



The LifeWatch ERIC Biodiversity
& Ecosystem eScience Conference

PLOVDIV, BULGARIA

7-10 JULY 2026

ITSoneDB v1.144 and BioMaS@ITSoneWB: two ELIXIR-IT main resources for amplicon based mycobiome investigation

10 July 2026

Presenter: Giuseppe Defazio, Ph.D.

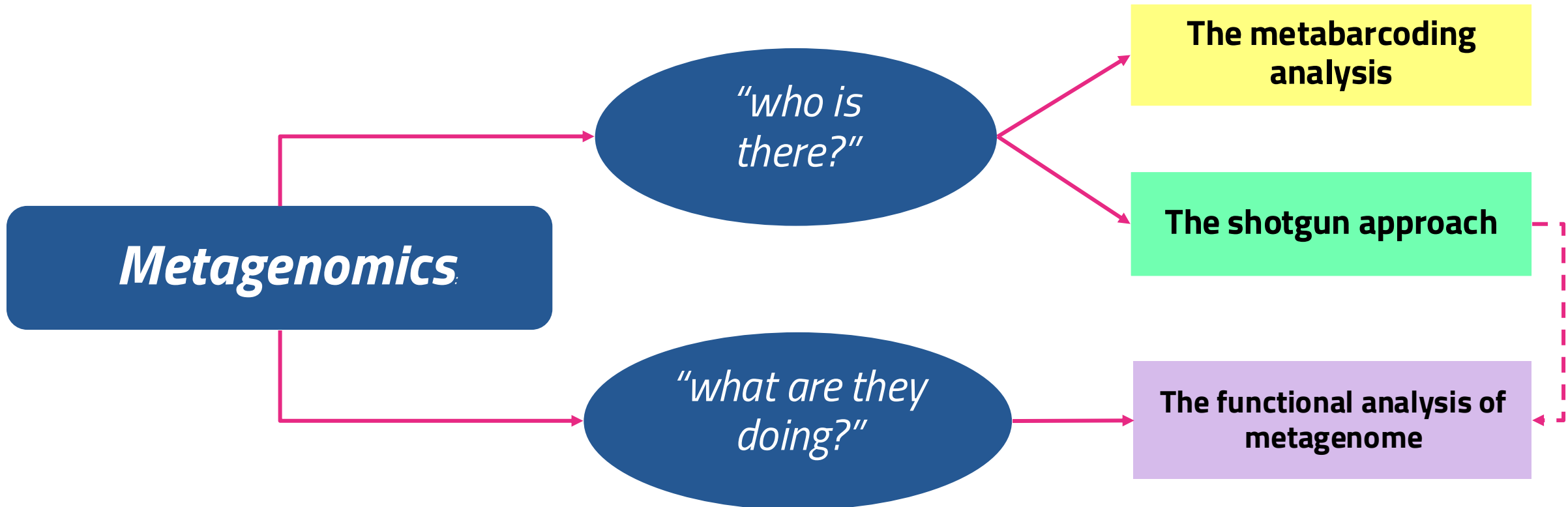


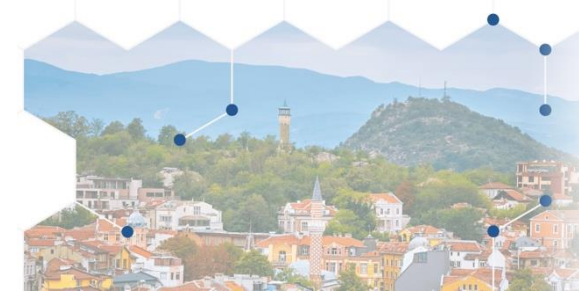
UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO





DEFINITION: if genomics is the study of "the complete set of the DNA molecules of an organism", **metagenomics** is literally "*beyond genomics*" ("meta" = "beyond") indeed, is the study of all microorganisms genomes present in an environmental or host sample.





Shotgun

Metagenomics



Metabarcoding



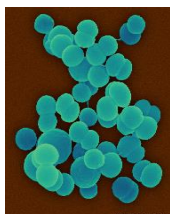
Reference: all available prokaryotic, eukaryotic and viral genomes

Reference: curated barcode collections **SILVA**, **RDP**, **UNITE**, **ITSoneDB**



The amplification and sequencing of 'barcode' DNA regions

Prokaryota:



Eubacteria,
Archaeobacteria



16S rRNA gene

Eukaryota:



Phytoplankton



18S rRNA gene



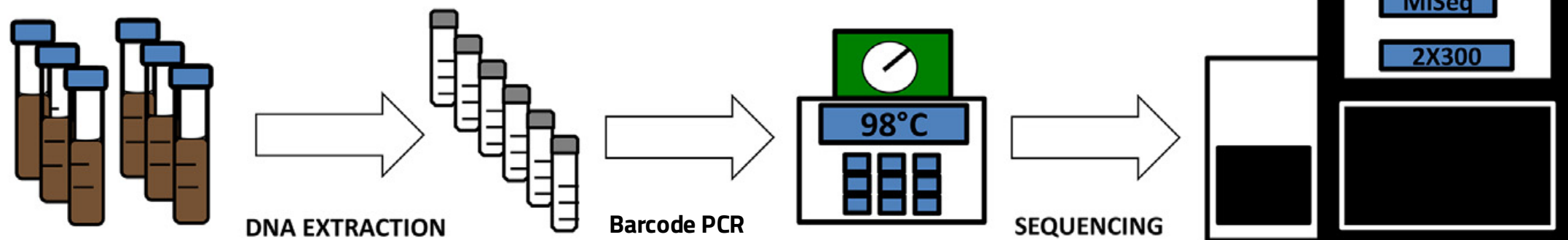
Fungi

ITS
Internal Transcribed
Spacers

Barcode features

- it is **ubiquitous** in the taxonomic research field;
- it is **variable** enough to allow the **discrimination** at the **deepest taxonomic levels**;
- it is **flanked by conserved regions** suited for the PCR primers design;
- it has a **size** compatible with currently available NGS technologies;

The Metabarcoding Analysis workflow



Paired – end reads

```
@ERR000589.41 EAS139_45:5:1:2:111/1
CTTTCCTCCCTGCTTTCCTGGCCCCACCATTTCCAGGGAACATCTTGTCAT
+
3IIIIIIIIIIII>1IIIFF9BG08E00I%IG+&?(4)%00646.C1#&(</pre>

```

Raw data

Merging

Dereplication

Denoising

Chimera removal

Alignment vs. reference

Taxonomic assignment

Here we present...



ITSoneDB

release 1.144 (June 2024) is
based on ENA April 2024 data

<https://itsonedb.cloud.ba.infn.it/index.jsp>

 **Galaxy**
COMMUNITY HUB

 **docker**

BioMaS @ *ITSoneWB*

Bioinformatic analysis of Metagenomic AmpliconS

```
docker pull ibiomcncr/biomas2
```



Who we are...



CNR Milano:

~1200 CPU cores and
5 PB storage (mirror)

CNR Milano:

5 PB storage (backup)



CNR (Bari) and University of Bari
Sequencing facility

CNR Bari:

12.320 CPU cores, 10 GPUs
7,2 PB storage; 25Gb network

INFN Bari:

4.192 CPU cores
2,1 PB storage; 10Gb network

UniBa (Physics Dept.)

4000 CPU cores and 16 GPUs
5.5 PB Cloud Storage
2 PB Posix Storage
20 PB Tape Library

Uniba (Physics Dept.)

5 PB Storage HPC/HTC
250 CPU cores and 8 GPUs
1.5 PB Storage Cloud



UNIVERSITÀ
DEGLI STUDI
DI MILANO



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

UniMi:

INDACO update

UniBo:

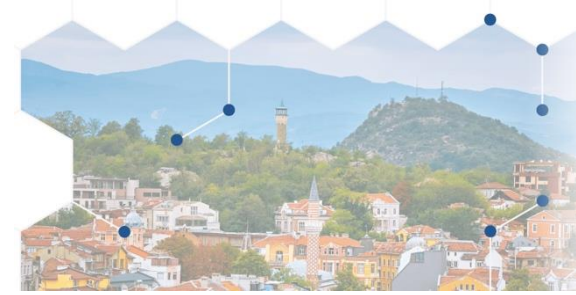
IT infrastructure update

UniPd:

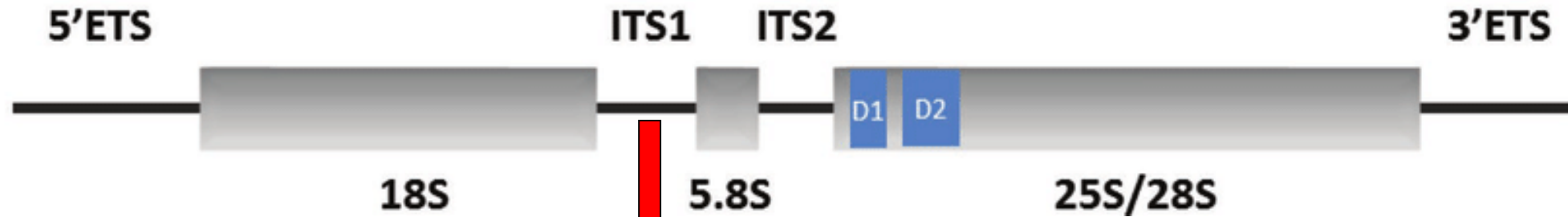
IT infrastructure update



The ITSoneDB v. 1.144



Eukaryotes



ITSoneDB

The database collects
ITS1 sequences of
eukaryotic
microorganisms

The Internal Transcribed Spacer 1 (ITS1) is the most reliable '*barcode*' for fungal communities survey.



European Nucleotide Archive (ENA)

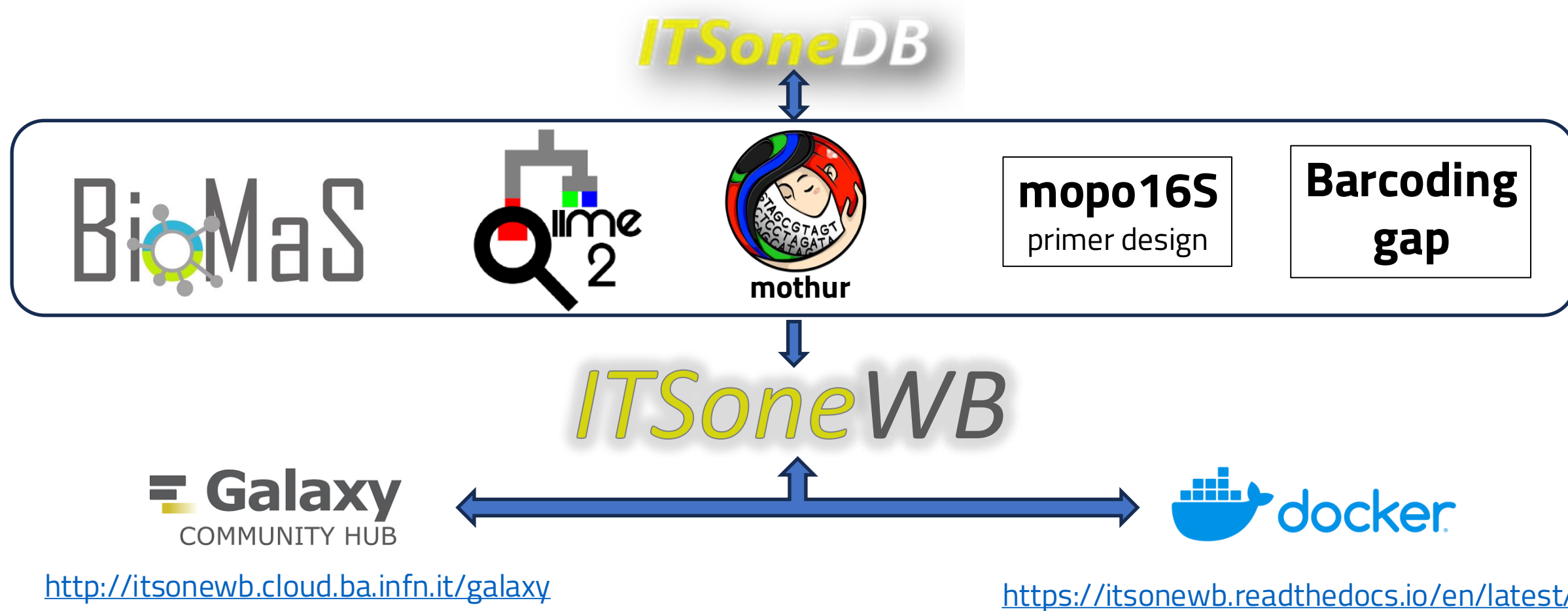
**Sequences
download**

Parsing of the '*Feature Table*' in locally downloaded entries to find already annotated ITS1.

De novo inference of ITS1 position based on 18S and 5.8S Hidden Markov Models (HMM) mapping

Validation of mapping information

The ITsone Work Bench ITsoneWB

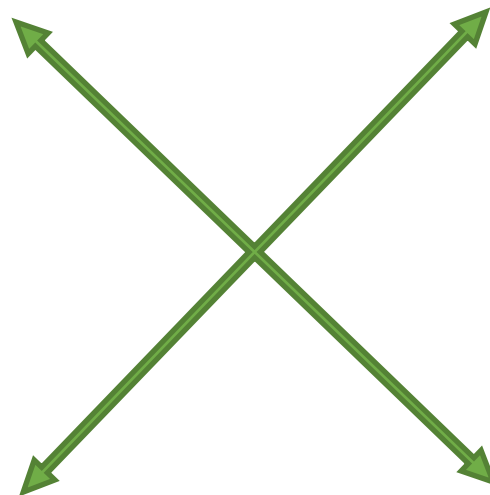




ITSoneDB

versus

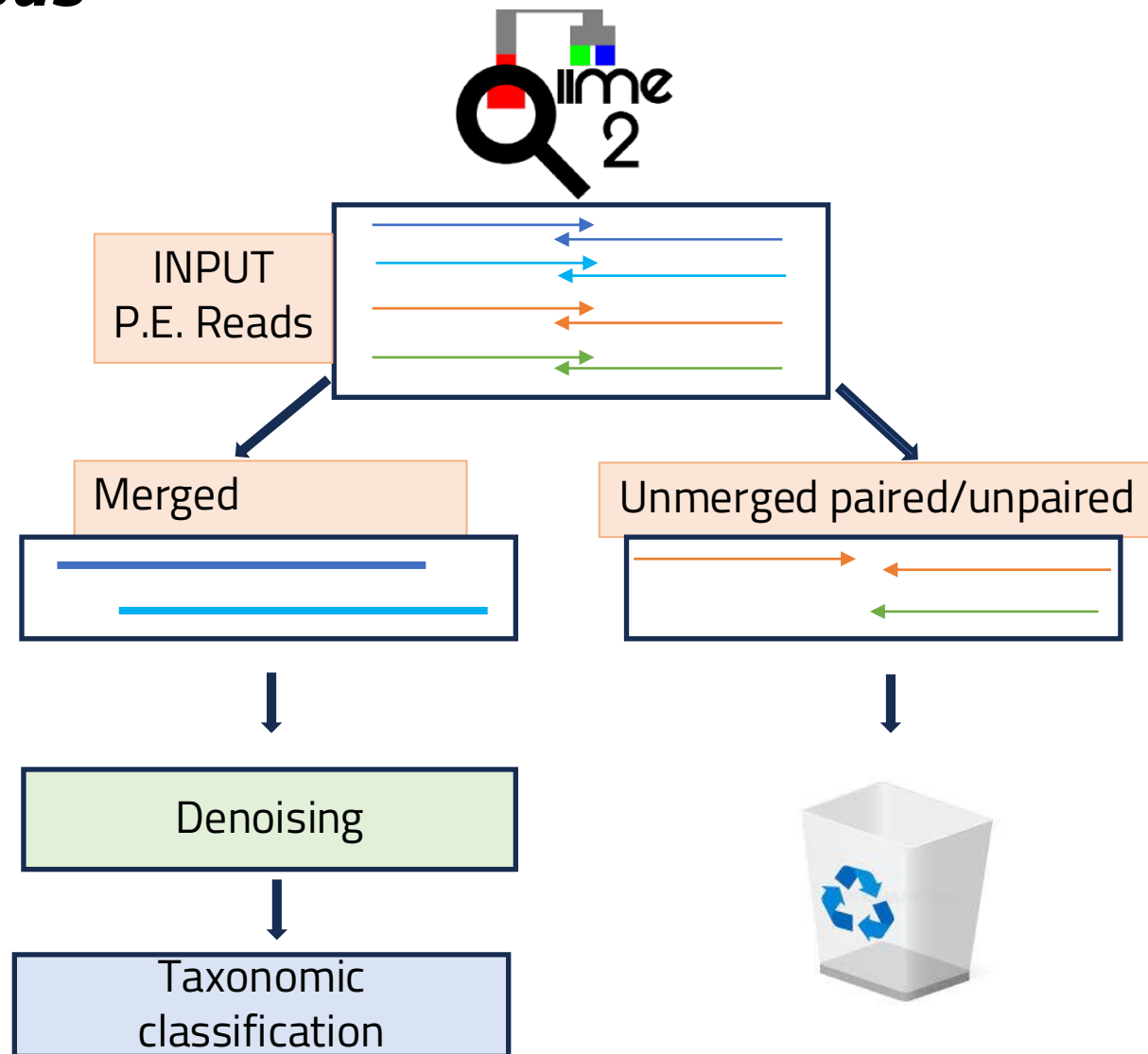
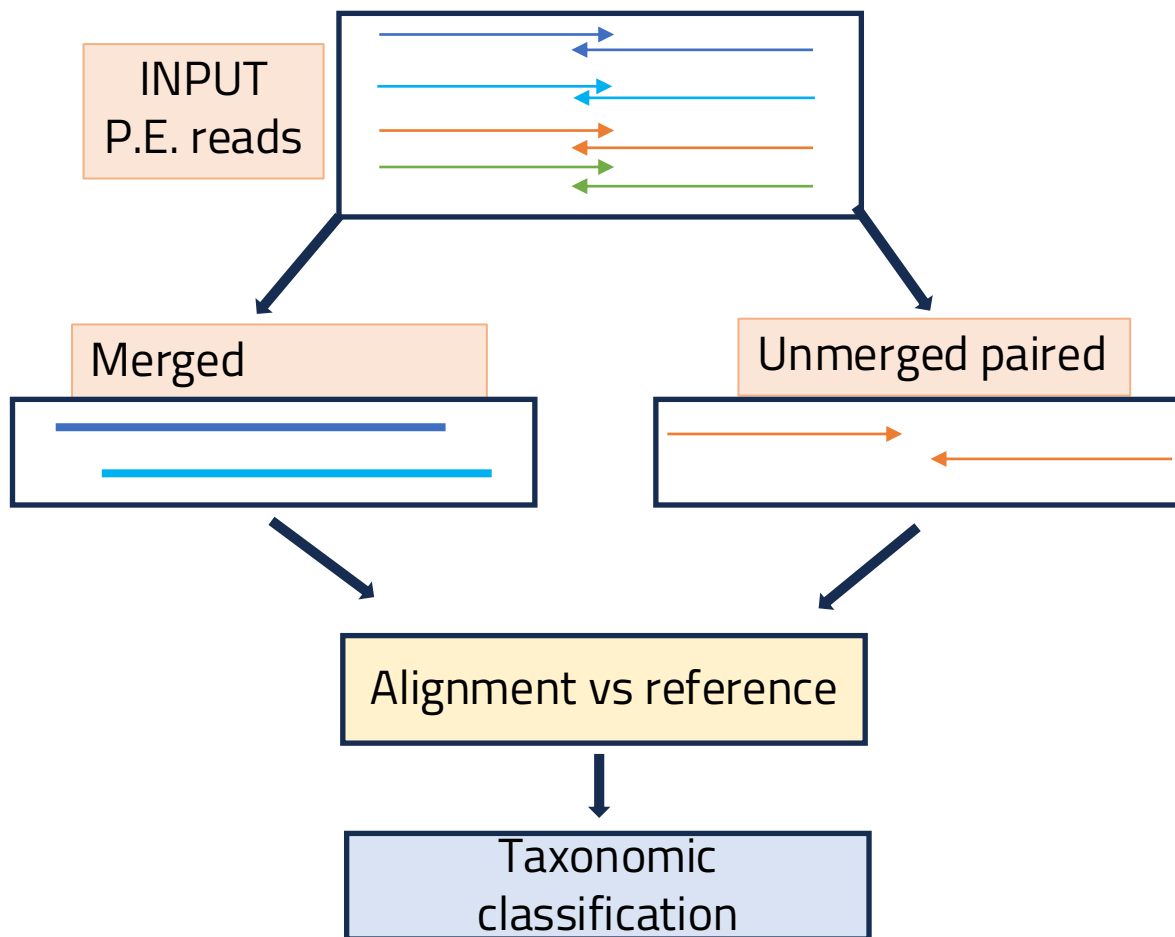
unite





versus

BioMaS @ *ITSoneWB*
Bioinformatic analysis of Metagenomic AmpliconS



Benchmark of ITSoneDB 1.144 and BioMaS@ITSoneWB docker



SAMPLE 1

- Real Mock community
- MiSeq platform
- Paired End 2 x 250
- P.E. reads 412,262

SAMPLE 2

- Sawdust sample of *Picea jezoensis* var. *hondoensis* deadwood
- MiSeq platform
- Run: DRR144721 Bioproject: PRJDB7194
- Paired End 2 x 300
- P.E. reads 290,585

Mock Community composition



SAMPLE 1

Percentage	Organism Name	ATCC code
10	<i>Aspergillus fumigatus</i>	ATCC MYA-4609D-5
10	<i>Cryptococcus neoformans var. grubli</i>	ATCC 208821D-2
10	<i>Trichophyton interdigitale</i>	ATCC 9533D-5
10	<i>Penicillium chrysogenum</i>	ATCC 10106D-5
10	<i>Fusarium keratoplasticum</i>	ATCC 36031D-5
10	<i>Candida albicans</i>	ATCC 10231D-5
10	<i>Nakaseomyces glabratus</i> (alias <i>Candida glabrata</i>)	ATCC 2001D-5
10	<i>Malassezia globosa</i>	ATCC MYA-4612D-5
10	<i>Saccharomyces cerevisiae</i>	ATCC 201390D-5
10	<i>Cutaneotrichosporon dermatis</i>	ATCC 204094D-5



The ITSoneDB + BioMaS@ITSoneWB mock results at genus level



Selecting Items > than 0.01 %

SAMPLE 1

Sensitivity 90%
Precision 90%

	PE reads nr	PE reads %
Total PE reads	412,262	100
Merged	137,897	33.45
Classified	379,600	92.08
Correctly classified	322,395	78.2

Taxon Name	Total Assigned	Percent	
<i>Cutaneotrichosporon</i>	63,070	16.61	TP
<i>Fusarium</i>	56,327	14.83	TP
<i>Cryptococcus</i>	50,598	13.32	TP
<i>Trichophyton</i>	50,359	13.26	TP
<i>Saccharomyces</i>	36,081	9.50	TP
<i>Nakaseomyces</i>	25,701	6.77	TP
<i>Aspergillus</i>	17,624	4.64	TP
<i>Candida</i>	13,316	3.50	TP
<i>Penicillium</i>	9,319	2.45	TP
<i>Trichosporon</i>	7,397	1.94	FP (*)
<i>Malassezia</i>	0	0	FN

* *Trichosporon* is the basionym of *Cutaneotrichosporon*



BEEs
2026

The ITSoneDB + QIIME2 mock results at genus level



Selecting Items > than 0.01 %

Taxon Name	Percentage	
<i>Fusarium</i>	50.09	TP
<i>Cryptococcus</i>	11.68	TP
<i>Saccharomycetales_genus</i>	11.36	FP
<i>Cutaneotrichosporon</i>	8.37	TP
<i>Trichophyton</i>	7.37	TP
<i>Malassezia</i>	5.41	TP
<i>Aspergillus</i>	3.39	TP
<i>Penicillium</i>	2.29	TP
<i>Saccharomyces</i>	0	FN
<i>Nakaseomyces</i>	0	FN
<i>Candida</i>	0	FN

SAMPLE 1

Sensitivity 70%
Precision 87.5%

	PE reads nr	PE reads %
input	412,262	100
filtered	412,238	99.99
denoised	410,603	99.99
merged	318,858	77.34
non-chimeric	79,855	19.37
Correctly classified	38,011	9.22



The UNITE + BioMaS@ITSoneWB mock results at genus level



SAMPLE 1

**The combination of
UNITE + BioMaS@ITSoneWB produces
no classification for any reads pair**



The UNITE + QIIME2 mock results at genus level



Selecting Items > than 0.01 %

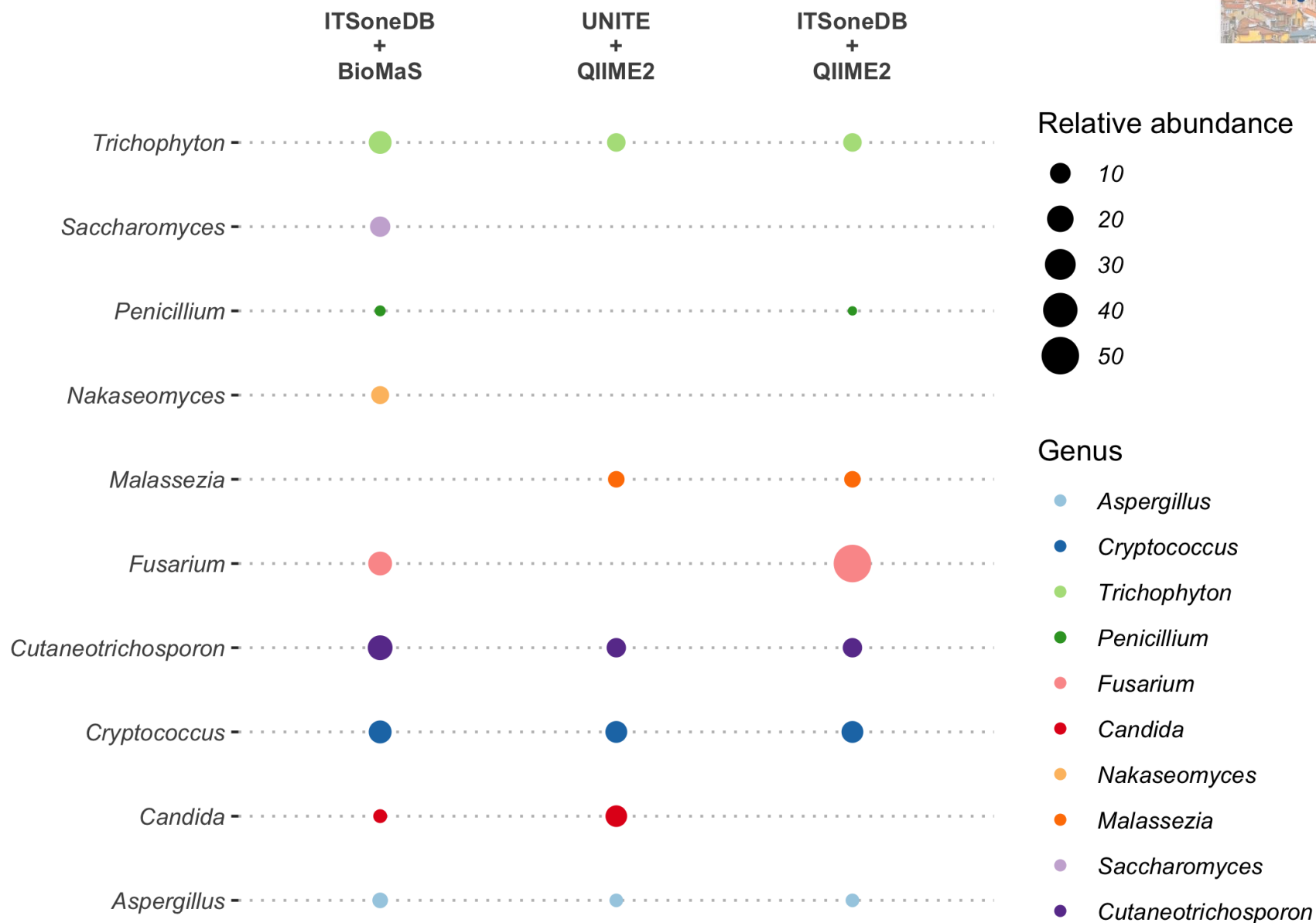
Taxon Name	Percentage	
<i>Fungi_genus</i>	52.39	FP
<i>Cryptococcus</i>	11.68	TP
<i>Candida</i>	11.36	TP
<i>Cutaneotrichosporon</i>	8.37	TP
<i>Trichophyton</i>	7.37	TP
<i>Malassezia</i>	5.41	TP
<i>Aspergillus</i>	3.39	TP
<i>Saccharomyces</i>	0	FN
<i>Nakaseomyces</i>	0	FN
<i>Penicillium</i>	0	FN
<i>Fusarium</i>	0	FN

SAMPLE 1

Sensitivity 60%
Precision 85.7%

	PE reads nr	PE reads %
input	412,262	100
filtered	412,238	99.99
denoised	410,603	99.99
merged	318,858	77.34
non-chimeric	79,855	19.37
Correctly classified	38,011	9.22

Observed abundance of expected taxa





BEEs
2026

The sawdust sample of *Picea jezoensis* var. *hondoensis* deadwood result comparison for taxonomic classification



BioMaS@ITSoneWB docker 498 genera found

Unclassified P.E. reads 11

Input P.E. reads 290,585

SAMPLE 2

Name	Reads	Percentage
<i>Brachysporium</i>	18,323	7.568
<i>Curvichaeta</i>	15,320	6.328
<i>Hyphoderma</i>	12,744	5.264
<i>Acrodontium</i>	10,870	4.490
<i>Aniptodera</i>	10,674	4.409
<i>Leptodontidium</i>	6,901	2.850
<i>Ascocoryne</i>	4,183	1.728

Brachysporium nigrum **wood** Saugatucket River (Lamore & Goos 1978)

Curvichaeta curvispora New Zealand **wood** (Reblova 2004)

Hyphoderma obtusifforme The **Corticaceae** of North Europe (Erikss 1975)

Acrodontium **Fagus sylvatica decadey leaf** (de Hoog 1972)

Aniptodera from **Wood** in Freshwater Habitats (Shearer 1989)

Leptodontidium **root**-associated fungi (Melin 1922)

Ascocoryne **deadwood** associated fungi (Leonhardt et al. 2019)

QIIME2

Feature ID	ITSoneDB + QIIME2	UNITE + QIIME2	Reads
ASV1	Unassigned	k_Fungi	10
ASV2	g_Rhizophydium	k_Fungi	18

Discarded P.E. reads 290,557

Rhizophydium decomposing fungi



Defazio et al. *BioData Mining* (2026) 19:35
<https://doi.org/10.1186/s13040-026-00532-6>

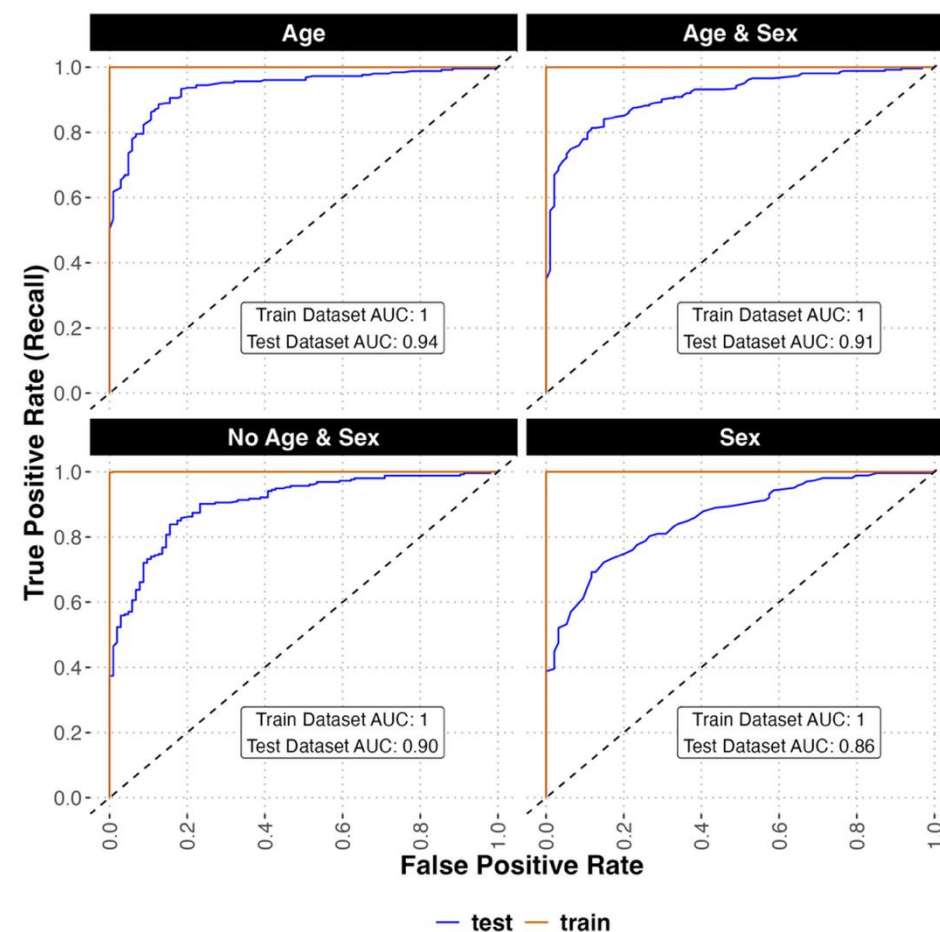
BioData Mining

Machine learning-based assessment of the healthy human gut mycobiota landscape using ITS1 DNA metabarcoding data

Giuseppe Defazio¹, Erika Lorusso¹, Mariangela De Robertis^{1,2}, Tommaso Mello³, Andrea Galli³,
Graziano Pesole^{1,2*} and Bruno Fosso^{1*}

Keypoints

- 1.500 human gut samples
- Health status prediction with Random Forest models relying on mycobiome profiling
- Accuracy in health status prediction > 80%





Nakaseomyces is appreciable exclusively with *ITSoneDB 1.144 + BioMaS@ITSoneWB docker* combination because of *Nakaseomyces grabratus* ITS1 sequence 862 nt long.

The highest Sensitivity and Precision for mock community are obtained with *ITSoneDB 1.144 + BioMaS@ITSoneWB docker*.

In both mock community and sawdust sample *ITSoneDB 1.144 + BioMaS@ITSoneWB docker* catches the highest diversity.

In real samples genera observed with *ITSoneDB 1.144 + BioMaS@ITSoneWB docker* are coherent with literature findings.



Despite the large application in Prokaryotes metabarcoding analysis, Denoising and ASV inference applicability to ITS1 surveys is limited due to barcode length variability (100-1000 nt) in eukaryotes.

BioMaS relying on sequence mapping and direct classification, as higher sensitivity and accuracy for fungal/eukaryotes ITS1-based surveys.

ITSoneDB shows a higher reliability and classifier tools adaptability than UNITE

Use case demonstrate ITSoneDB & BioMaS@ITSoneWB are suitable for Machine Learning applications



The LifeWatch ERIC Biodiversity
& Ecosystem eScience Conference
PLOVDIV, BULGARIA 7-10 JULY 2026



**GIUSEPPE
DEFAZIO**
PostDoc Researcher



**MARCO ANTONIO
TANGARO**
Technologist



**BRUNO
FOSSO**
Associate Professor



**GRAZIANO
PESOLE**
Full Professor In
Molecular Biology

THANK YOU!

Questions?

bruno.fosso@uniba.it
graziano.pesole@uniba.it



**UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO**

