Methodology

Amplicon-based metagenomics, also known as metabarcoding, involves PCR amplification of specific marker genes followed by high-throughput sequencing. This approach targets conserved regions that vary sufficiently between species to allow taxonomic discrimination.



Common Marker Genes in Amplicon Metagenomics

The selection of appropriate marker genes is crucial for successful amplicon-based metagenomic studies. These genetic regions must balance conservation (for reliable PCR amplification) with variation (for taxonomic discrimination).

Taxonomic Group	Common Marker Genes	Key Characteristics
Bacteria	16S rRNA gene	Nine hypervariable regions (V1-V9); V3-V4 and V4 commonly targeted
Archaea	16S rRNA gene	Different primer sets from bacteria; often requires specific primers
Fungi	ITS region, 18S rRNA, 28S rRNA	ITS1 and ITS2 regions most common; variable length can cause bias
Plants	rbcL, matK, trnH-psbA	Chloroplast genes used for plant DNA barcoding and identification
Animals	COI, 18S rRNA, 12S rRNA	COI is standard for metazoan DNA barcoding; 12S for vertebrates
Protists	18S rRNA, COI	Highly diverse group requiring multiple markers for comprehensive coverage

Reference Databases for Amplicon Metagenomics

The accuracy of taxonomic assignments in amplicon-based metagenomics depends heavily on the quality, comprehensiveness, and curation of reference databases. Several specialised databases have been developed for different taxonomic groups and marker genes.

SILVA Database

Provides quality-checked, aligned ribosomal RNA sequences for bacteria, archaea, and eukaryotes. Contains over 9 million 16S/18S sequences and is regularly updated. Particularly strong for environmental sequences.

Greengenes

Specialised for bacterial and archaeal 16S rRNA genes with approximately 1.8 million sequences. Uses a hierarchical taxonomy based on de novo phylogenetic tree construction, though updates have been infrequent.

UNITE Database

Focused on fungal ITS sequences with approximately 1.5 million sequences. Implements the concept of species hypotheses (SH) to address the incomplete fungal taxonomy and provides dynamic taxonomic assignment thresholds.

RDP (Ribosomal Database Project)

Provides ribosomal RNA sequences and tools for analysis. Contains bacterial and archaeal 16S rRNA, fungal 28S rRNA, and fungal ITS regions with approximately 3.4 million sequences.

Shotgun Metagenomics: Methodology

Shotgun metagenomics involves sequencing all DNA present in an environmental sample without targeting specific genes. This approach provides a comprehensive view of both taxonomic composition and functional potential of microbial communities.

DNA Extraction and Fragmentation

Total genomic DNA is extracted from the sample and randomly fragmented to sizes appropriate for the sequencing platform (typically 300-500 bp).

Library Preparation

DNA fragments are processed to create sequencing libraries by adding platform-specific adapters and barcodes for sample identification.

Sequencing

High-throughput sequencing generates millions to billions of reads covering the entire genomic content of the sample, including all organisms present.

Quality Control and Preprocessing

Raw sequences undergo quality filtering, adapter trimming, and removal of host DNA contamination if relevant (e.g., plant DNA in rhizosphere samples).

Computational Analysis

Sophisticated bioinformatic pipelines perform taxonomic classification, functional annotation, and potentially metagenomic assembly to reconstruct partial or complete genomes.

Reference: Quince et al. (2017). Shotgun metagenomics, from sampling to analysis.

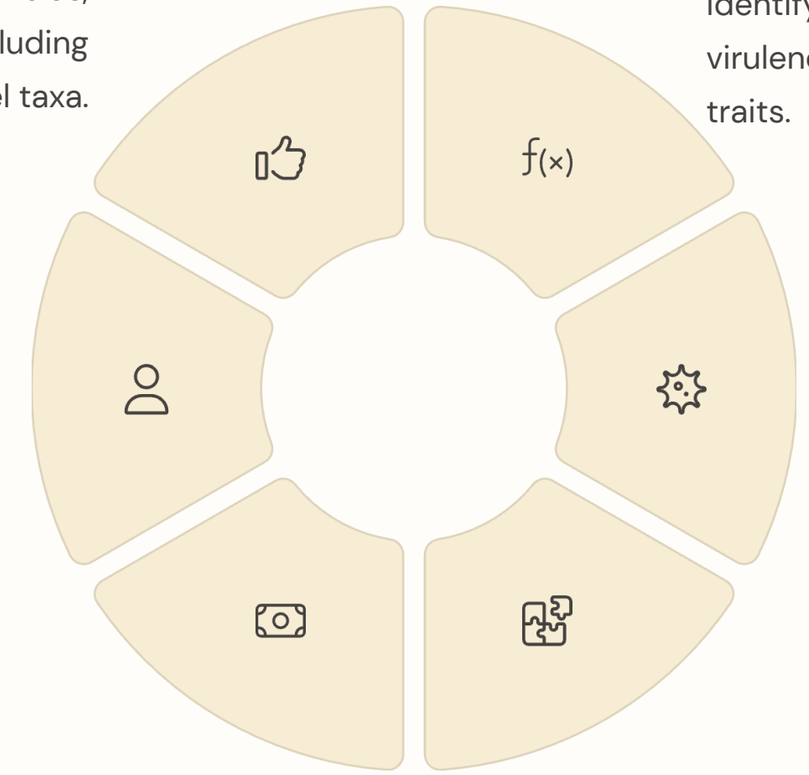
Advantages and Limitations of Shotgun Metagenomics

Unbiased Community Sampling

No PCR amplification step means no primer bias, allowing detection of all organisms including those with divergent sequences or novel taxa.

Functional Insights

Reveals the metabolic potential of communities by identifying genes involved in biogeochemical cycles, virulence factors, antibiotic resistance, and other functional traits.



Complex Analysis

Data analysis is challenging, requiring sophisticated bioinformatic pipelines and significant expertise in computational biology.

Virus Detection

Captures viral communities that lack conserved marker genes and would be missed by amplicon approaches.

Higher Cost

Requires substantially more sequencing depth and computational resources than amplicon approaches, limiting sample throughput.

Genome Reconstruction

Enables assembly of metagenome-assembled genomes (MAGs) from uncultured organisms, expanding genomic databases and ecological understanding.

Computational Challenges in Shotgun Metagenomics

The computational analysis of shotgun metagenomic data presents substantial challenges due to the volume and complexity of sequence information. Several key computational bottlenecks must be addressed for successful analysis.

1 Data Volume

A single shotgun metagenomic sample can generate hundreds of gigabytes of raw sequence data, requiring substantial computational infrastructure for storage and processing. High-performance computing clusters or cloud computing resources are often necessary.

2 Sequence Assembly

De novo assembly of metagenomic data is complicated by uneven coverage of different organisms, strain-level variation, and the presence of repetitive regions. Specialised assemblers like MEGAHIT and metaSPAdes use novel algorithms to address these challenges.

3 Binning

Assigning assembled contigs to their source organisms (binning) is difficult in complex communities. Methods rely on sequence composition statistics, coverage patterns across samples, or phylogenetic markers to group related sequences.

4 Functional Annotation

Predicting gene functions in novel organisms is challenging, particularly for genes without close homologs in reference databases. Machine learning approaches are increasingly used to improve annotation accuracy.

Applications in Environmental Sciences

Metagenomics has revolutionised environmental sciences by providing unprecedented insights into microbial communities across diverse ecosystems. These approaches enable researchers to monitor environmental health, track pollution impacts, and understand biogeochemical cycles at the molecular level.

Soil Health Assessment

Aquatic Ecosystem Monitoring

Climate Change Research

Applications in Plant Health Sciences

Early Pathogen Detection

Metagenomic surveillance can identify plant pathogens before visible symptoms appear, enabling preventive action. This is particularly valuable for slow-developing diseases where early intervention is critical for control.

Cryptic Pathogen Identification

Metagenomics can distinguish morphologically similar pathogens with different virulence profiles, such as closely related Phytophthora species or Fusarium strains with different host ranges.

Microbiome Characterisation

Analysis of phyllosphere (leaf surface), rhizosphere (root zone), and endosphere (internal tissue) microbial communities reveals complex interactions that influence plant health and productivity.

Disease Suppression

Identification of beneficial microorganisms that contribute to natural disease suppression informs biological control strategies and the development of microbial inoculants.

Biodiversity and Biosecurity Applications

Biodiversity Assessment

Metagenomics enables comprehensive biodiversity surveys across all domains of life, from microorganisms to macrofauna, without taxonomic bias. This approach is particularly valuable in remote or inaccessible environments where traditional sampling is challenging.

Environmental DNA (eDNA) metabarcoding can detect rare or cryptic species from trace DNA in soil, water, or air samples, providing non-invasive monitoring for conservation.

Biosecurity Applications

- Detection of non-native pathogens in imported plant material
- Monitoring of insect pests in wood packaging and shipping containers
- Surveillance of ballast water for invasive aquatic organisms
- Assessment of seed lots for contamination with weed seeds or pathogens
- Early warning systems at points of entry and high-risk sites
- Tracking the origin of invasive species through population genomics

These applications are increasingly integrated into national biosecurity frameworks, providing rapid, sensitive detection capabilities that complement traditional inspection methods. Next-generation sequencing platforms now enable on-site testing at ports and borders, reducing response times for potential threats.

Challenges and Considerations in Metagenomic Studies

Despite its transformative potential, metagenomics presents significant challenges that must be carefully addressed in experimental design and data interpretation. Awareness of these limitations is essential for generating reliable, reproducible results.

Sampling Strategies

Environmental heterogeneity can lead to significant variation between samples, even from the same site. Careful consideration of spatial and temporal sampling designs, including appropriate replication, is crucial for capturing representative community profiles and detecting meaningful patterns.

DNA Extraction Biases

Different extraction methods preferentially recover DNA from certain organisms whilst potentially missing others. Cell wall composition, particularly in fungi and gram-positive bacteria, affects lysis efficiency. Standardised protocols and positive controls are essential for comparing across studies.

Reference Database Limitations

Taxonomic and functional assignments depend heavily on reference database quality and completeness. Underrepresented groups, particularly in non-model environments like tropical forests or extreme habitats, may be misclassified or remain unidentified.

Statistical Challenges

Metagenomic datasets are high-dimensional with complex correlation structures. Appropriate statistical approaches must account for compositional data constraints, batch effects, and multiple testing issues to avoid spurious conclusions.

Metabarcoding used in Sentinel systems

- Analysis of non-symptomatic material in sentinel arboreta, nurseries, plantations.
- Establishment of baseline in centinela plantings
- Study the biodiversity of insects from traps
- Study spore tramps

Pyrosequencing 454 GS-FLX and Illumina MiSeq

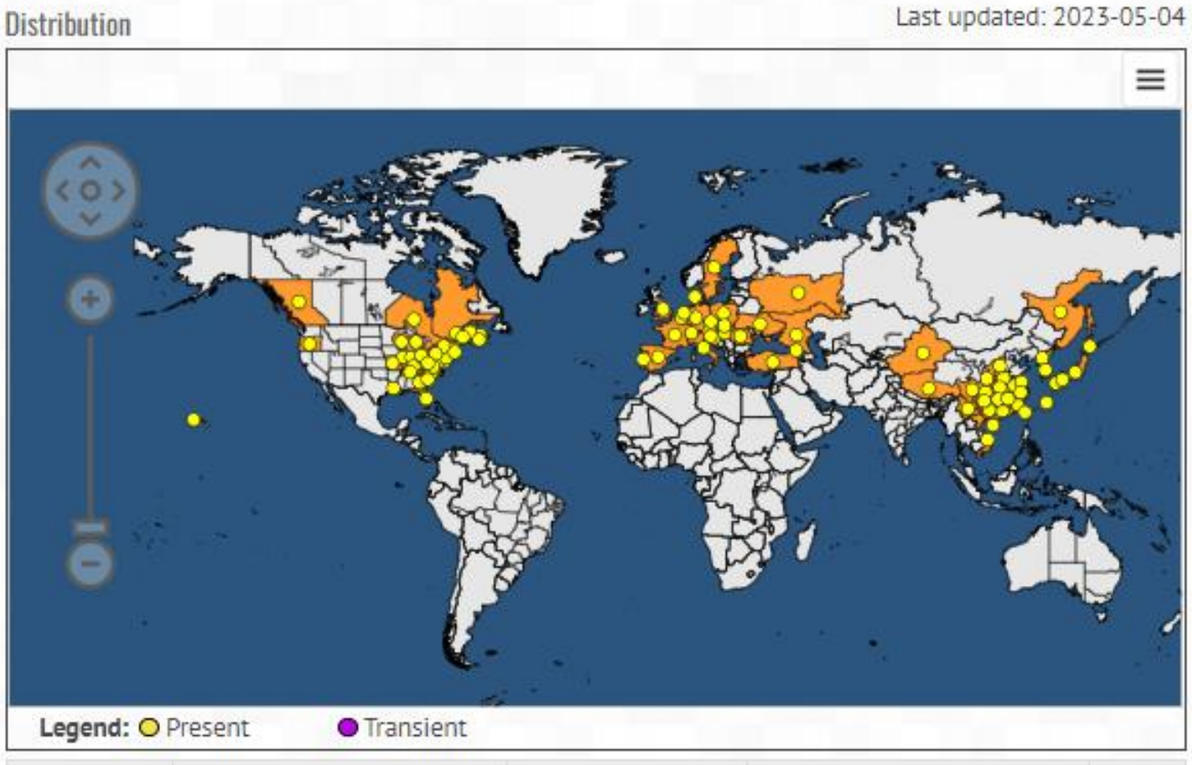
Alien invasive inside the alien invasive

Xylosandrus germanus - Chestnut

- *X. germanus*, was first reported in northern Italy in 1992 in stands dominated by chestnuts (*Castanea sativa* Miller; Bernabo, 2000; Dutto et al., 2018), then in walnut plantations (Juglans spp.), in mixed broadleaf stands and in apple orchards (Stergulc et al., 1999).

- *X. germanus* is currently not regulated in the EPPO region but is a quarantine pest in the Comite de Sanidad Vegetal del Cono Sur (COSAVE) region (EPPO, 2018).

Xylosandrus germanus distribution



EPPO Global Database

Castanea sativa distribution

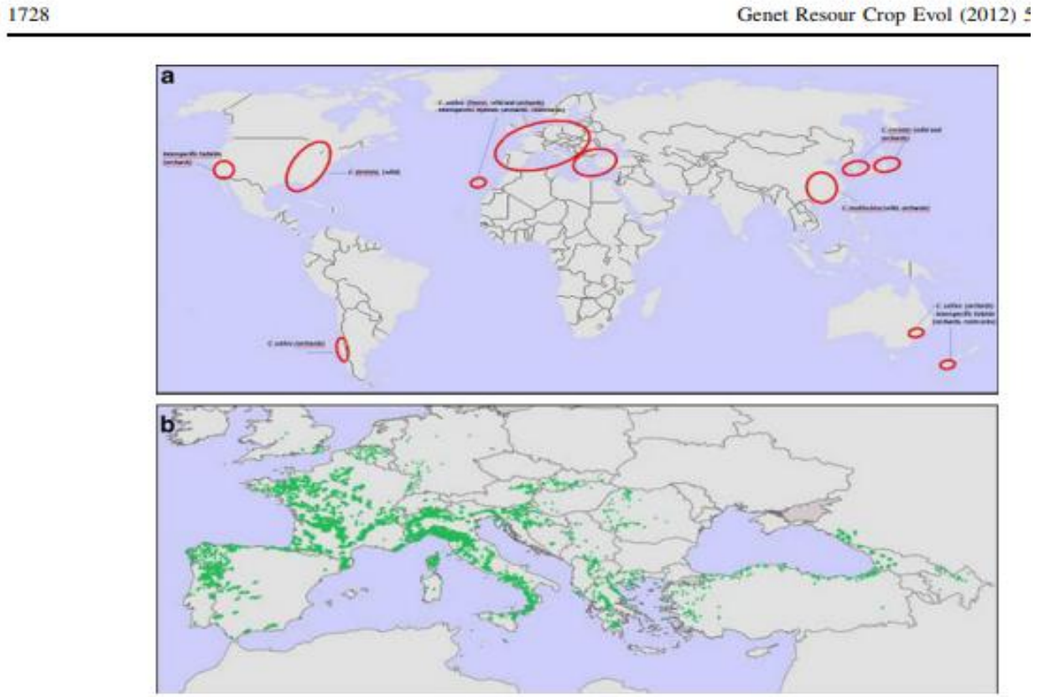


Fig. 1 Main areas of chestnut (*Castanea* spp.) distribution in the world **a** and in Europe **b**

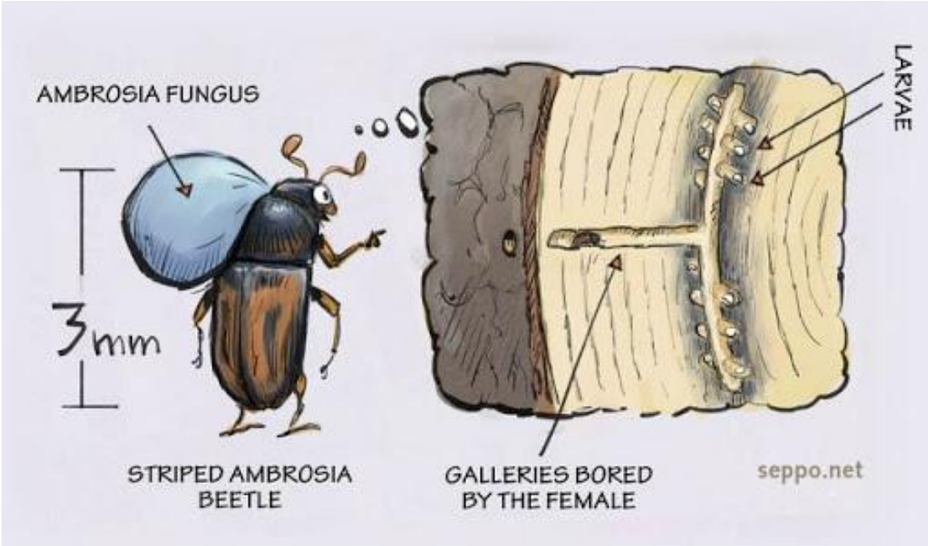
Mellano et al., 2012

- The distribution of the insect coincides with the distribution of *Castanea* sp.
- Few reports

Ambrosia beetles

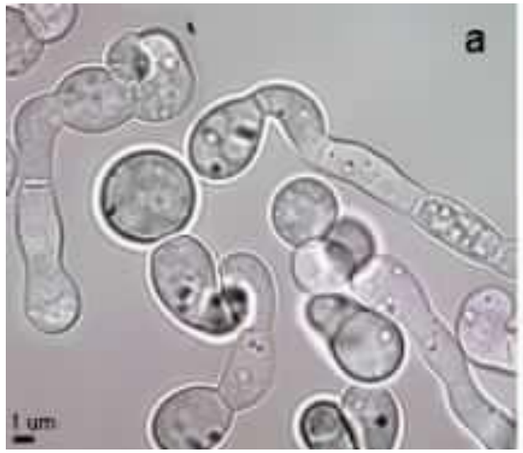
An example of mutualistic symbiosis

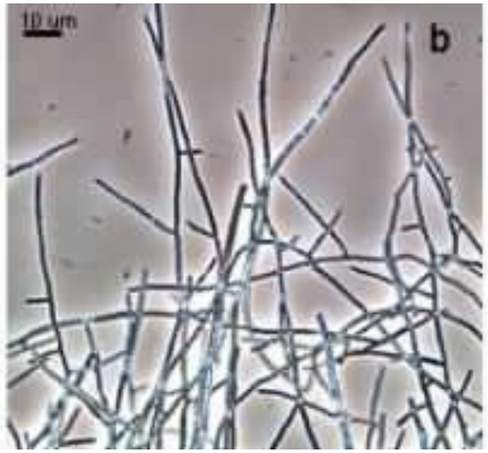
Ambrosia beetles carry tightly symbiotic fungi to a specialized structure called mycangium

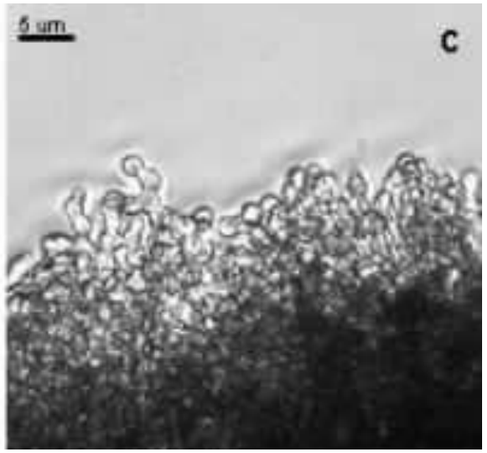


Fungus farming

An example of polymorphism: *Ambrosiella hartigi* associated with *Xylosandrus germanus*.



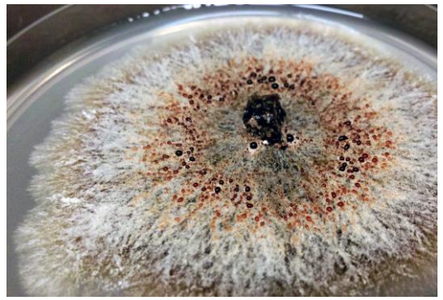




Micangio

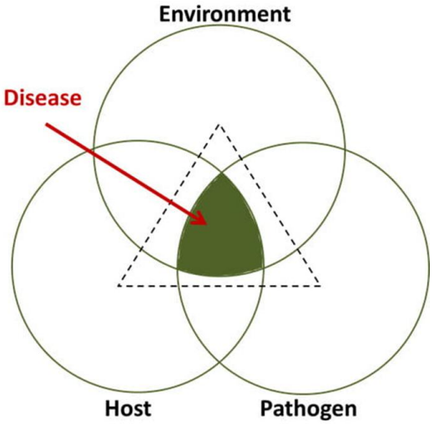
Wood

Gallery



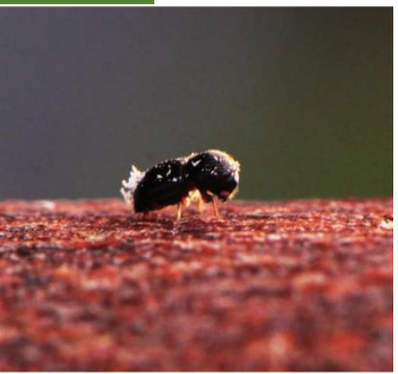
- What if one of the symbiotic fungi is also a plant pathogen?

Xyleborus glabratus - Harringtonia lauricola



- *Both the insect and the fungus were introduced to the US in the early 2000s from Southeast Asia, probably through the timber trade.*
- *Harringtonia lauricola* is lethal to Lauraceous species:
  - *Persea borbonia* (Laurel/redbay)
  - *Persea palustris* (Swamp Bay)
  - *Persea americana* (Avocado)
- In a very short time, they spread from Georgia to North Carolina, Texas and Florida, killing hundreds of thousands of trees.

- In addition to the fungi with which it maintains this symbiotic relationship, they can act with the fungal community in other ways.



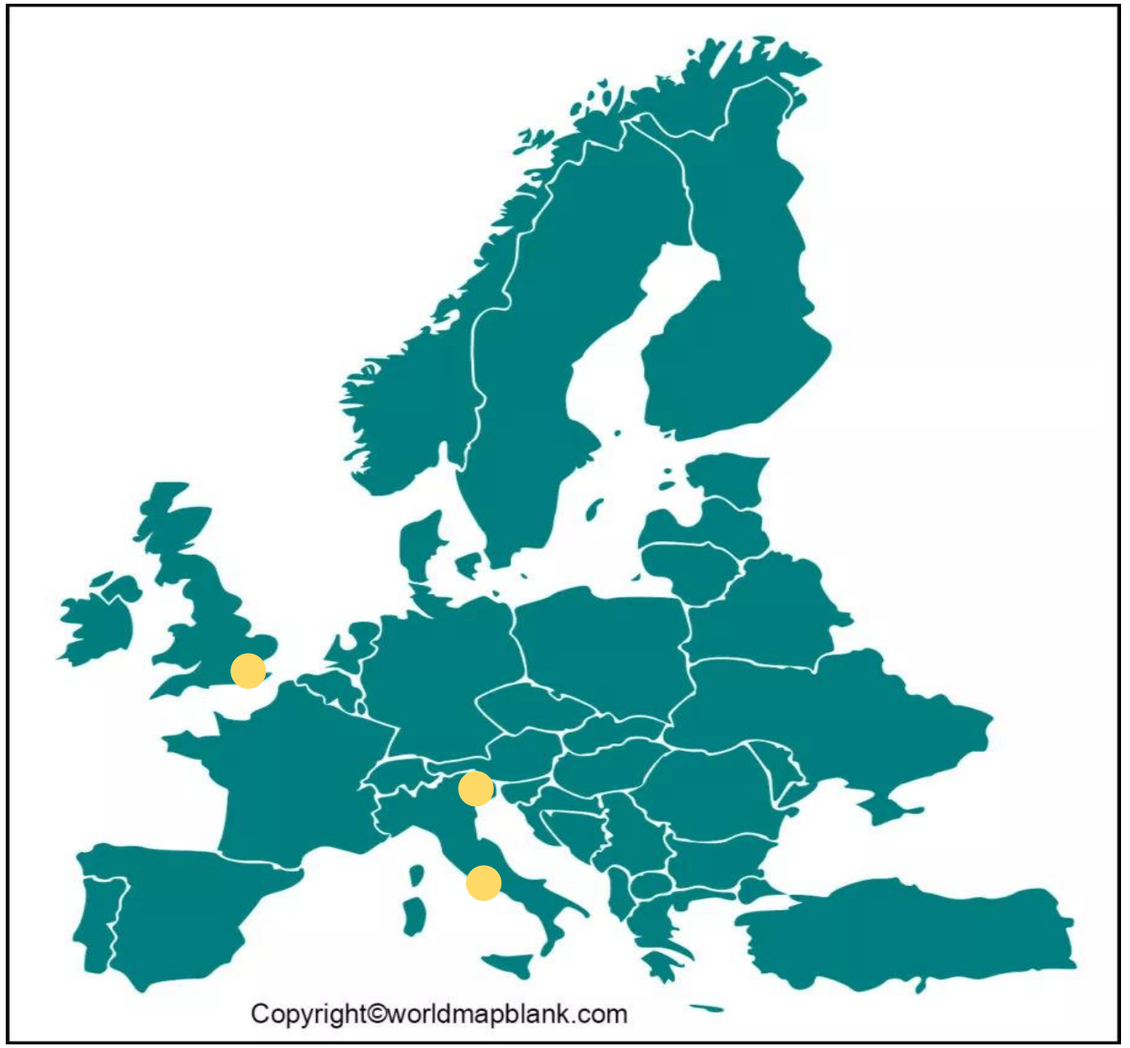
- Ability to act as a vector (sensu stricto, sensu lato)
- Ability to facilitate inoculation.
- Ability to facilitate penetration / colonization

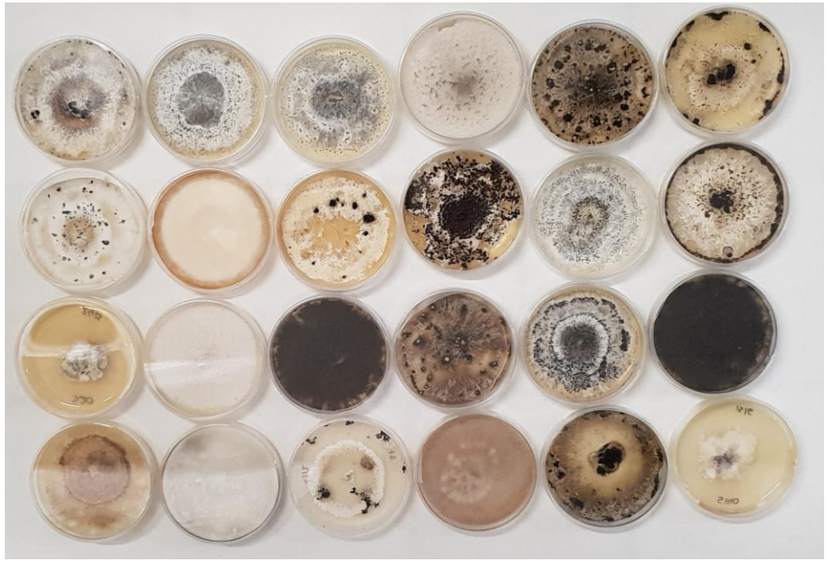
Objetive: Study the fungal community associated to *Xylosandrus germanus* in chestnut groves

UK: Gosdenheath
Hollist Common

Italy: Pagnacco (Friuli-Venezia Giulia)
Monte Cimini (Lazio)

DNA from 36 insects was extracted, amplified (ITS1F/ITS2) and sequenced Illumina 2x300 bp

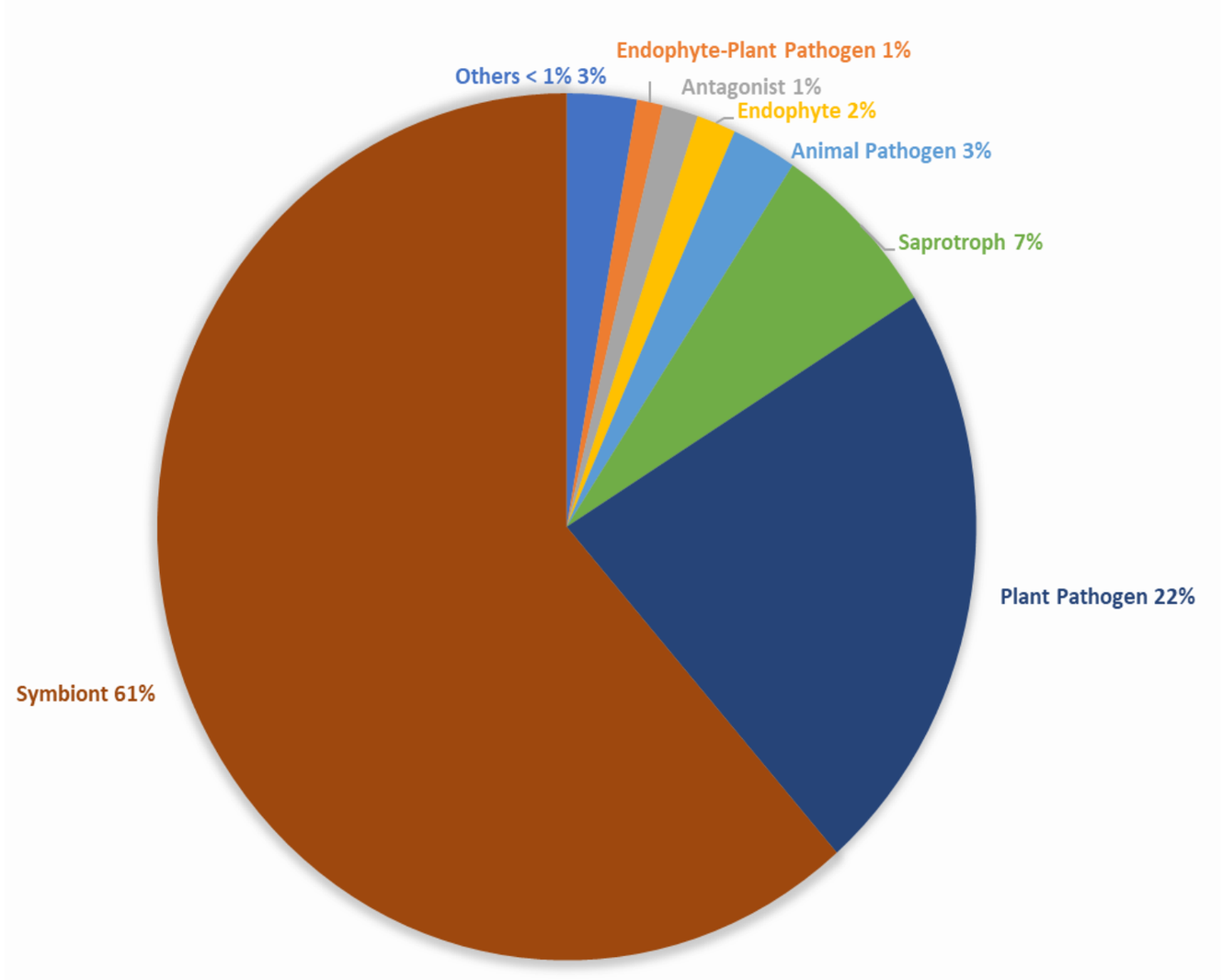




From the 36 insects analyzed 3.961.968 reads were clustering with a 98 % similarity in 677 OTUs

Of these, 355 OTUS were identified to the species level.

Funtional guilds



Symbiont: *Ambrosiella grosmanniae*

Plant pathogen: *Alternaria alternata*

Saprotroph: *Cladosporium dominicanum*

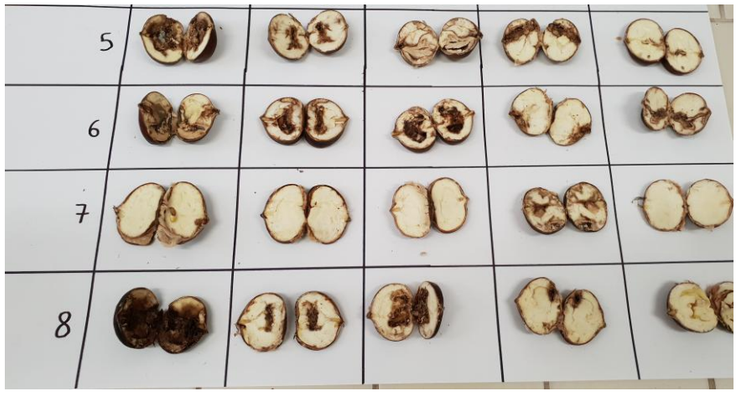
Animal pathogen: *Candida vrieseae*

Endophyte: *Cladosporium sphaerospermum*

Antagonist: *Trichoderma citrinoviride*

Endophyte-Plant Pathogen: *Gnomoniopsis smithogilvyi*

The analysis of the fungal population associated with *X. germanus* also allows us to hypothesize the movement of the insect population on a local, regional and global scale, also providing new information on plant hosts.



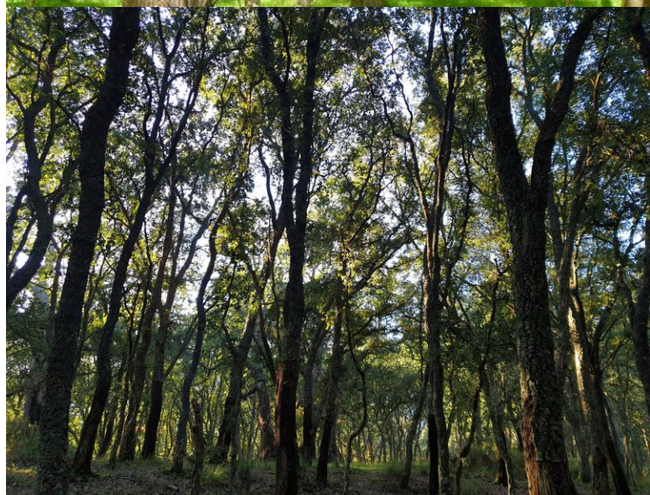
Fungi associated to the chestnut ecosystem:

- *Gnomoniopsis smithogilvyi* 'brown rot of fruits'
- *Cryphonectria parasitica* 'chestnut blight'
- *Ramularia endophylla* 'leaf necrosis'
- *Sclerotinia pseudotuberosa* 'black rot of fruits'
- *Colletotrichum acutatum* 'pink rot of fruits'

Fungi associated to other European ecosystems

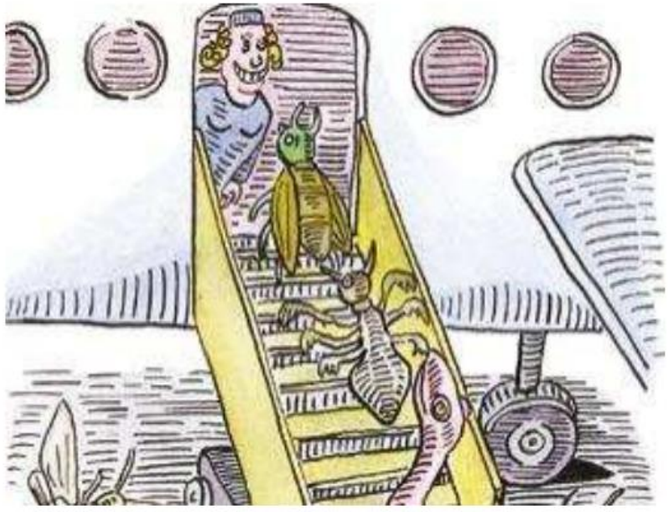
- *Biscogniauxia mediterranea: Quercus, Fraxinus*
- *Peniophora quercina: Castanea, Quercus….*
- *Diaporte foeniculina: Castanea, Citrus, Persea….*
- *Ophiostoma ips*: *Pinus*
- *Calonectria pseudonaviculata* (*Cylindrocladium buxicola*): *Buxus blight*
- *Taphrina carpini*: *Carpinus*
- Diaporthe amygdali. *Prunus* canker
- *Coniella quercicola*: *Quercus*
- Ophiostoma quercus: *Quercus*





Fungi associated to the insect's origin ecosystem news for Europe

- *Paracamarosporium hawaiiense*: Korea and China conifers
- *Cladosporium dominicanum*: Asia and South America
- *Mycosphaerella shimabarensis*: Japan
- *Devriesia pseudoamericana:*



Nursery Monitoring

- About 57% of invasive fungi were introduced into Europe through the plant trade
- Nurseries are one of the main pathways for the dispersal of pests and pathogens

SAMFIX
SAving Mediterranean Forests from Invasions of Xylosandrus beetles and associated pathogenic fungi

Xylosandrus germanus

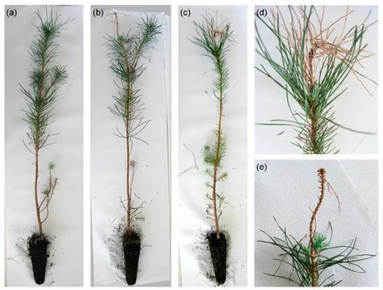
- The traps were placed near nurseries near the Circeo National Park (Italy)

- Only specimens of *Xylosandrus germanus* were captured from which the associated fungi was isolated

Fungal species not reported In Italy

- *Peniophora rufomarginata*
  - *Quercus faginea*: Portugal
  - *Quercus ilex*: Spain
  - *Quercus robur, rotundifolia*: Portugal
  - *Salix aurita:* Scotland
  - *Tilia* sp. : Irlanda

- *Pestalotiopsis pini*
  - Reported in Portogallo (2017) on *Pinus pinea*



- *Pseudosydowia eucalypti*
  - Reported in Africa (South Africa), Asia (Myanmar), Australia, Europa (Portugal) su *Eucalyptus* spp.
  - Substrato**: Leaves**

Fungi species alien in Europe

- ***Cladosporium dominicanum***

Citrus sp.: Repubblica Domenicana e Iran

Dracaena fragrans: Filippine

- ***Cladosporium iranicum***
  - *Citrus aurantium*: Iran
  - *Citrus sinensis* Iran

- ***Coniothyrium palmicola:***
  - *Cocos nucifera*: SudAfrica
  - Syagrus oleracea: Brasile

"It doesn't seem to be covered in our invasive species management plan."

- *Xylosandrus* spp. present a new risk for the natural ecosystems in Europe

- During invasion processes, *Xylosandrus* beetles modulate their associated fungal community, which includes exotic species and native : local or recently acquired species.

- The associated fungal community can give information om invasion Pathways.