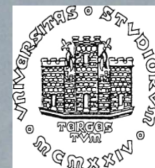
# Development of NGS meta-barcoding for the characterization of aerobiological samples

**Lucia Muggia**

Alberto Pallavicini, Elisa Banchi, Claudio G. Ametrano, David Stankovic, Silvia Ongaro, Enrico Tordoni, Mauro Tretiach

DIPARTIMENTO DI SCIENZE DELLA VITA



UNIVERSITÀ DEGLI STUDI DI TRIESTE

# **PROGETTO FRA-2016** (Finanziamento per la Ricerca di Ateneo, Università di Trieste)

## The ARPA collaborators

| | |
|---|---|
| Dott. Verardo Pierluigi | ARPA Friuli Venezia Giulia (Pordenone) |
| Dott.ssa Tassan Francesca | ARPA Friuli Venezia Giulia (Trieste) |
| Dr.ssa Gabrielli Francesca | ARPA Marche |
| Dott.ssa Nadia Trobiani | ARPA Marche |
| Dott.ssa Lazzarin Stefania | ARPA Veneto |
| Dott.ssa Moretti Olga | ARPA Umbria |
| Dr.ssa Borney Francesca | ARPA Valle d' Aosta |

# Pollen monitoring

mainly based on microscopic analyses
identification and quantification and requires well trained personnel
procedure is laborious and time consuming
Identification up to genus level, usually higher taxonomic ranks

# Fungal spores - AEROMYCOLOGY

huge diversity of morphologies
high seasonal variation
difficulties in identifying fungal species using spores alone

10µm

# Project aims

**Aim 1.** Set up and standardization of DNA extraction, amplification and sequencing of aerobiological samples for routine analyses

**Aim 2.** Standardization of a methodological and bioinformatic pipeline for meta-barcoding analysis of aerobiological samples

**Aim 3.** Establishment of reference sequences database

# DNA barcoding

provides a fast and thorough identification of organisms.
species-specific DNA regions = species-specific tag.
DNA barcode (nuclear, mitochondrial or plastidial DNA):

1) be relatively conserved within species but variable enough to discriminate between them;
2) contain sufficient phylogenetic information;
3) be flanked by highly conserved regions

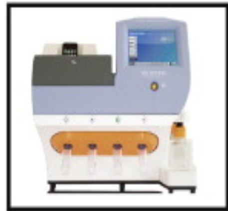
<u>!!!!! no universal DNA barcode exist !!!!</u>

DNA BARCODING
101

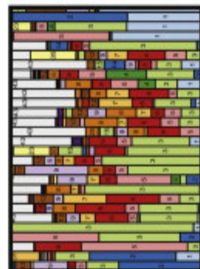
# **DNA Meta-barcoding**: in microbiome studies for the characterization of soil and water microbial communities in diverse ecological conditions, or of plants and animal associated microorganisms.

Mixed species environmental DNA sample

Lab processing



Next-generation sequencing



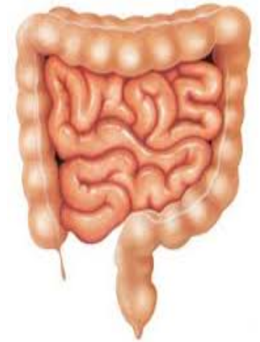
Data analysis

Species identification

**Minibarcodes** that target at 200-300 bp are used.

- ✓DNA barcoding
- ✓Next generation sequencing (NGS)

**Environmental samples**



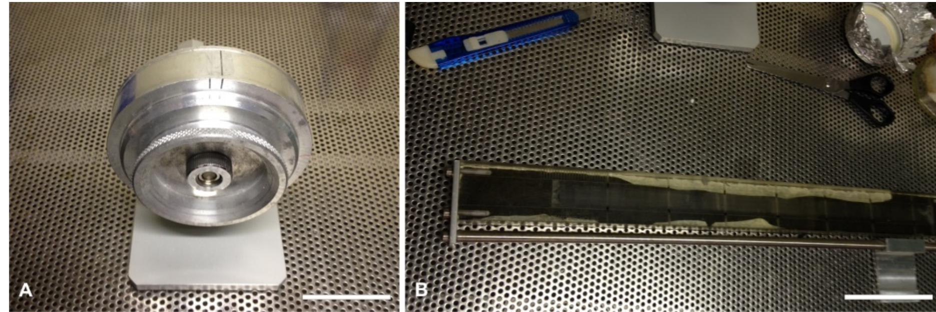
Ion Torrent PGM®

# Sampling and methodology

5 localities = cities ★

Valle d'Aosta
Trentino - Alto Adige
Friuli Venezia Giulia
Lombardia
Veneto
Piemonte
Emilia - Romagna
Liguria
Toscana
Marche
Umbria
Abruzzo
Lazio
Molise
Campania
Puglia
Basilicata
Sardegna
Calabria
Sicilia

The Twenty Regions of the Republic of Italy




total material DNA extraction

March-November 2017

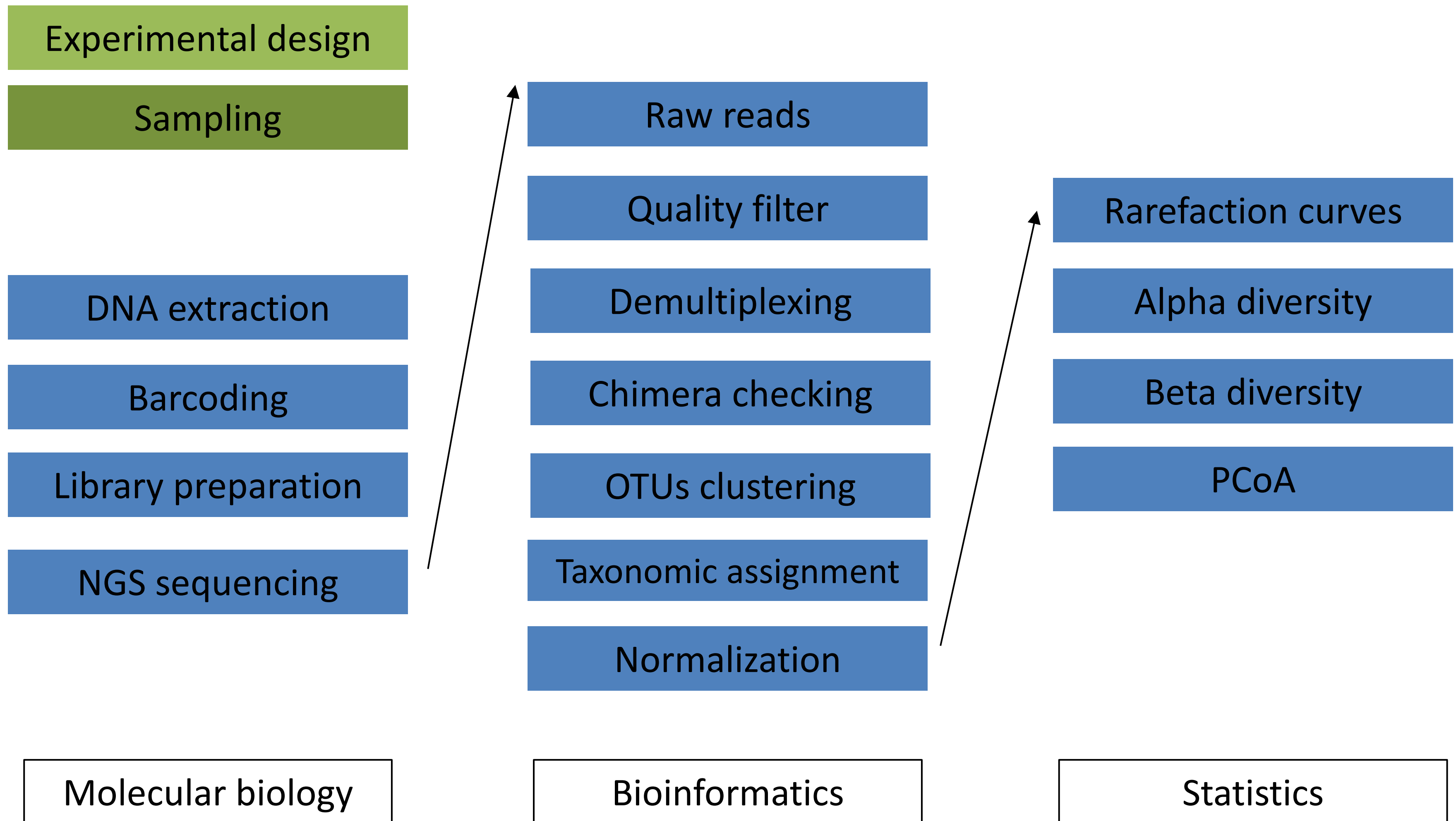
PCR amplifications:

**trnL** (plastidial) for plants

others?? **nuclear**??

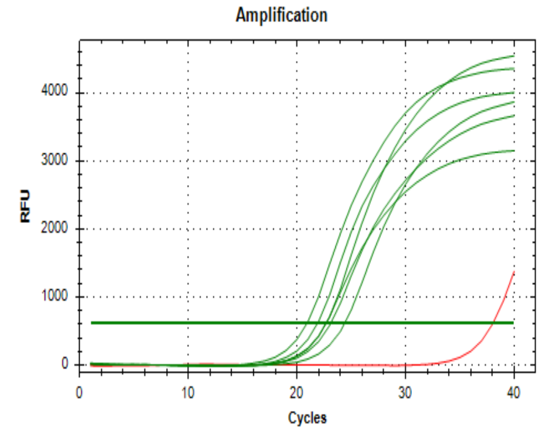
**ITS2** (nuclear) for fungi

# Metabarcoding standard pipeline:

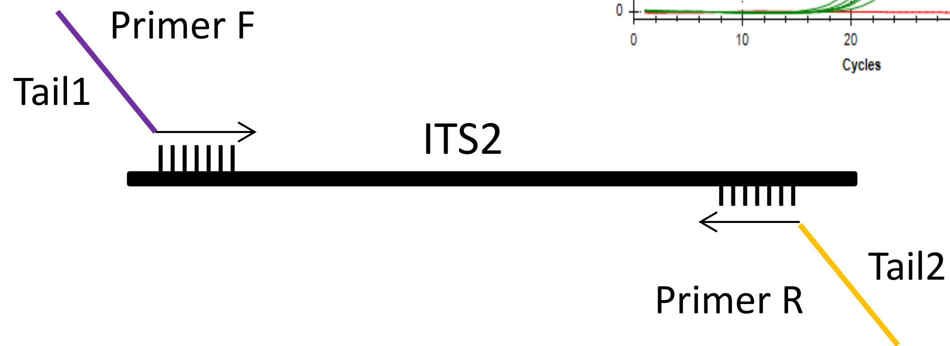



WHO IS THERE???

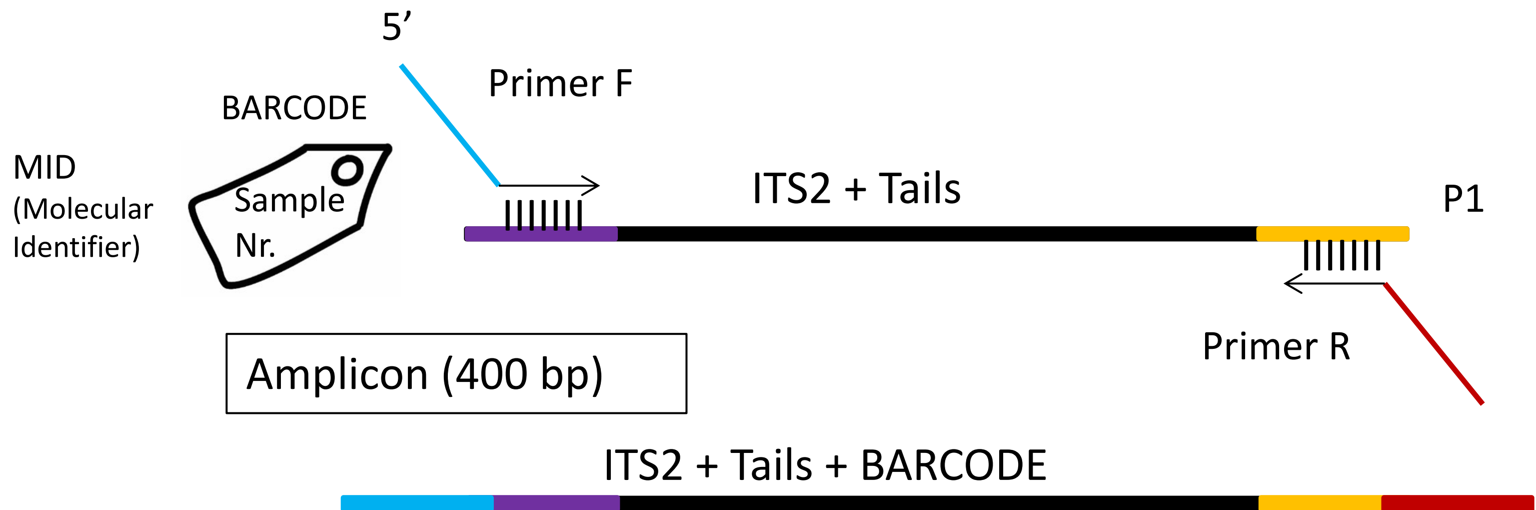
# Library preparation



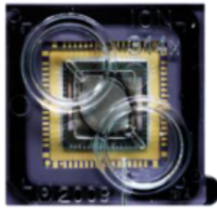

1st PCR



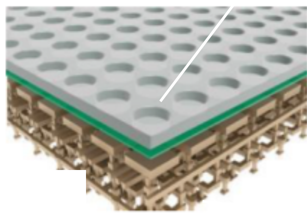

2nd PCR



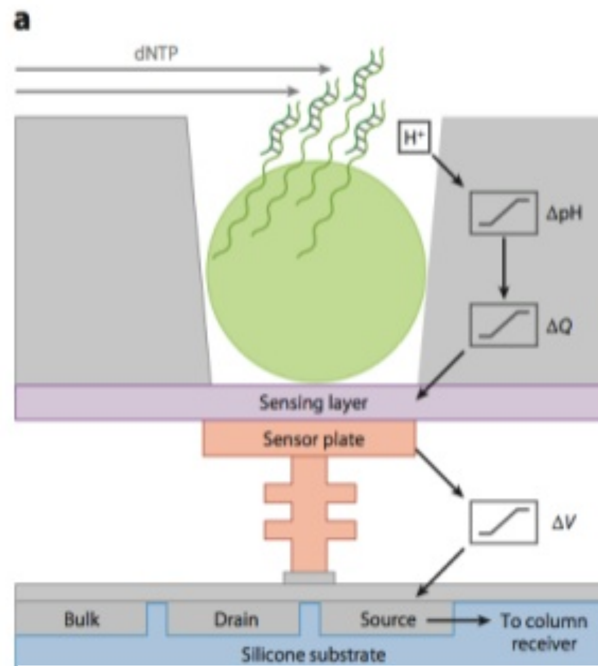
# Ion Torrent pH Based Sequencing

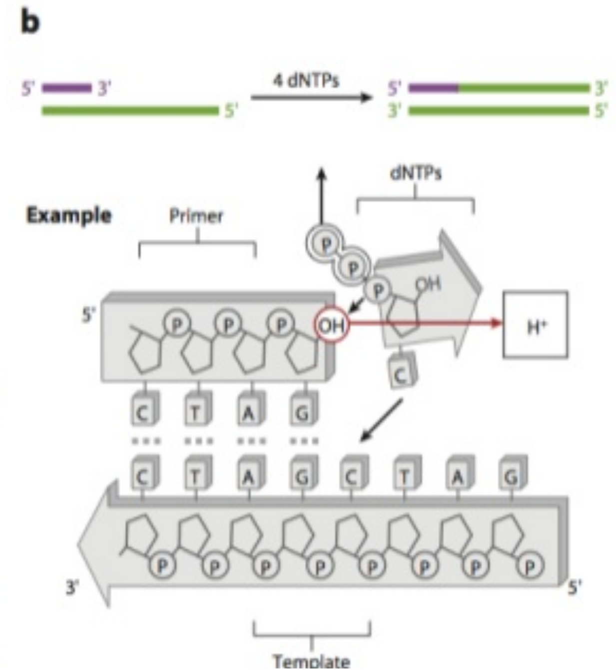


Chip
Semiconductor Packaging



Millions of Sensors
Semiconductor Design

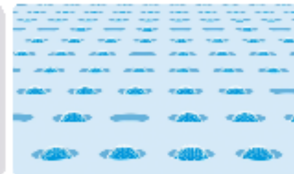



**Copy DNA**

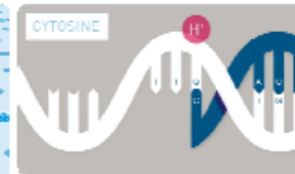
- DNA fragments
- Attach to a beads
- Fragment is copied by PCR

**Load chip**



- Load bead into a well
- Load one of the four nucleotides

**Incorporate nucleotide**




H ion changes the pH in the well

**Detect and call**

An ion-sensitive layer beneath the well measures that pH change and converts it to voltage

This voltage change is recorded, indicating the nucleotide has been incorporated and the base is called

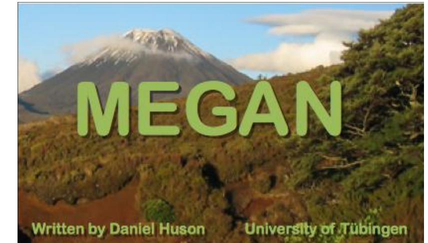
# Bioinformatic analysis (1)

QIIME Pipeline (Caporaso et al., 2011)

- file conversion (.fastq to .fna and .qual)
- clean, trim and split the library (-l 150  -s 15 -h 8 -m 3)
- ITSx extraction (Bengtsson-palme et *al*., 2013)
- chimera checking with qiime-usearch 6.1 (UNITE reference dataset)
- 97% similarity OTUs picking (open reference strategy, UNITE database and blast)
- calculate the core taxonomy
- filter out taxonomy the taxa belonging to lichen hosts
- recalculate the core taxonomy to focus only on intrathalline diversity
- statistical analysis (rarefaction curves, reads distribution, alpha and beta diversity)

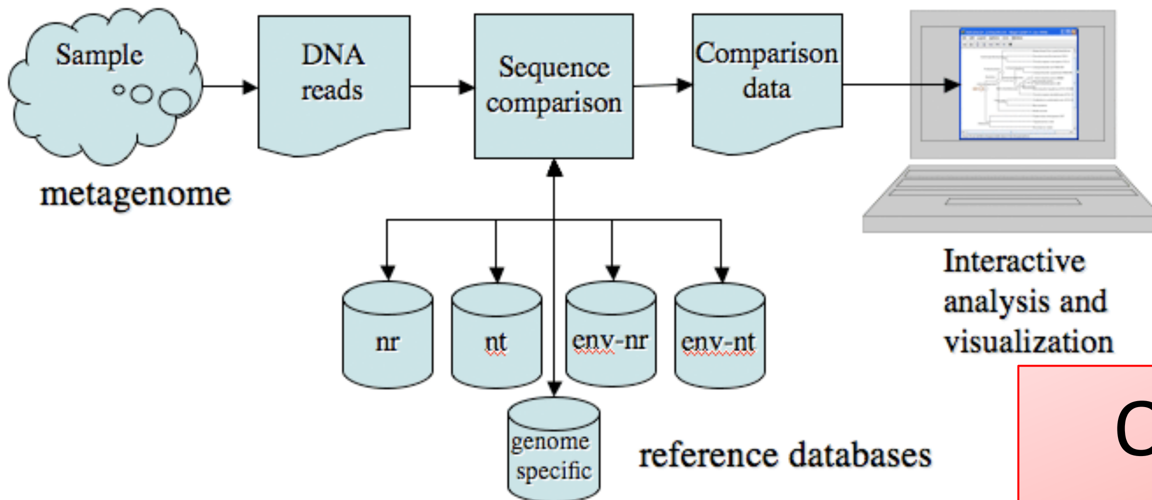
```
gzahn$ print_qiime_config.py -t

System information
==================
         Platform:      linux2
   Python version:      2.7.10 |Anaconda 2.3.0 (64-bit)| (default, May 28 2015, 17:02:03)  [GCC 4.4.7 20120313 (Red Hat 4.4.7-1)]
Python executable:      /home/gzahn/anaconda/bin/python

QIIME default reference information
===================================
For details on what files are used as QIIME's default references, see here:
 https://github.com/biocore/qiime-default-reference/releases/tag/0.1.3

Dependency versions
===================
          QIIME library version:        1.9.1
           QIIME script version:        1.9.1
qiime-default-reference version:        0.1.3
                  NumPy version:        1.9.2
                  SciPy version:        0.15.1
                 pandas version:        0.16.2
             matplotlib version:        1.4.3
            biom-format version:        2.1.5
                   h5py version:        2.5.0 (HDF5 version: 1.8.15)
                   qcli version:        0.1.1
                   pyqi version:        0.3.2
             scikit-bio version:        0.2.3
                 PyNAST version:        1.2.2
                Emperor version:        0.9.51
```

# Bioinformatic analysis (2)




MEGAN Pipeline (Huson et al., 2011)

- blast and taxonomic assignment with MALT and MEGAN
- malt build - download as fasta file the sequences of interest, in ncbi, with an appropriate query (i.e. "ITS2" OR "ITS" OR "transcribed" AND "Fungi" OR "uncultured fungus" OR "fungus" – 96428 seqs)
- malt run (produces .rma files, one per sample) i.e. -id 90
- import .rma files in MEGAN and taxonomic analysis (LCA algorithm)




OUTPUT = **reference sequences databases** for plants and fungi

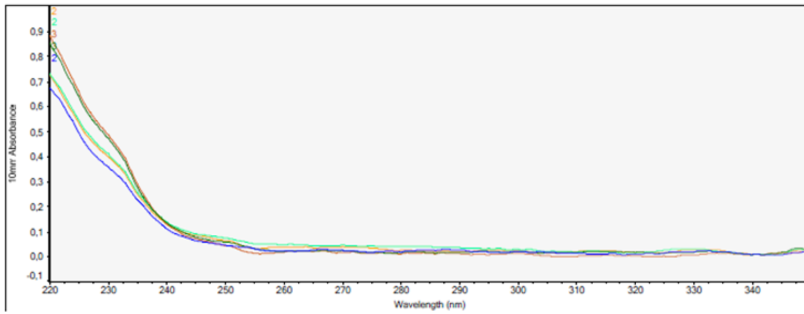
# The time schedule: 01.01.2017 - 31.12.2018

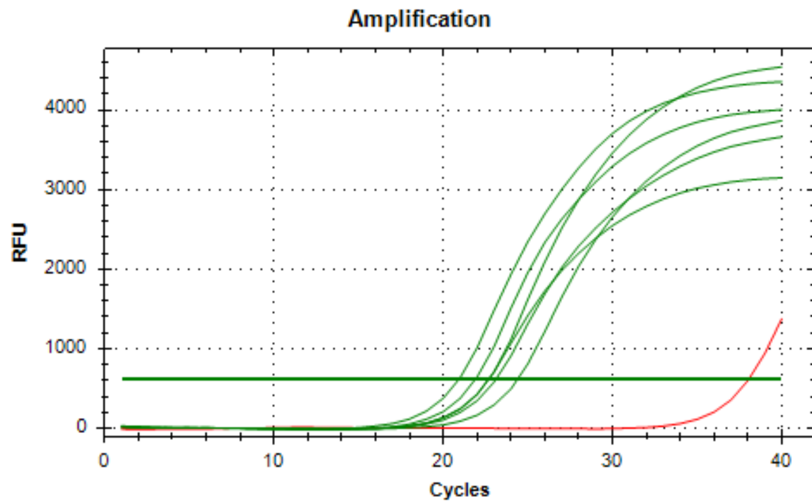
| | | Time schedule (months) | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Aims** | **Activities** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** |
| 1 | 1. Preparatory phase | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 2. Sampling | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 3. Standardization of set up protocols | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | 4. Bioinformatic pipeline standardization | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 | 5. Sequence reference database | | | | | | | | | | | | | | | | | | | | | | | | |
| | Preparation of scientific publications | | | | | | | | | | | | | | | | | | | | | | | | |

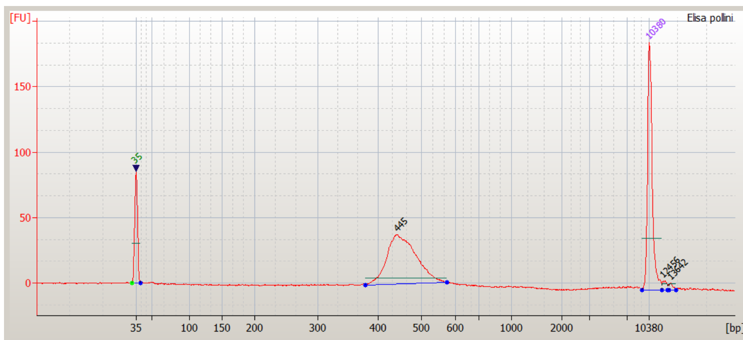
Collaboration with ARPA, sampling and collection of preliminary results since July 2016.

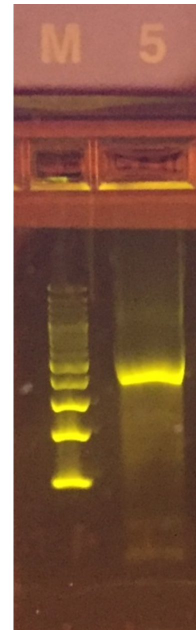
## Preliminary study

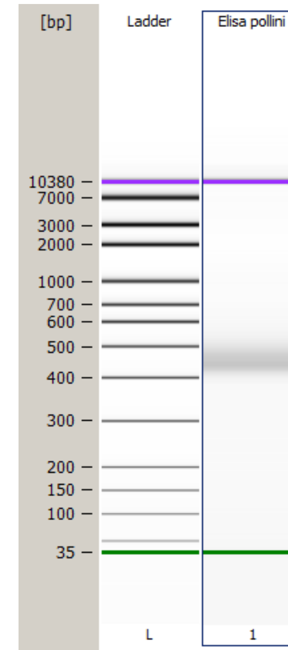
- 48 samples (2 weeks of sampling)
- ARPA Marche, FVG, Umbria, Veneto
- ✓ DNA extraction
- ✓ Amplification with RT-PCR
- ✓ Library preparation
- ✓ NGS sequencing with Ion Torrent®

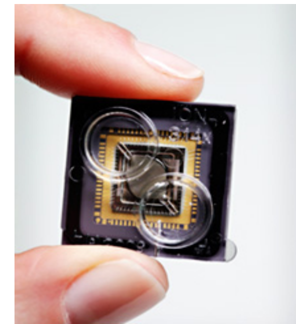


140.105.153.83/report/116/

App Bookmarks Crochet Dottorato Ricette Viaggi Varie Save to Mendeley Altri Preferiti

All Data Management Auto Actions are disabled and /results/ is 73.14% full Visit Data Management

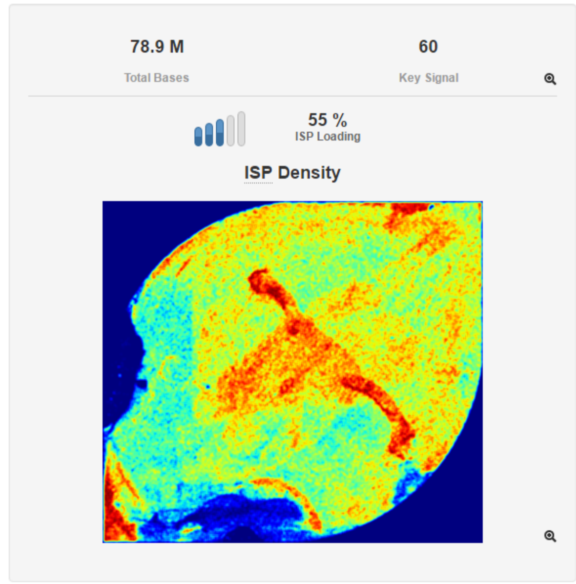
Output Files Plugin Summary

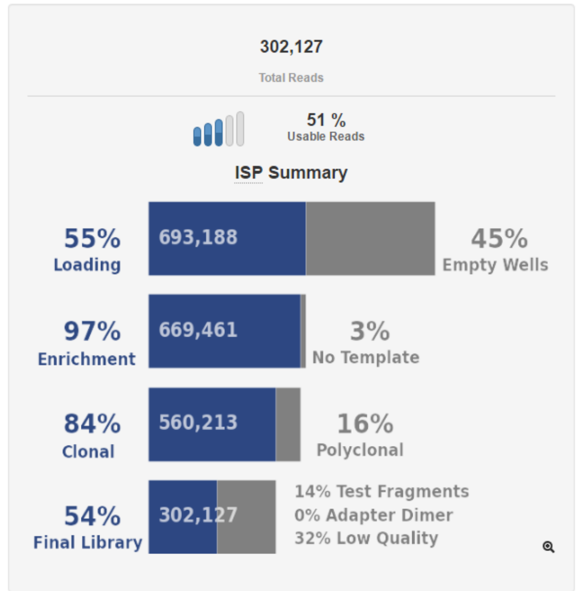
Upload to IR Report Actions Reports

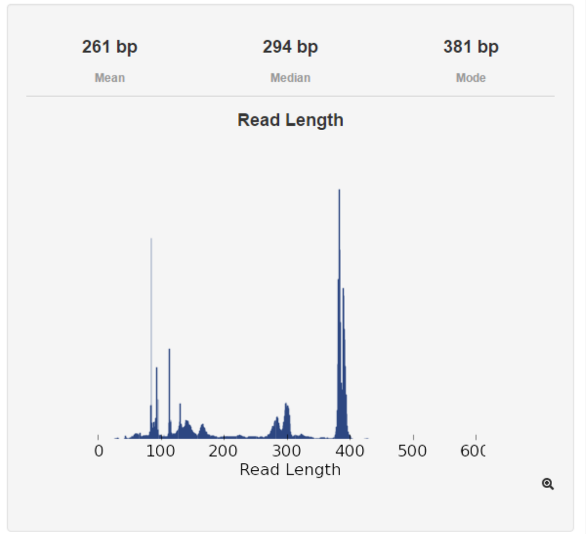
**Run Summary: R_2017_02_23_12_56_30_ionadmin_SN2-84-Pollen_Fungi_2_ITS2_314_23.02.17**

Reports : Auto_ionadmin_SN2-84-Pollen_Fungi_2_ITS2_314_23.02.17_256 (116)

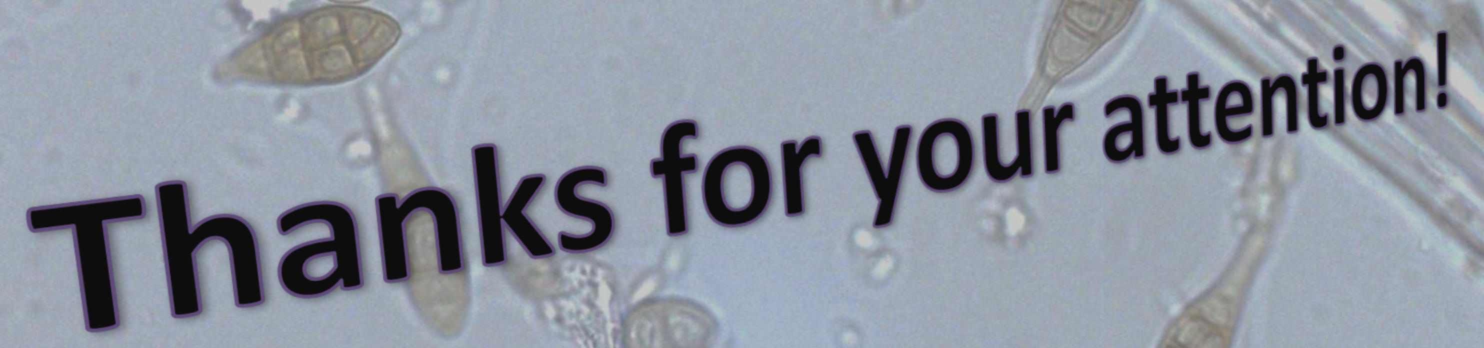
**Read Summary: Unaligned**




**No Alignment Reference selected**



# Thanks for your attention!

# Looking forward to our future colaboration!!!!

10μm