Metagenomics: challenges, pitfalls and perspectives

Graziano Pesole

IBIOM
Istituto di Biomembrane, Bioenergetica e Biotecnologie Molecolari

elixir ITALY

UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Workshop on Computational Metagenomics: Methods, Standards and Experimental Procedures

19-20 June, 2017 - University of Bari "Aldo Moro", Bari, Italy

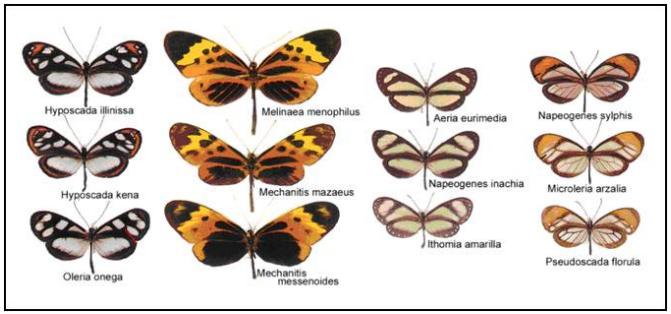
METAGENOMICS: AN EXTRAORDINARY TOOL TO INVESTIGATE BIODIVERSITY

Biodiversity, i.e. the richness and variety of life forms and ecosystems of our planet. Since every living organism has a genetic outfit specific of both the species and the individual, it follows that the molecular approach based on the characterization of the ensemble of the genetic material extracted from a given environment (e.g. Metagenomics) is a powerful and objective tool to determine the variety and the relative abundance of species (and their genome potential) populating it.

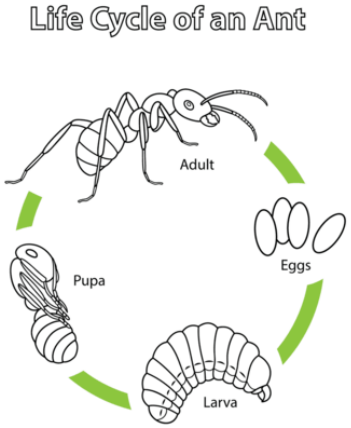
MOLECULAR BIODIVERSITY

Investigating Biodiversity through Molecular Approaches offers a number of advantages over traditional methodologies, that still may provide relevant complementary information

Criptic species



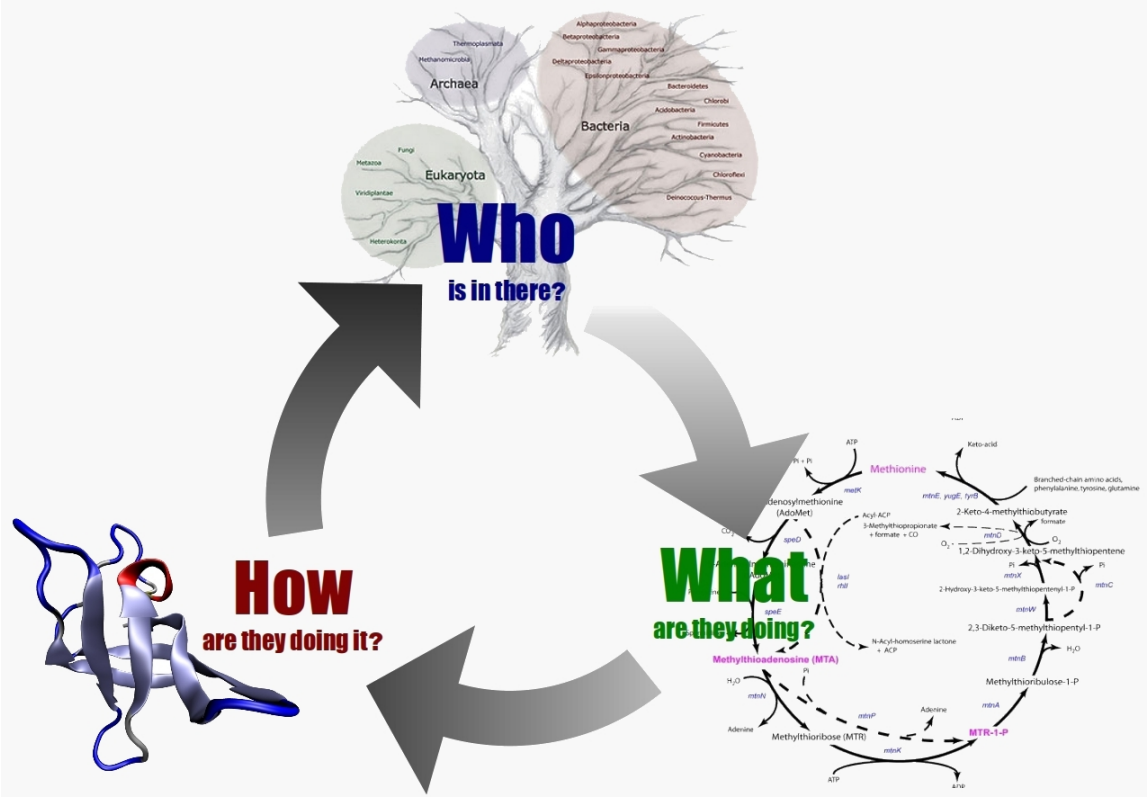
Development stages



Microbial Biodiversity

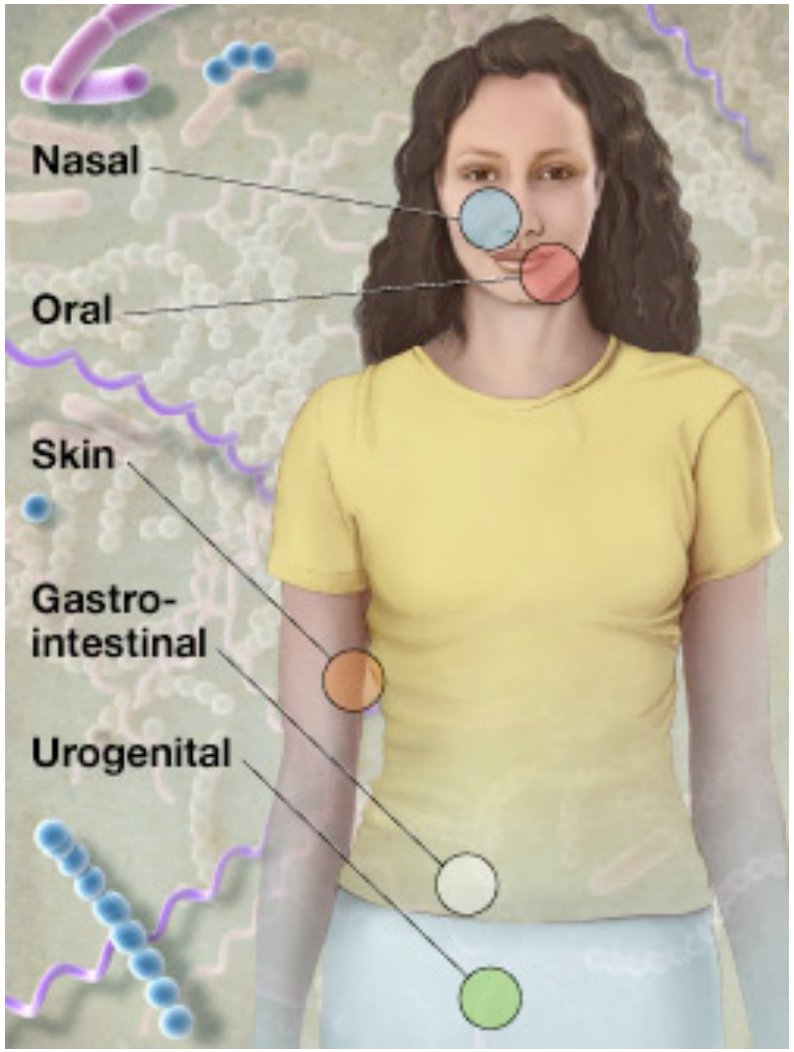
METAGENOMICS

A deep-wide view of the Microbiome Biodiversity



Metagenomics allows to unveil the composition and function of mixed microbial communities in any environmental niche, **including those which could not be isolated and cultivated in the lab (≥ 98%)**. Indeed, the Biodiversity of each environment, i.e. all living organisms (mostly microbial) can be fully represented by their genetic material.

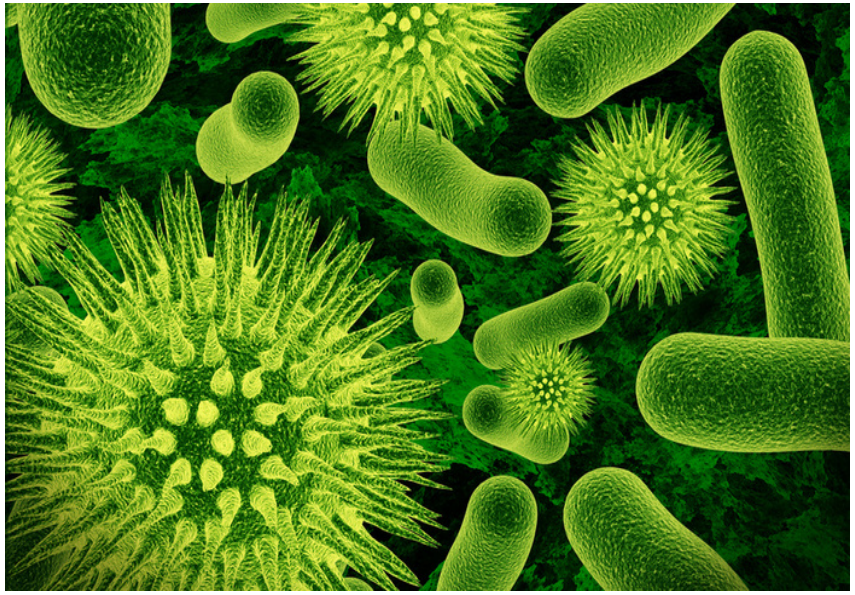
The Human Microbiome: a forgotten organ



Our body is an extremely complex environment, made of about 10^{13} cells, but containing at least tenfold more (about 10^{14}) bacterial cells. The so called "human microbiota" has a profound influence on **physiology**, **nutrition**, and it is crucial to our **health**. In fact, it provides nutrients and vitamins, supports the response to infection and detoxification of several toxic substances.

Some microbes are **native**, normally found in the body, Some other microbes are **introduced**, suddenly arriving at a new residence in the body

The Human Microbiome: a forgotten organ

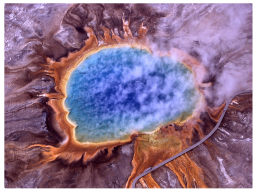


- More than **10,000 microbial species** (including bacteria, fungi, protists and viruses) have been identified as living in human bodies thus far.
- Microbial cells outnumber human cells in a human body **10:1**!
- Microbes account for **1-3% of human body weight.**
- The microbiome contributes **8 million protein-coding genes**.

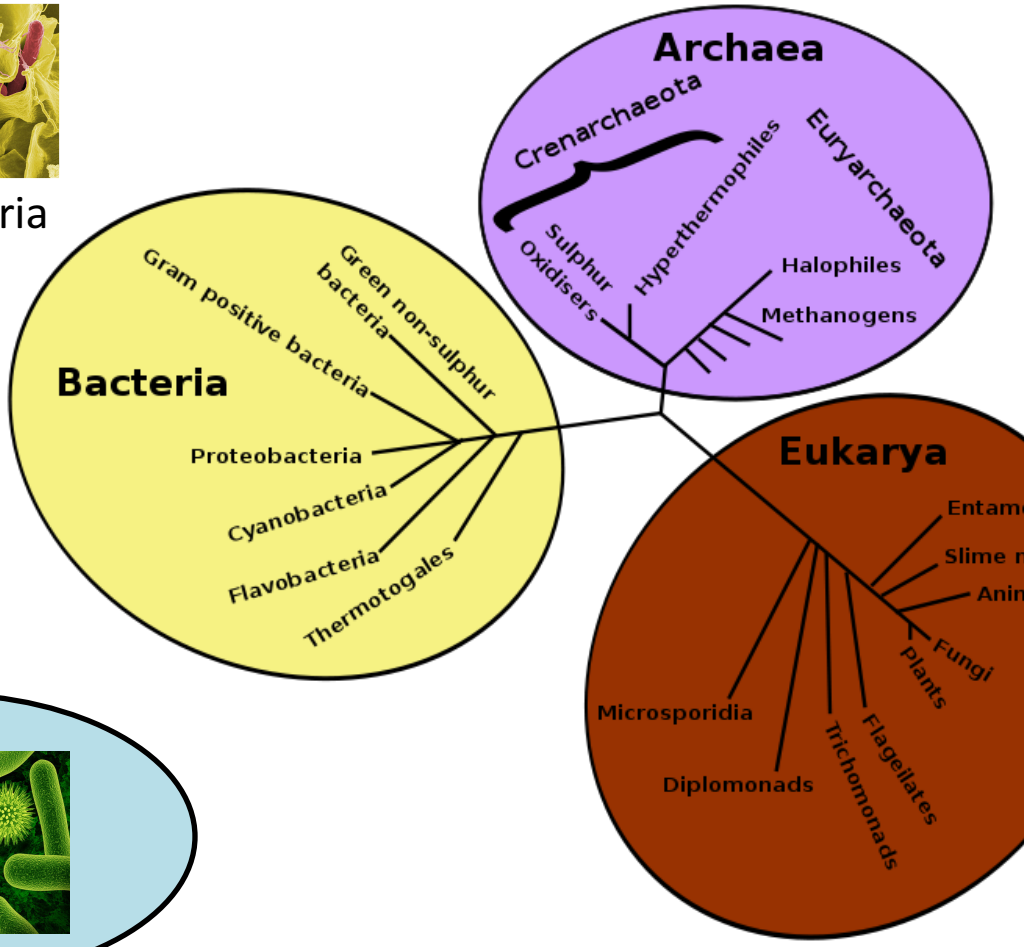
The Human Microbiome Project is studying these microbial communities in healthy individuals at different sites on the body, including nasal passages, mouth, skin, GI tract and UG tract (http://commonfund.nih.gov/hmp/)

Microbes in the Human Microbiome include species from each major domain

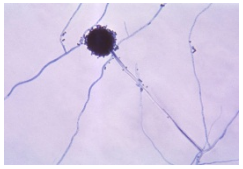
Bacteria



“Extremophile” Archaebacteria



Viruses



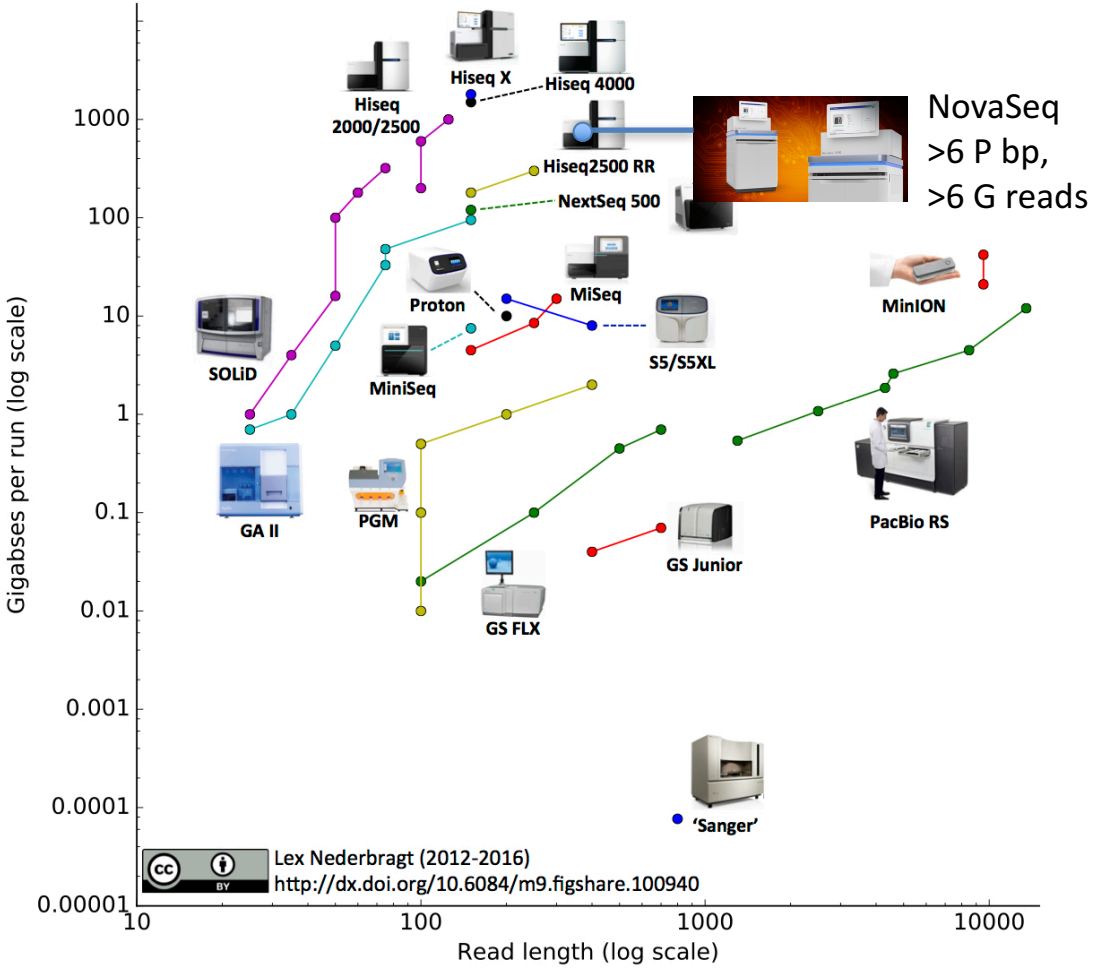
Fungi

Gut Microbiota and pathogenesis

The gut microbial composition may affect the metabolism and energetic homeostasis (e.g. obesity) and has been found associated to several disease states.

- **obesity**
- **type 1 diabetes**
- **childhood asthma**
- **inflammatory bowel disease (e.g. Cronh disease)**
- **colorectal cancer**
- **cardiovascular disease**
- **human immunodeficiency**
- **anxiety**
- **respiratory infections**

Second and Third Generation Massive Sequencing



Volume of Data

Cost / bp
(about 30 cent / Mb)

Molecular Approaches for investigating Biodiversity

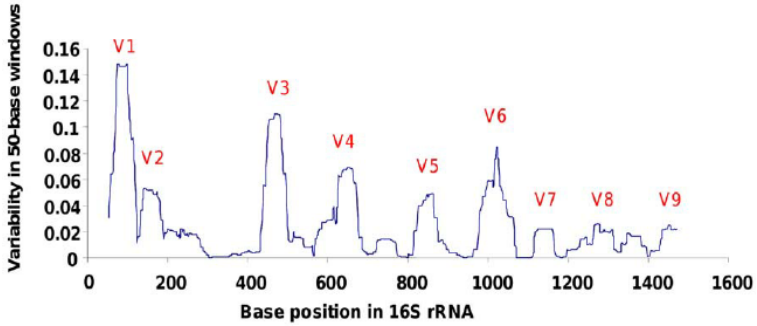
DNA barcoding

1

Sanger

DNA metabarcoding

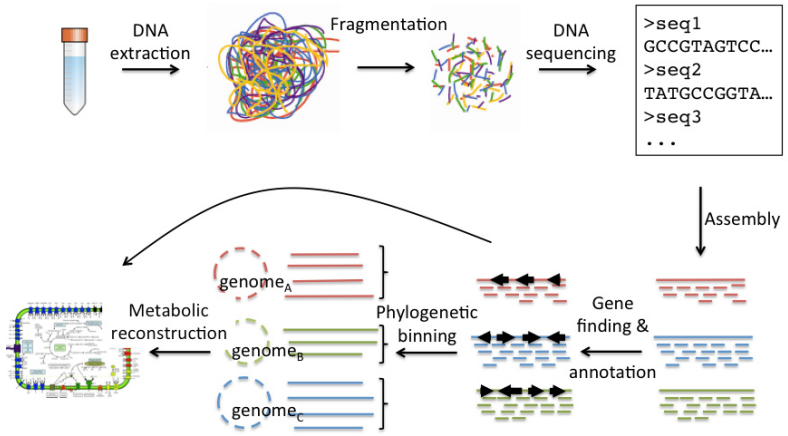
2



HTS

Shotgun Metagenomics

3



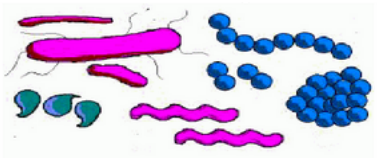
HTS

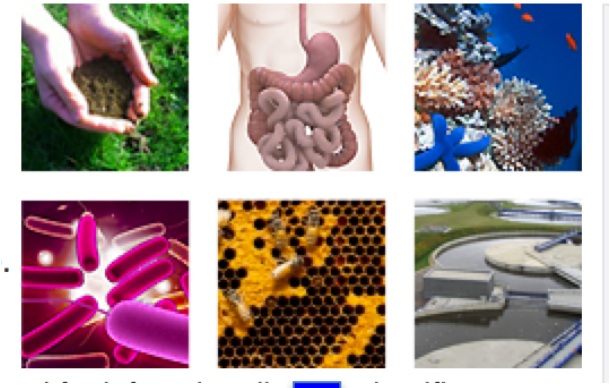
Functional Metagenomics

4

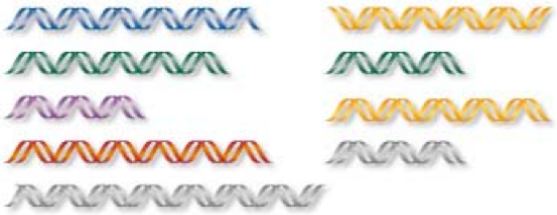
HTS

DNA Metabarcoding and Shotgun Metagenomics





Clinical or Environmental samples



DNA/RNA extraction

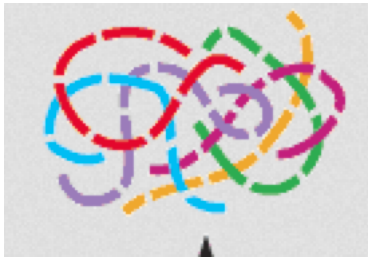
DNA Metabarcoding approach

Shotgun approach

metabarcode

(e.g. 16S rRNA, ITS, Cox1, ..)

NGS library
Deep sequencing
NGS data analysis



Target-oriented Metagenomics or DNA Metabarcoding

We need to define a target DNA sequence which is **ubiquitously** present in the taxonomic group under investigation, and fits some peculiar features:

Highly conserved flanking regions which allow amplification and sequencing in a broad range of taxa with universal primers

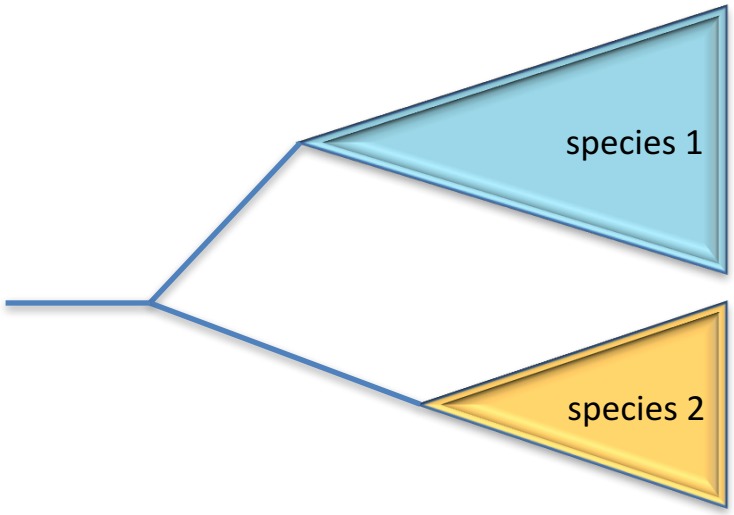
Suitable variability of the internal region to easily discriminate between intra-specific and intra-specific differences.

Amplicon size fulfilling NGS platform technical features

Metabarcoding

Different target regions have been defined for different taxonomic groups. Specific recommendations have been issued by CBOL (Consortium of the Barcode of Life, http://www.barcodeoflife.org/).

Taxon	Target
Bacteria	variable regions of 16S rRNA
Fungi	Internal transcribed spacer regions of the rRNA cluster (ITS)
Metazoa	mitochondrial Cox1
Plants	plastid matK and rbcL

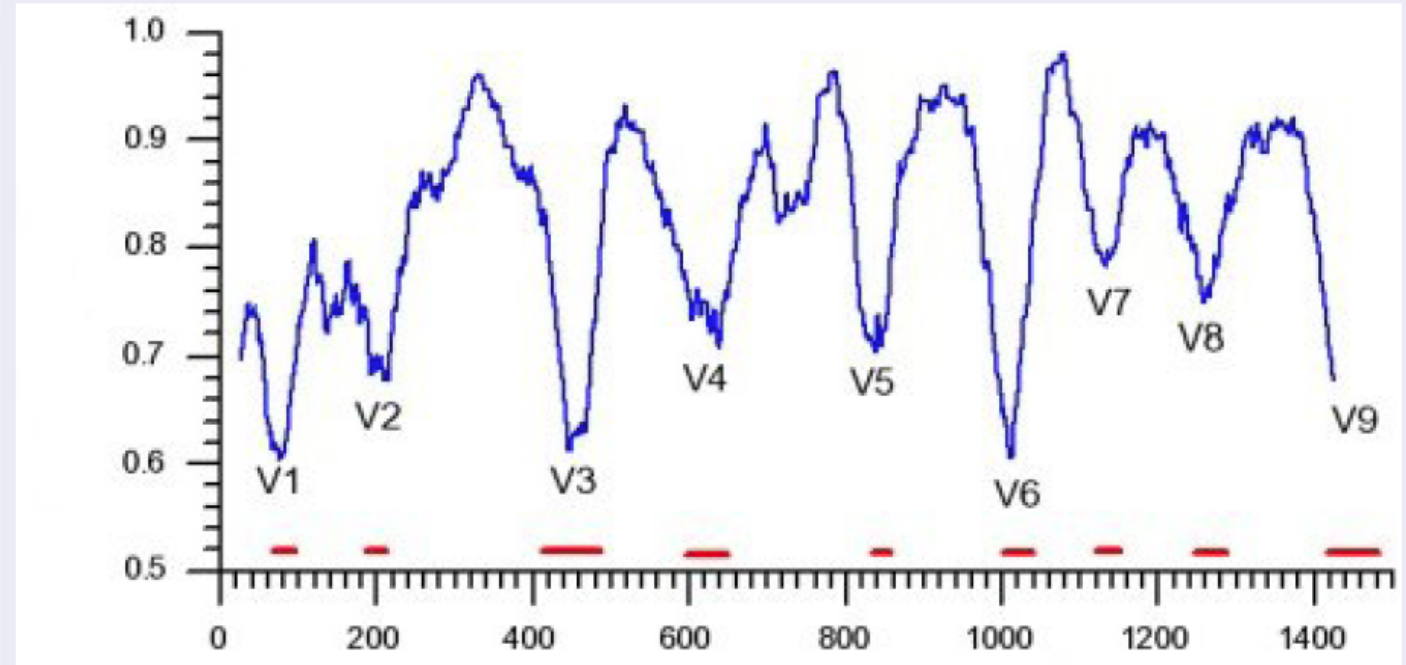


In order to work as reliable barcodes the selected target should be easily amplifiable by using primers universal for all species in the taxon under investigation, and the minimum interspecific variability (d_{K2P}) larger that the maximum intraspecific variability.

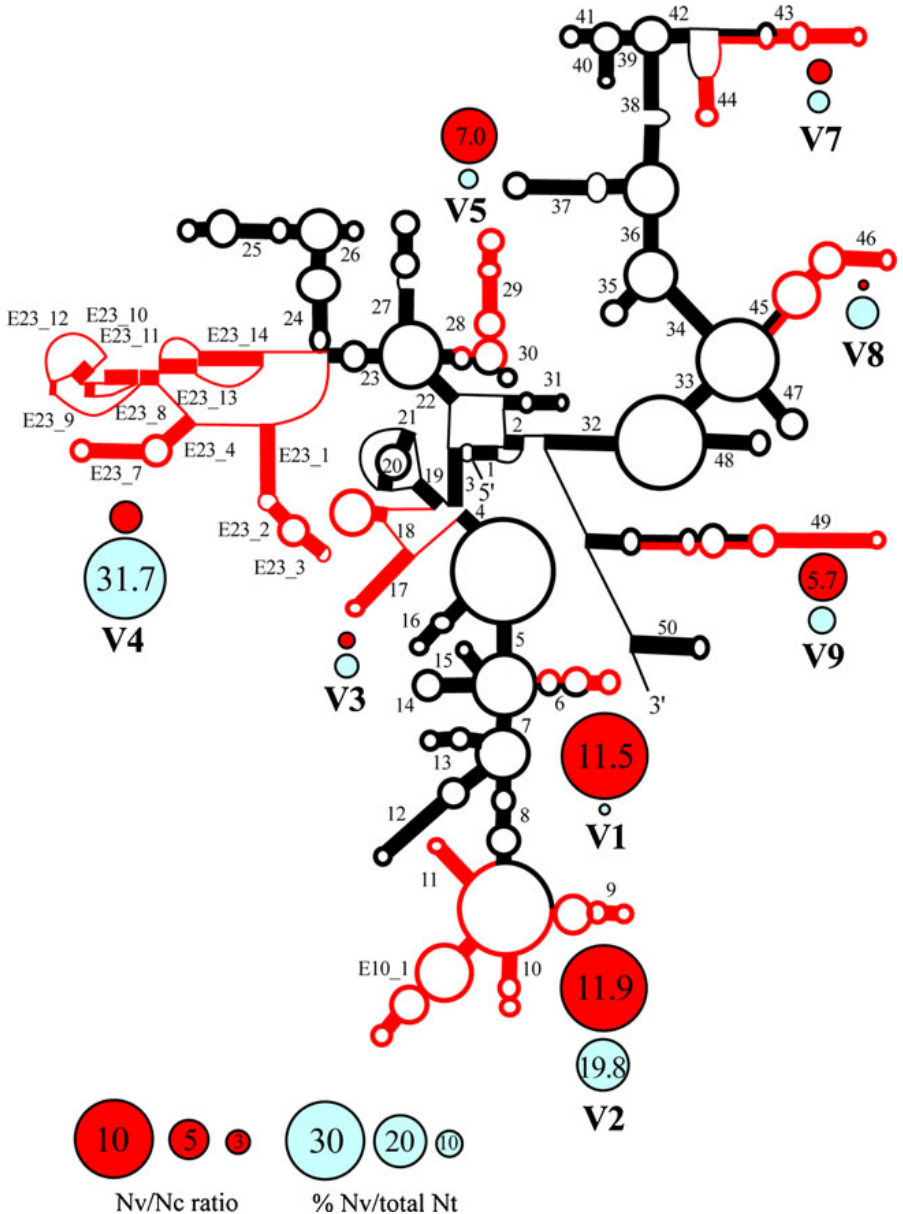
Hypervariable 16S rRNA regions for Bacterial Metabarcoding

The bacterial 16S rRNA is composed of nine hypervariable regions (V1-V9) interspersed with highly conserved regions. The sequence of the 16S rRNA gene and its hypervariable regions has been determined for a large number of organisms, and is available from multiple databases such as the Ribosomal Database Project. For taxonomic classification, it is sufficient to sequence individual hyper-variable regions instead of the entire gene length.

Small subunit 16S ribosomal RNA structure



Hypervariable 18S rRNA regions for protists meta-barcoding



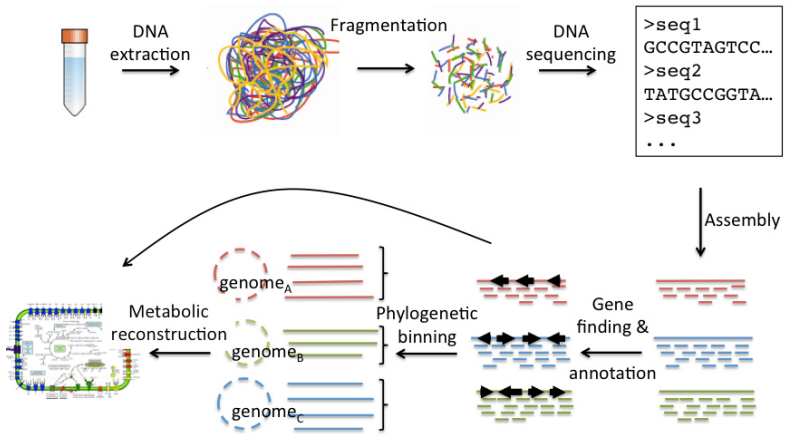
A similar conservation pattern is observed in protists 18S rRNA where the same hypervariable regions (V1-V9) interspersed with highly conserved regions are observed. The **V4** and **V9** regions are the most commonly used.

Problematic issues with 16S/18S rDNA metabarcoding

1. Copy-number variation bias
2. Variability in amplification efficiency
3. Inconsistencies when targeting different variable regions of the gene
4. Problems with accurately and consistently delineating prokaryotic/eukaryotic species

Shotgun Metagenomics

Shotgun metagenomics of total nucleic acids extracted by an environmental samples can be carried out at level of DNA or RNA and may give simultaneously a snapshot of the taxonomic and functional composition of the microbiome. When there are no suitable targets (e.g. viruses) and the amplicon-based strategy is unfeasible, the shotgun is the only possible approach.



Uncovering Earth's virome

David Paez-Espino[1], Emiley A. Eloe-Fadrosh[1], Georgios A. Pavlopoulos[1], Alex D. Thomas[1], Marcel Huntemann[1], Natalia Mikhailova[1], Edward Rubin[1,2,3], Natalia N. Ivanova[1] & Nikos C. Kyrpides[1]

Viruses are the most abundant biological entities on Earth, but challenges in detecting, isolating, and classifying unknown viruses have prevented exhaustive surveys of the global virome. Here we analysed over 5 Tb of metagenomic sequence data from 3,042 geographically diverse samples to assess the global distribution, phylogenetic diversity, and host specificity of viruses. We discovered over 125,000 partial DNA viral genomes, including the largest phage yet identified, and increased the number of known viral genes by 16-fold. Half of the predicted partial viral genomes were clustered into genetically distinct groups, most of which included genes unrelated to those in known viruses. Using CRISPR spacers and transfer RNA matches to link viral groups to microbial host(s), we doubled the number of microbial phyla known to be infected by viruses, and identified viruses that can infect organisms from different phyla. Analysis of viral distribution across diverse ecosystems revealed strong habitat-type specificity for the vast majority of viruses, but also identified some cosmopolitan groups. Our results highlight an extensive global viral diversity and provide detailed insight into viral habitat distribution and host-virus interactions.

BIOINFORMATICS TOOLS FOR METAGENOMICS ANALYSIS

1. DNA Metabarcoding

BioMaS

Bioinformatic analysis of Metagenomic AmpliconS

2. Shotgun Metabarcoding

MetaShot

Workflow for taxon classification of host-associated microbiome from shotgun metagenomic data

3. Functional Metagenomics

A Galaxy Suite for Applied targeted metagenomics

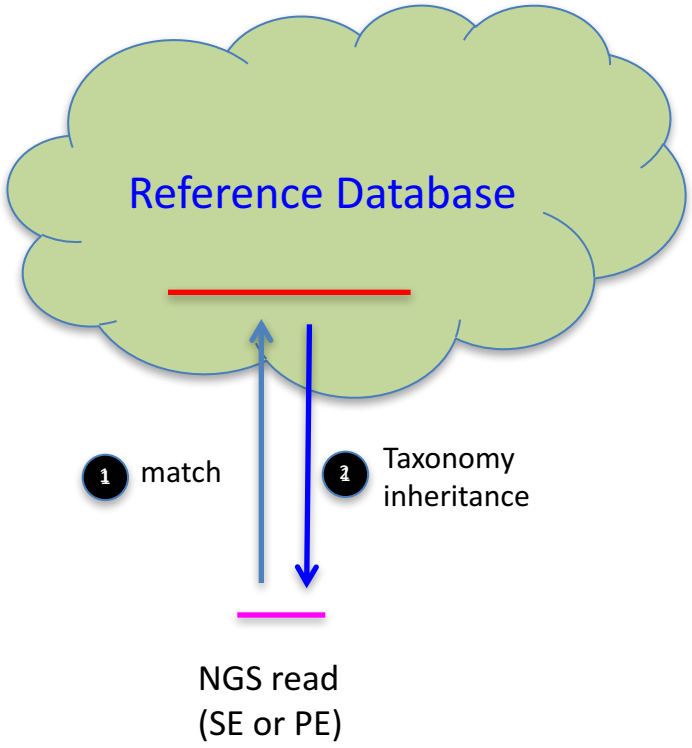
Bioinformatics for meta-barcoding data analysis

The accessibility to proper and user friendly computational systems for large-scale bioinformatic analysis of metabarcoding data is still a challenge.

BioMaS (Bioinformatic analysis of Metagenomic AmpliconS) is a new cloud based pipeline designed to provide biomolecular researchers, without specific computer skills and involved in taxonomic studies of environmental microbial communities, with a completely automatic workflow, comprehensive of all fundamental steps, from raw sequence data upload and cleaning to final taxonomic identification.

Bioinformatics for metabarcoding data analysis: taxonomic assignment of NGS reads

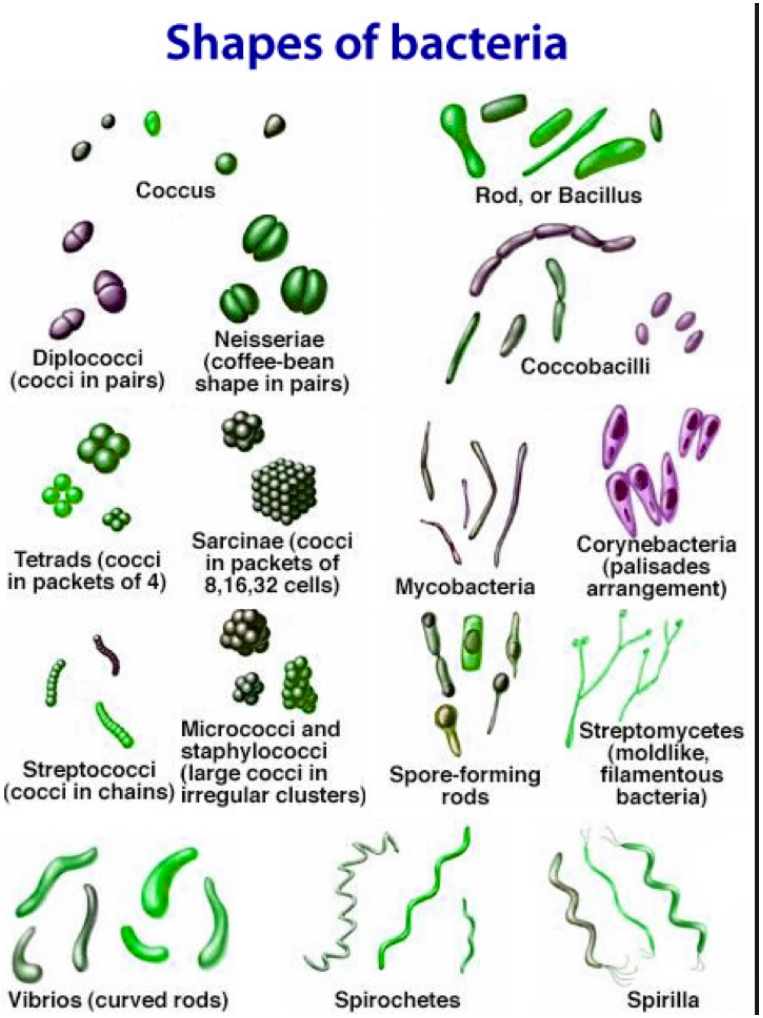
A critical step of metagenomics analysis is the taxonomic assignment of NGS reads. This is typically done by using a **similarity-based approach** against a reference databases.



This is not trivial for two major reasons:

- A NGS read may typically show several “best matches” against the reference database (also depending on the selected cutoff)
- Some of the reference database entries may have a wrong or incomplete taxonomic assignment (e.g. uncultured bacterium)

Defining a bacterial species in the genomic era



The assignment of microbial organisms to species or higher taxonomic ranks is traditionally done based on stable phenotypic characters (e.g. cellular morphology, growth requirement and other metabolic traits).

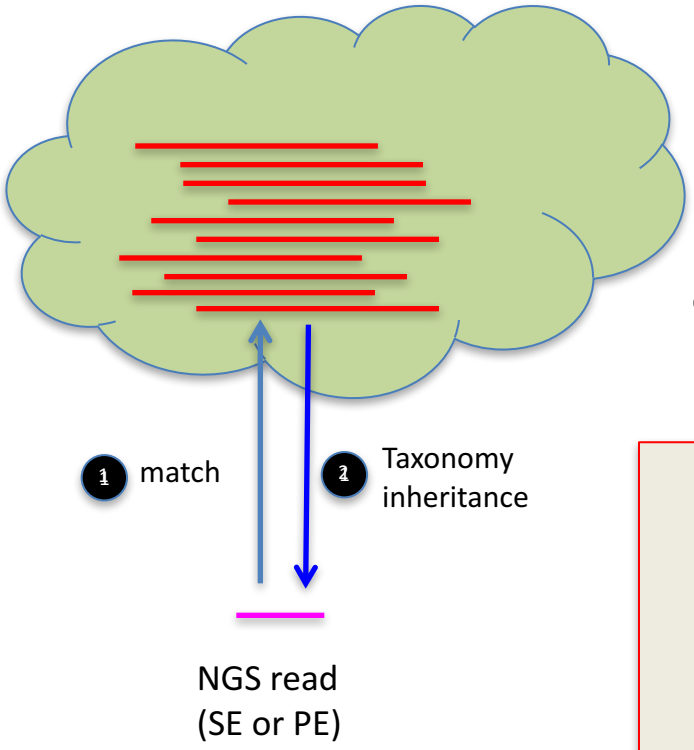
In consideration that most of microbial strains cannot be cultured, a genomic-based classification is currently used:

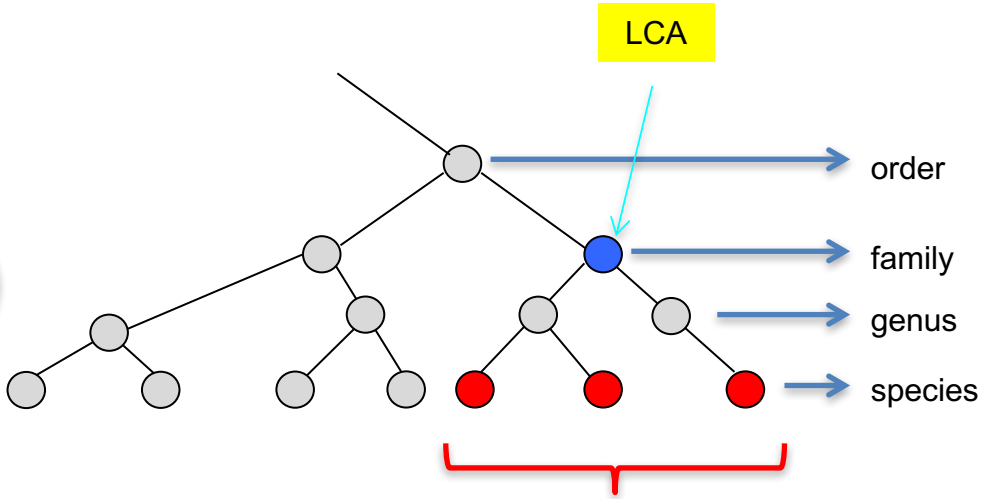
- **≥97% identity of 16S/18S rRNA**
- **≥95% identity at whole genome level**

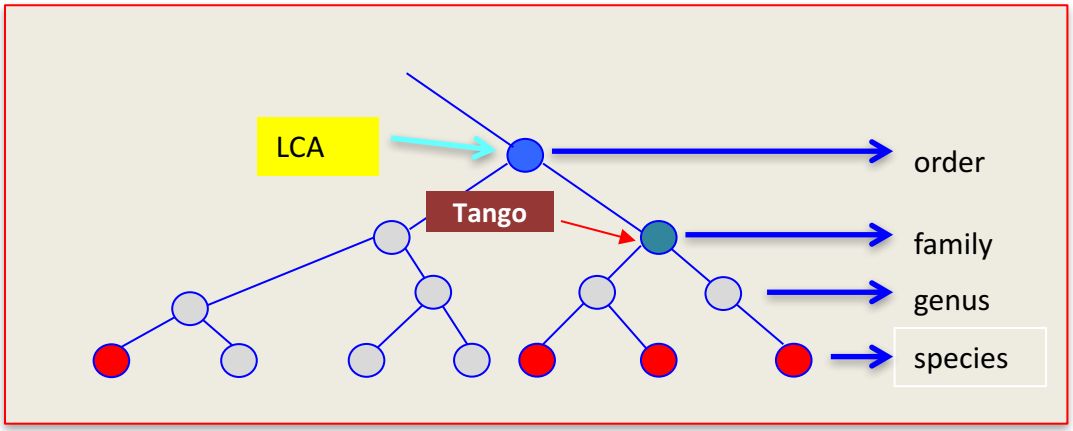
See: Chan et al. BMC Microbiology 2012:12:302

NGS reads taxonomic assignment

If the read shows a “single” best match against the Reference database, the taxonomic assignment is trivial. In general, more than a single best match is found, and the assignment is done either to the “last common ancestor” (LCA) or to the “most likely” taxon rank (i.e. **TANGO**).

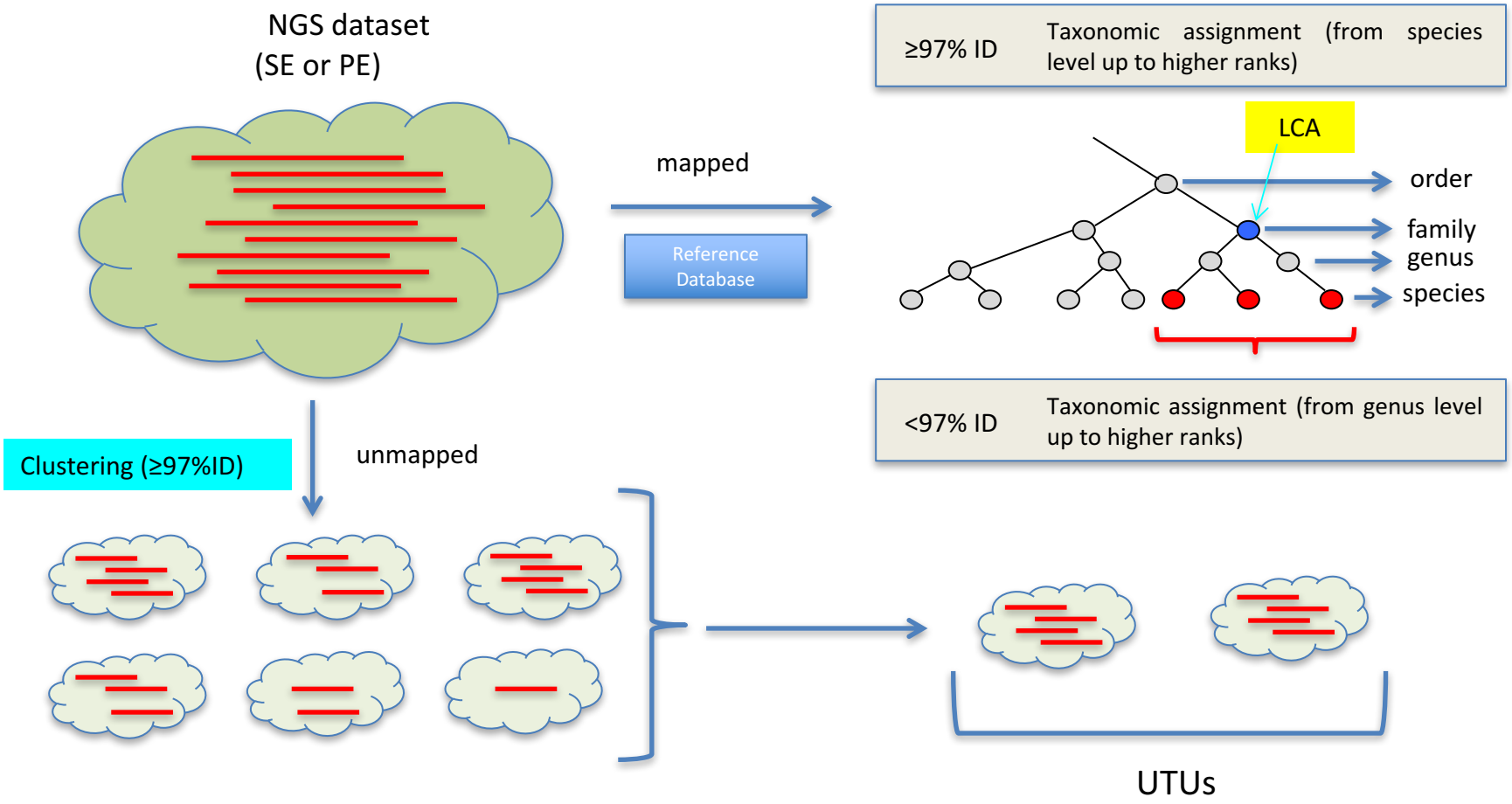






NGS reads taxonomic assignment

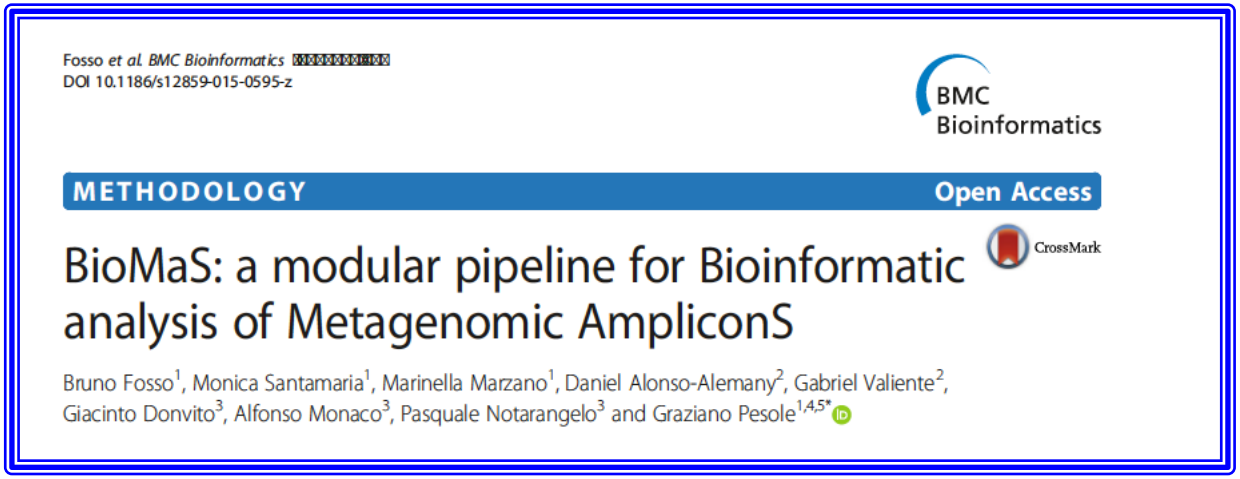
The **97% sequence identity cutoff** is routinely applied to assign a read to a species based on the 16S/18S rRNA analysis. Residual mapped reads may be classified at higher taxonomic ranks. Finally, unmapped reads may represent novel species which can be named UTUs (unknown taxonomic units)



BioMaS

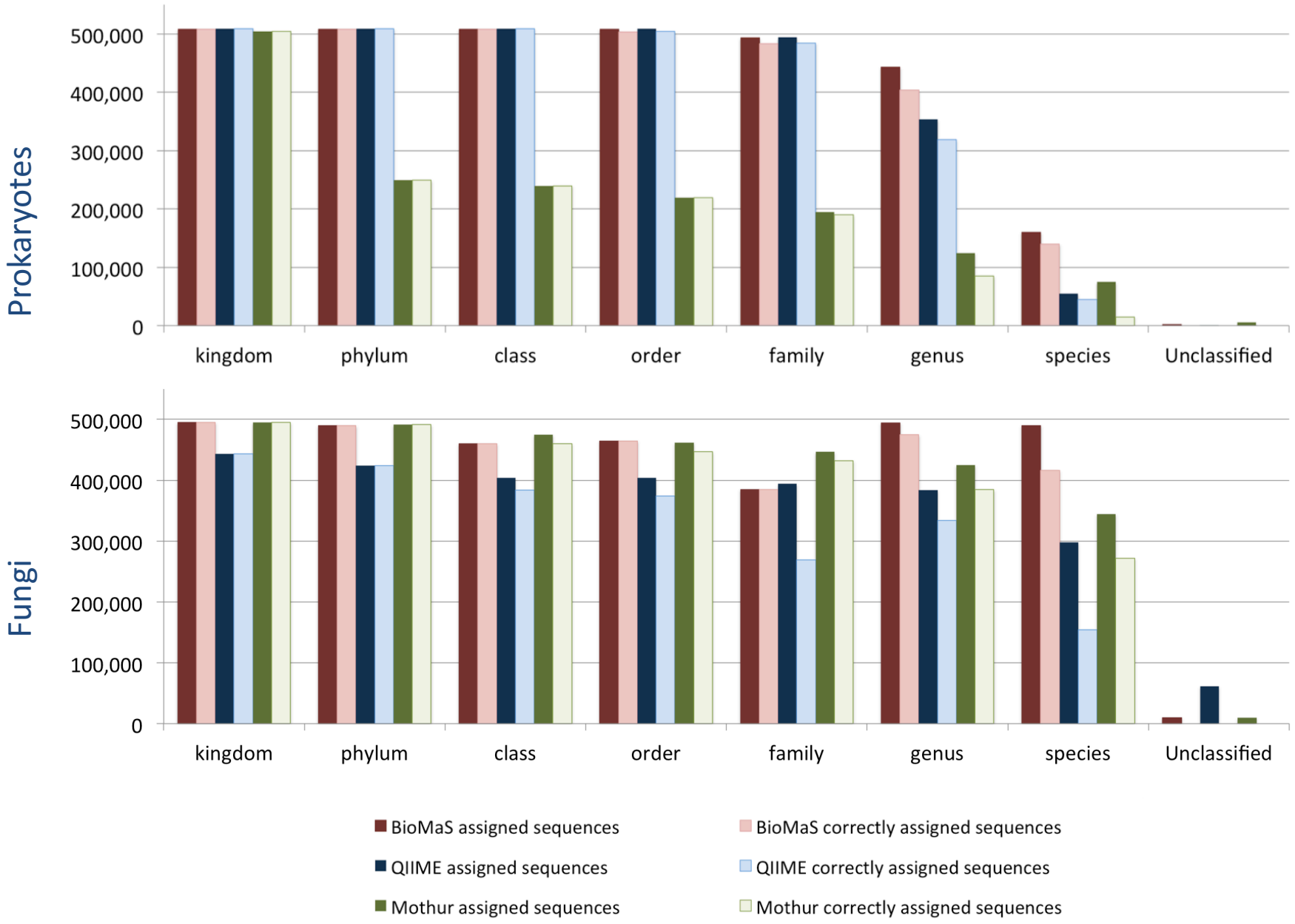
Bioinformatic analysis of Metagenomic ampliconS

BioMaS **(Bioinformatic analysis of Metagenomic ampliconS)** is a modular pipeline able to convert the Illumina raw data in taxonomic assignments.

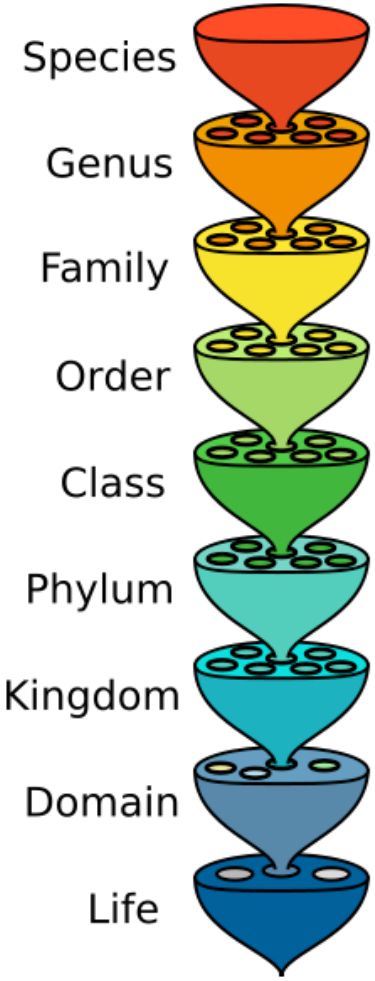


1. Data quality check, statistical evaluation and consensus generation
2. Similarity assessment (NGS reads vs Reference sequences)
3. Taxonomic assignment by TANGO (Alonso-Alemany et al. 2014)
4. Statistical analysis and Results visualization

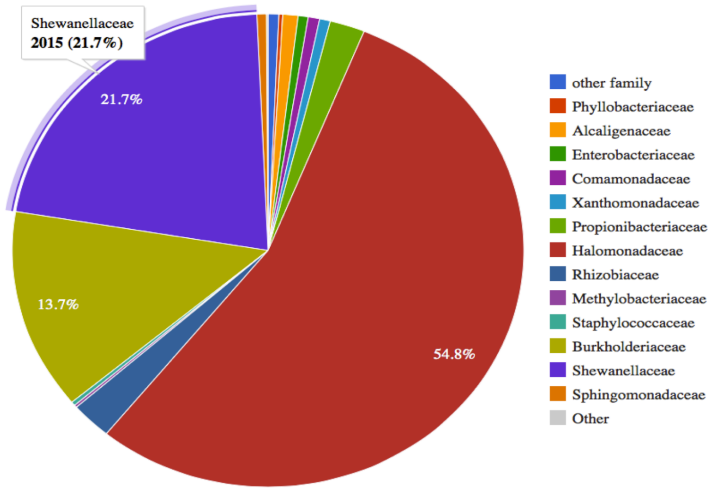
BioMaS performance benchmark



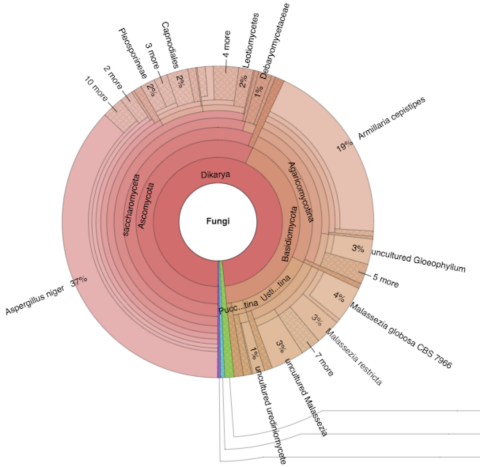
BioMaS report files



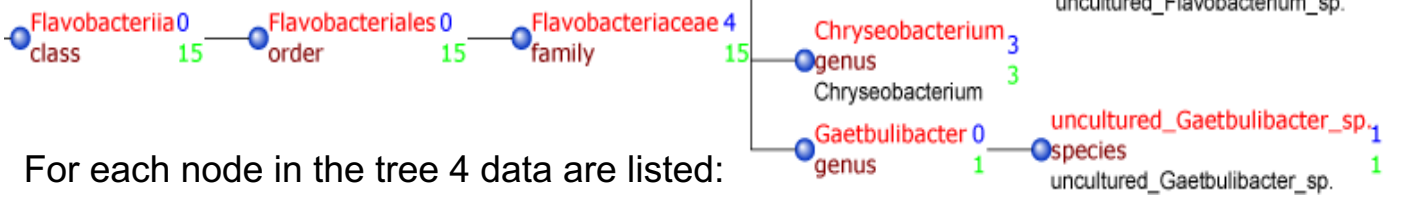
Pie chart taxonomic (Family) distribution



Krona Graph



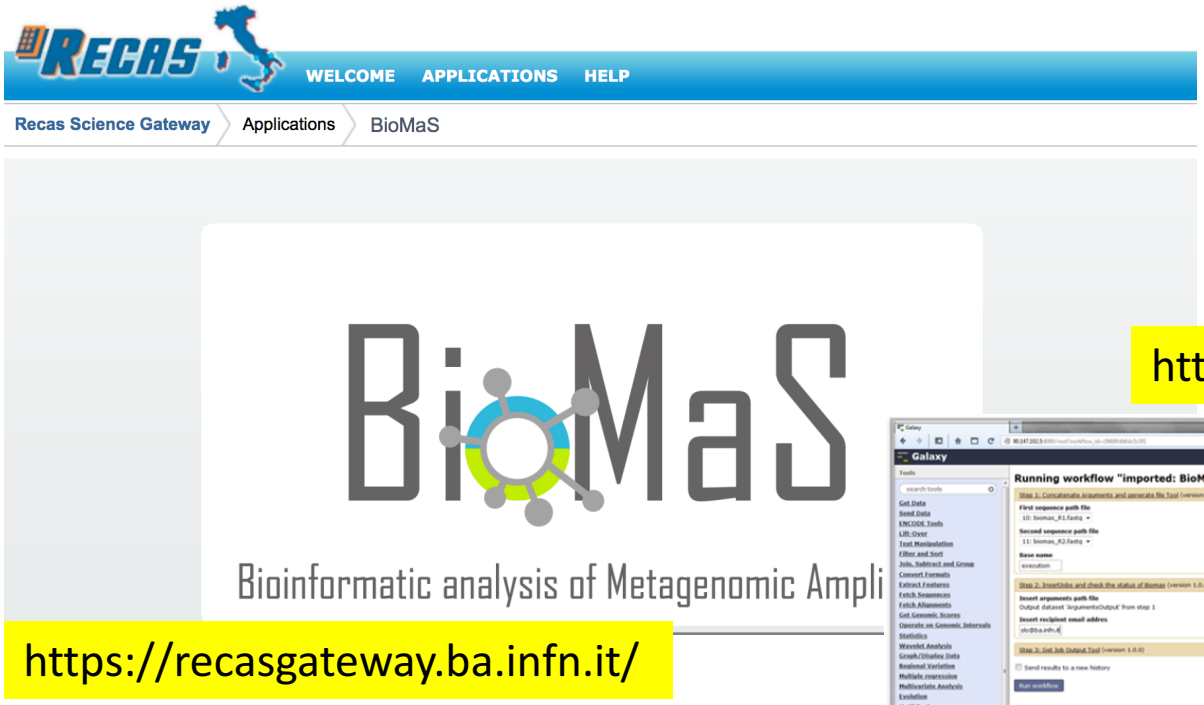
Tree-based taxonomic representation



For each node in the tree 4 data are listed:

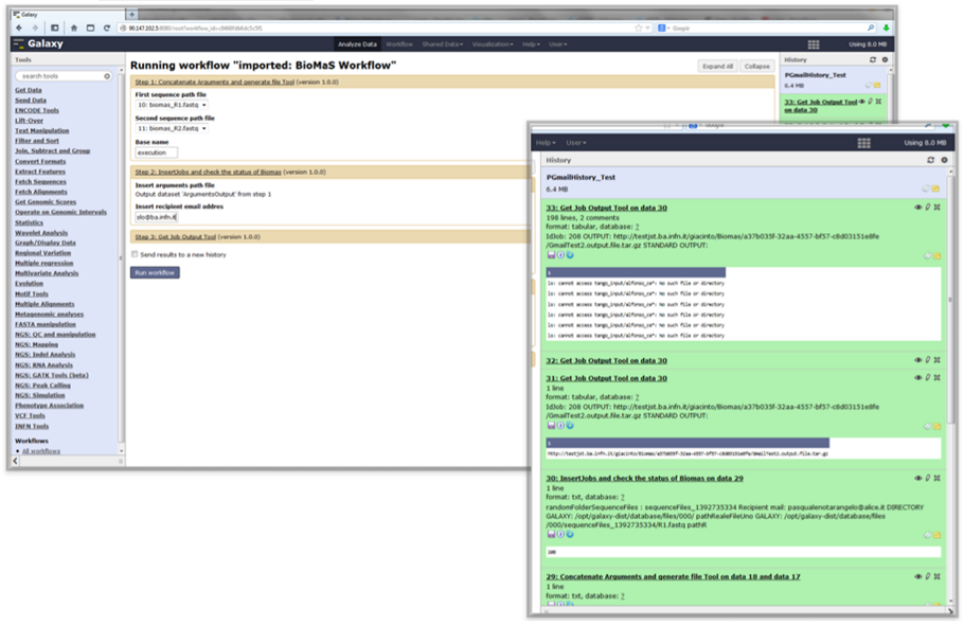
- Node scientific name (red)
- Taxonomic rank (brown)
- Number of reads directly assigned to the node (blue)
- Number of reads assigned to the the node and to its descendants (green)

BioMaS cloud-based web services



https://recasgateway.ba.infn.it/

http://galaxy.cloud.ba.infn.it:8080

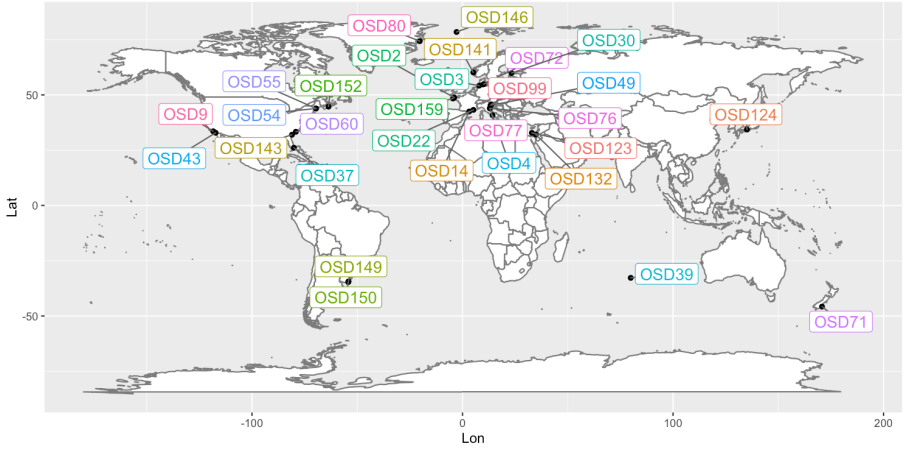


Furthermore, BioMaS Post-Processing Tools are also available, in order to produce an output file suitable for the comparative analyses and differential abundance analysis through METAGENassist (http://www.metagenassist.ca/METAGENassist/)

TEST CASE 1 - ASSESSMENT OF V4 AND V9 REGIONS IN THE CHARACTERIZATION OF THE EUKARYOTIC MICROBIOME OF OSD SAMPLES BY 18S rRNA METABARCODING DATA

1. The V4 and V9 hypervariable regions of 18S rRNAs are both generally used for exploring the eukaryotic plankton diversity in oceans.

2. The aim of this investigation was to compare the performance of these two barcode regions in the taxonomic characterization, at qualitative and quantitative level, of the eukaryotic microbiome from metagenomic samples collected in the 2014 OSD summer campaign.



Under the support of Lifewatch Italy 30 of 191 metagenomic samples were investigated generating metabarcoding data for both V4 and V9 according to the standard NE08 OSD protocol.

>350,000 V4 PE reads (2x250 bp) and V9 PE reads (2x150 bp) per sample were generated and analyzed using both BioMaS and QIIME.

BioMaS

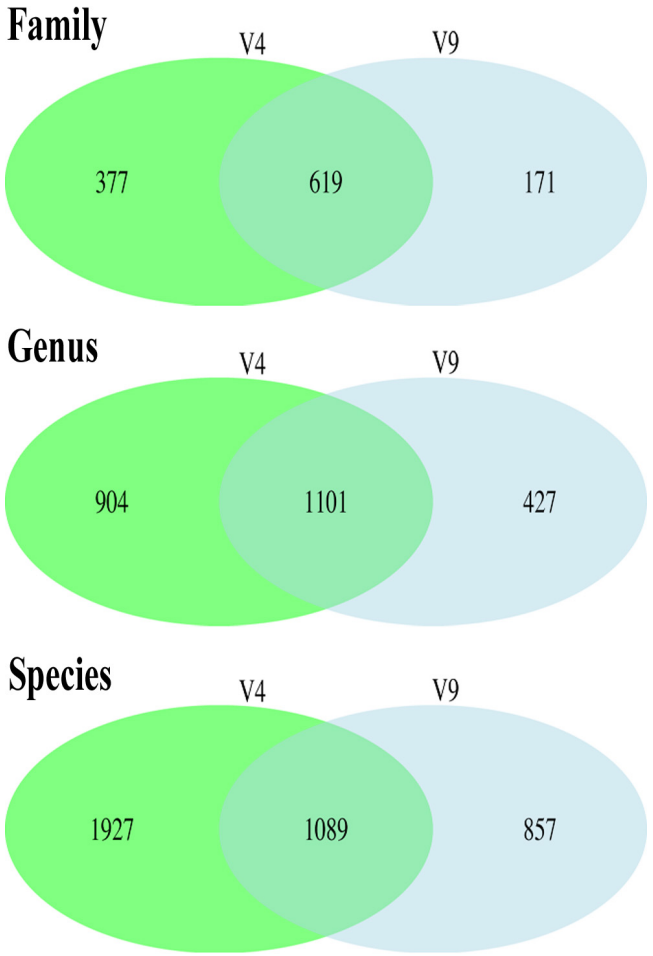
Bioinformatic analysis of Metagenomic AmpliconS

The BioMaS pipeline was applied to taxonomically annotate the V4 and V9 PE reads by using as reference the release 119.1 of the SILVA database.

BioMaS was able to classify about the 94% and 79% of V4 and V9 sequences, respectively.

Considering all the analysed samples a total of ***3,873*** species, ***2,432*** genera and ***1,167*** families were detected

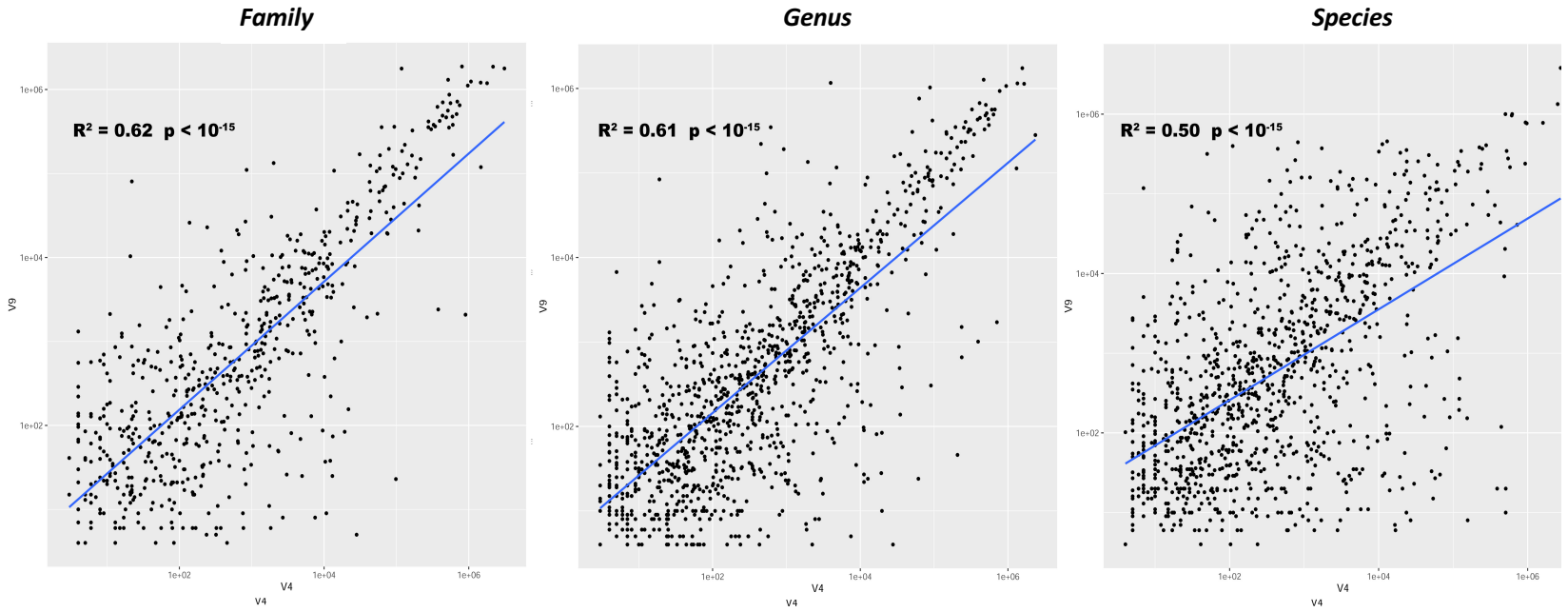
As expected, a higher number of taxa were detected by V4 than V9, but quite unexpectedly a remarkable number of taxa was detected by V9 only.



BioMaS

Bioinformatic analysis of Metagenomic AmpliconS

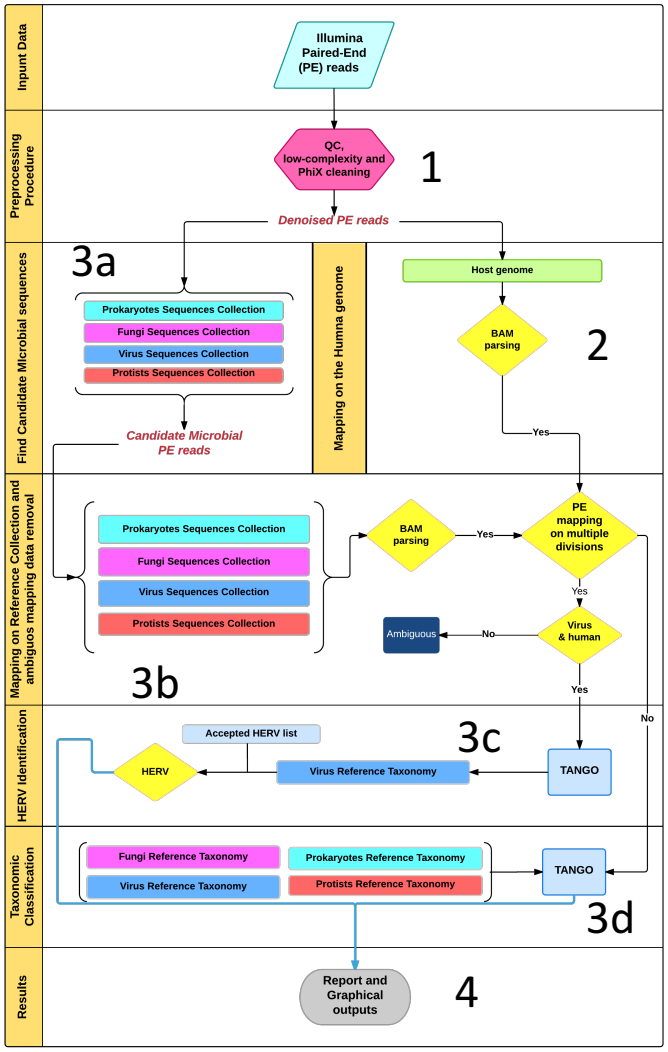
The taxon abundances as resulting by the normalized read count were compared for V4 and V9 data



Taxa coverages obtained from V4 and V9 normalized counts were significantly correlated. The qualitative (number of taxa) and quantitative (taxa abundances) observed biases may be related to:

- Different marker sizes
- Taxa representation in reference database

MetaShot



MetaShot analysis procedure can be divided in 4 main processes:

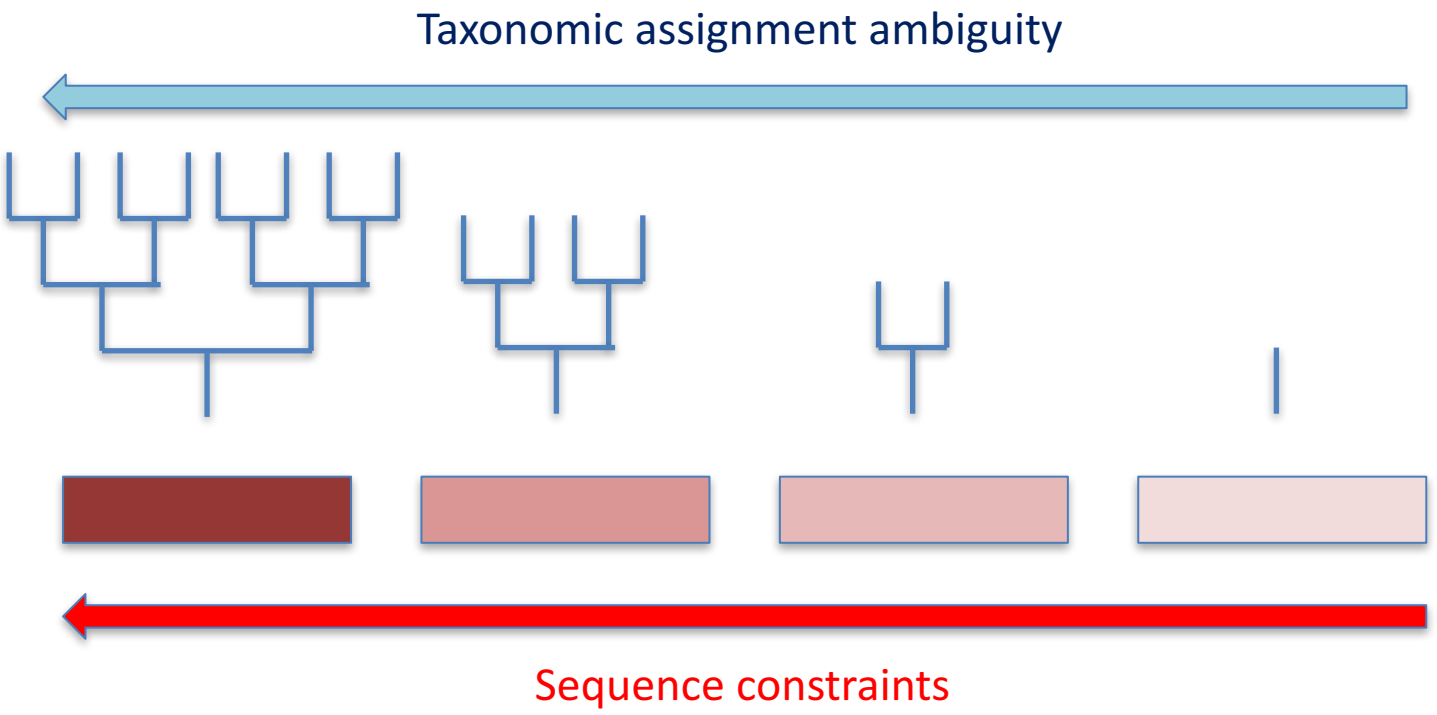
1. ***Pre-processing procedures***: input sequences containing low-quality/complexity regions and reads shorter than 50 nucleotides are removed.
2. ***Comparison with the human genome***: cleaned data are mapped against the human genome;
3. ***Comparison with reference databases and taxonomic annotation***: in ***MetaShot*** are implemented four reference collections for ***Prokaryotes***, ***Viruses***, ***Fungi*** and ***Protista***, obtained by processing data from GenBank and RefSeq databases. The taxonomic annotation is a 5 steps procedure:
   a) *Identification of candidate microbial sequences*;
   b) *Comparison with reference collections and exclusion of ambiguous reads*;
   c) *HERV identification;*
   d) *Taxonomic classification of unambiguous reads*.
   e) *Refinement of taxonomic assignments resolving ambiguities*
4. ***Report generation***: CSV files, HTML interactive tables and Krona Graphs summarizing the obtained taxonomy assignments.

MetaShot provides also a script to extract ambiguous, unclassified and taxon specific PE reads.

Fosso et al. MetaShot: an accurate workflow for taxon classification of host-associated microbiome from shotgun metagenomic data. **Bioinformatics**. 2017 Jun 1;33(11):1730-1732.

MetaShot - Taxonomic Refinement

We expect that actually occurring genomes in the metagenomic sample consist of both conserved and variable tracts, with the latter showing much less ambiguity in the taxon assignment than the latter.



MetaShot, after the initial taxonomic read assignment, carries out a refinement step where ambiguous reads are “resolved” assigning them to the highest ranking taxon (i.e. the one with the largest number of reads). Indeed, we expect that actual taxa are represented by both conserved and variable DNA tracts, and thus have an higher number of assigned reads.

MetaShot Benchmark

MetaShot has been benchmarked against ***Kraken*** and ***MetaPhlAn2*** using:

1. About 21M Illumina PE reads in silico simulated human microbiome :
   - 986,114 bacterial reads (4.8%)
   - 146,886 viral reads(0.7%)
   - 19,582,500 human reads (94,5%)

2. A mock community consisting in bacterial and viral species (Conceicao-Neto, et al., 2015):
   - 4 bacterial species
   - 9 viral species

MetaShot Benchmark (1)

The following statistics were measured:

1. For Family, Genus and Species ranks, the number of assigned and correctly assigned PE reads for Homo sapiens, Prokaryotes and Viruses were estimated;
2. For each species: Precision (P), Recall (R) and F-measure (F) were measured

	Assigned %			Correctly Assigned %		
	KR	MS	MP	KR	MS	MP
Huan (host)	100.00	99.18	0	100.00	99.99	0
Prokaryotes						
Family	57.41	97.91	5.16	96.77	98.37	97.59
Genus	55.01	98.14	4.96	95.92	98.17	98.02
Species	54.17	99.31	4.76	79.52	88.06	90.7
Viruses						
Family	74.78	97.74	49.32	99.16	98.53	98.48
Genus	101.88	97.39	66.85	99.37	99.75	99.30
Species	73.45	97.81	43.86	98.98	96.70	95.46

	Human (host)			Prokaryotes			Viruses		
	KR	MS	MP	KR	MS	MP	KR	MS	MP
P(%)	99.85	100.00	0	35.67	98.13	98.00	94.95	98.30	80.93
R (%)	100.00	99.97	0	35.16	84.52	87.31	92.77	98.19	79.32
F (%)	99.92	100.00	0	35.36	86.79	90.72	92.82	98.07	79.93
U (%)	0.00	1.04	99.99	55.28	2.44	94.50	4.25	3.94	30.74

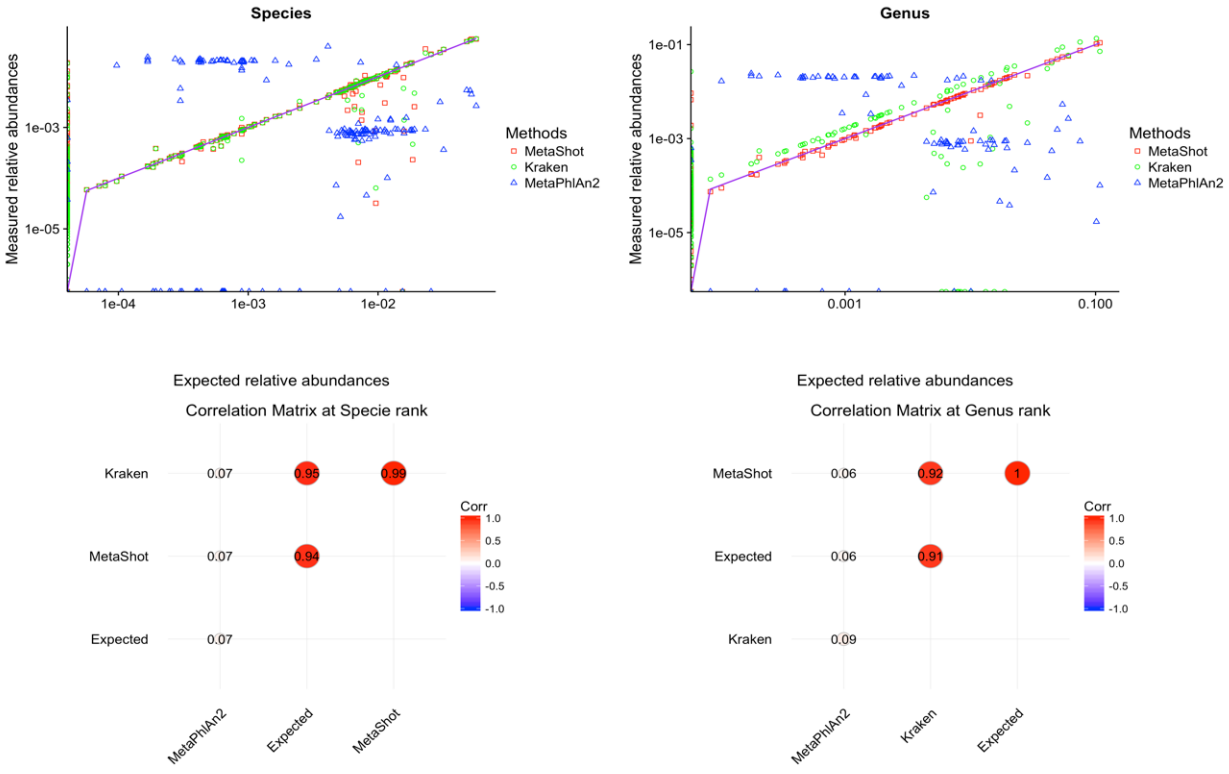
(U) Unclassified reads

MetaShot outperformed ***Kraken*** and ***MetaPhlAn2*** in terms of the overall accuracy of reads assignment for the Prokaryotes and Viruses at the Family, Genus and Species levels.

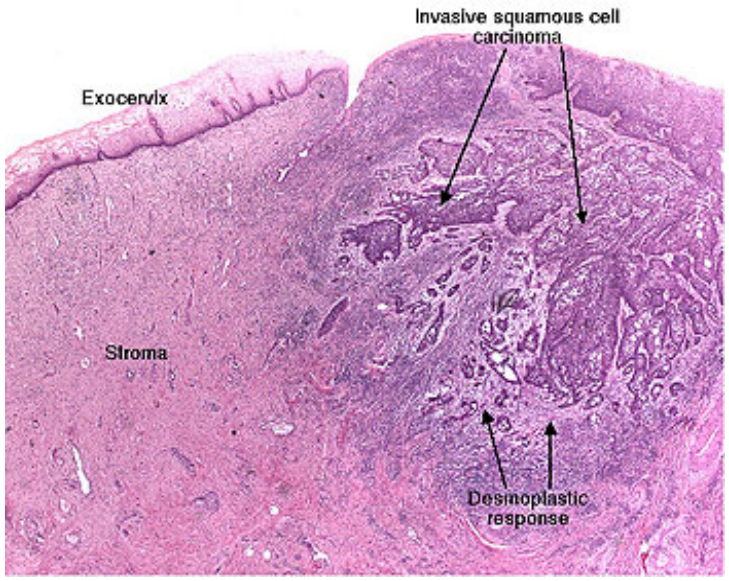
MetaShot Benchmark (2)

The observed taxon abundances, at Genus and Species level, were compared to the expected ones.

- Kraken and ***MetaShot*** estimations are quite close to the expected ones both at genus and species rank, while MetaPhlAn2 missed some of the expected taxa.
- The Pearson Correlation values (p-value ≤ e^{-16}) calculated comparing the abundances calculated by any of the considered tools and the expected abundances clearly show that the ***MetaShot*** quantitative estimates are overall the most accurate.



TEST CASE 2 – ANALYSIS OF THE CSC MICROBIOME



http://www.pathologyoutlines.com/topic/cervixSCC.html

Up to 9 out of 10 cervical cancers are **squamous cell carcinomas (CSC)**. These cancers develop from cells in the exocervix and the cancer cells have features of squamous cells under the microscope.

It is known that about 95% of these cancers harbor human papillomavirus (HPV) genomes, the specific serotype involved varies, the most common ones being HPV16, HPV18, and HPV31 (Growdon and Del Carmen, 2008).

We analyzed DNA-seq and RNA-Seq data from a sample of cervical squamous cell carcinoma of the uterus from a patient sample by applying ***MetaShot***.

Sample	PE reads	PP Pass	% PP Pass
DNA	528,034,456	512,253,714	97.01%
RNA	61,318,866	59,303,563	96.71%

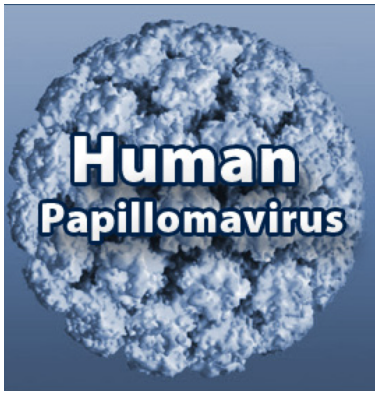
TEST CASE 2 – ANALYSIS OF THE CSC MICROBIOME

MetaShot Results

Taxonomic Analysis data					
Sample	Human	Prokaryotes	Virus	Fungi	Protists
DNA	501,609,424	4,242,577	25,368	14	28
RNA	52,312,428	63,567	14,150	91	1,836

Both for DNA and RNA data about the 98% of viral assignments were to **HPV serotype 31**. The presence of the HPV (Human Papilloma Virus) serotype 31 has been confirmed by PCR analysis.

The same data have been analysed by using Kraken, but it was unable to identify the presence of HPV serotype 31.



CHALLENGES

Metabarcoding and Shotgun Metagenomics poses specific challenges both in terms of the experimental design and of the amount and type of data to be produced,

1 Experimental design

- **Metabarcoding** (barcode choice) or **Shotgun**?
- # of replicates
- DNA or RNA

2 Amount and quality of data

- # of reads
  - 100-500 K /sample — 50-200 Mb
  - 100-300 M /sample — 20-90 Gb
- Read length (SE or PE)
  - 250 x 2
  - 100 x 2

3 ICT Infrastructure for data storage and computing

- 10-60 min CPU
- 20-50 CPU h

Problems and Pitfalls

Several errors and inconsistencies in nucleotide and taxonomy reference collections may severely bias results in both Metabarcoding and Shotgun data analysis.

1

Errors in ENA/NCBI nucleotide entries

(e.g. some reference genome contigs are contaminated with microbial / viral sequences)

2

Errors/Ambiguities/Differences in Reference Taxonomies

(e.g. Different reference taxonomies can be used in Data analysis such as NCBI Taxonomy, RDP, Greengenes, Silva which may bias taxonomic binning of generated reads)

PHIX CONTAMINATION IN REFERENCE GENOMES

The analysis of RNAseq shotgun data (240 M reads, SE 100 bp) from a skin biopsy of a patient with a disfiguring disease aimed at identify the possible etiological agent identified about **40k reads** assigned to **Babesia**, a intra-erytrocytic protist parasite.

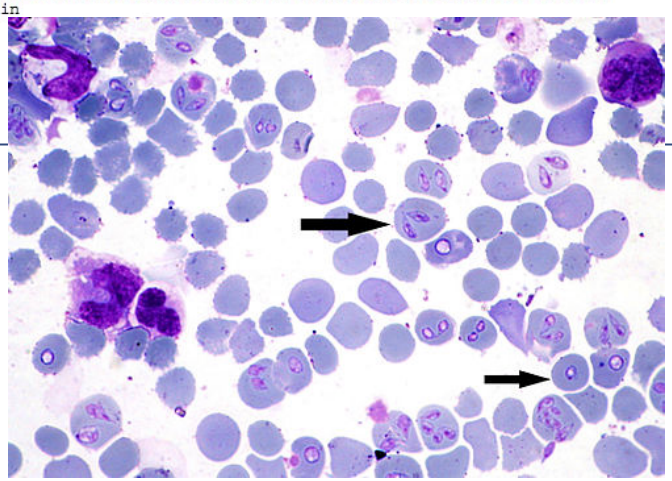
Babesia divergens genome assembly 454hybrid_PBjelly, scaffold Contig1323

GenBank: LK936033.1

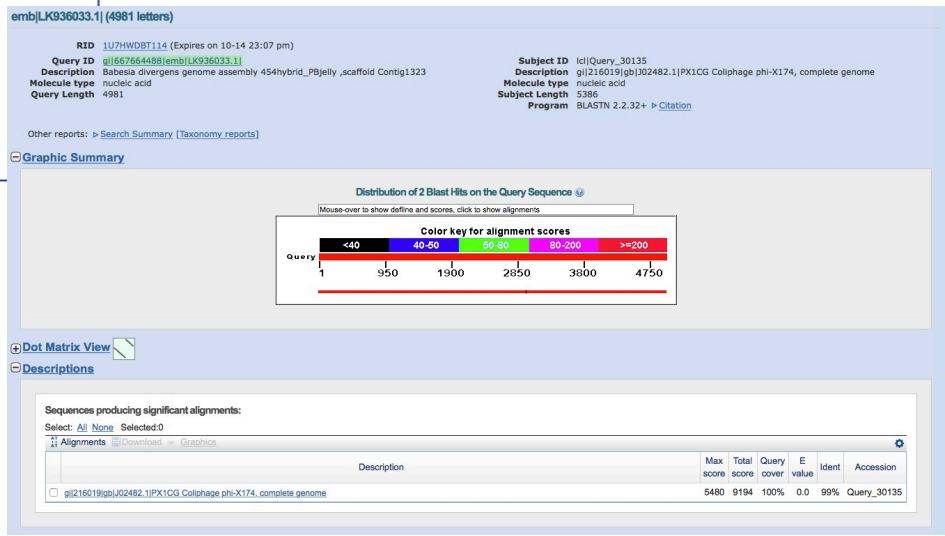
FASTA Graphics

Go to:

```
LOCUS       LK936033               4981 bp    DNA     linear   INV 07-OCT-2014
DEFINITION  Babesia divergens genome assembly 454hybrid_PBjelly, scaffold
            Contig1323.
ACCESSION   LK936033
VERSION     LK936033.1  GI:667664488
DBLINK      BioProject: PRJEB6536
            BioSample: SAMEA2612614
KEYWORDS    .
SOURCE      Babesia divergens
  ORGANISM  Babesia divergens
            Eukaryota; Alveolata; Apicomplexa; Aconoidasida; Piroplasmida;
            Babesiidae; Babesia.
REFERENCE   1
  AUTHORS   Montero,Estrella.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-2014) Spanish National Center for Microbiology,
            Carretera Majadahonda-Pozuelo, Km 2,2. 28220. Majadahonda. Madrid,
            Spain
FEATURES
     source
```



We found these reads mapping on a single contig of a ***Babesia divergence*** genome assembly, actually corresponding to the PhiX genome usually added to Illumina libraries. A similar contamination has been found (Mukherjee et al. 2015) in >1000 publicly available microbial genomes,



Other Bacterial contamination in Reference Genomes

Very often, in shotgun data analysis of human clinical sample we observe bizarre species intruders. For example, in the analysis of a human fecal sample we observed several reads assigned to the sea urchin ***Lichetinus variegatus***.

It is quite easy to check that it is, indeed, some bacterial contamination.

Contact Helpdesk

Sequence: JI425975.1

TSA: Lytechinus variegatus isotig12826.Lvarbgast mRNA sequence.

View: TEXT FASTA XML

Download: XML FASTA TEXT

Organism	Molecule type	Topology	Data class	Taxonomic Division
Lytechinus variegatus	mRNA	linear	TSA	INV

Show Version History

JI425975

RID M519EVH4014 (Expires on 06-16 21:01 pm)

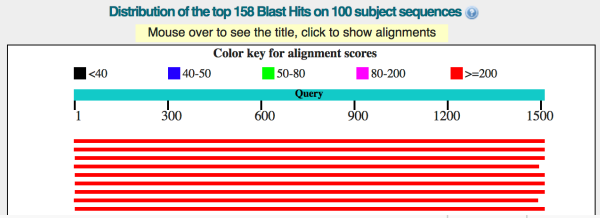
Query ID JI425975.1
Description TSA: Lytechinus variegatus isotig12826.Lvarbgast mRNA sequence
Molecule type rna
Query Length 1504

Database Name nr
Description Nucleotide collection (nt)
Program BLASTN 2.6.1+ ▹ Citation

Other reports: ▹ Search Summary [Taxonomy reports] [Distance tree of results] [MSA viewer]

Graphic Summary

oxopneustidae, Lytechinus



Sequences producing significant alignmer

Select: All None Selected:0

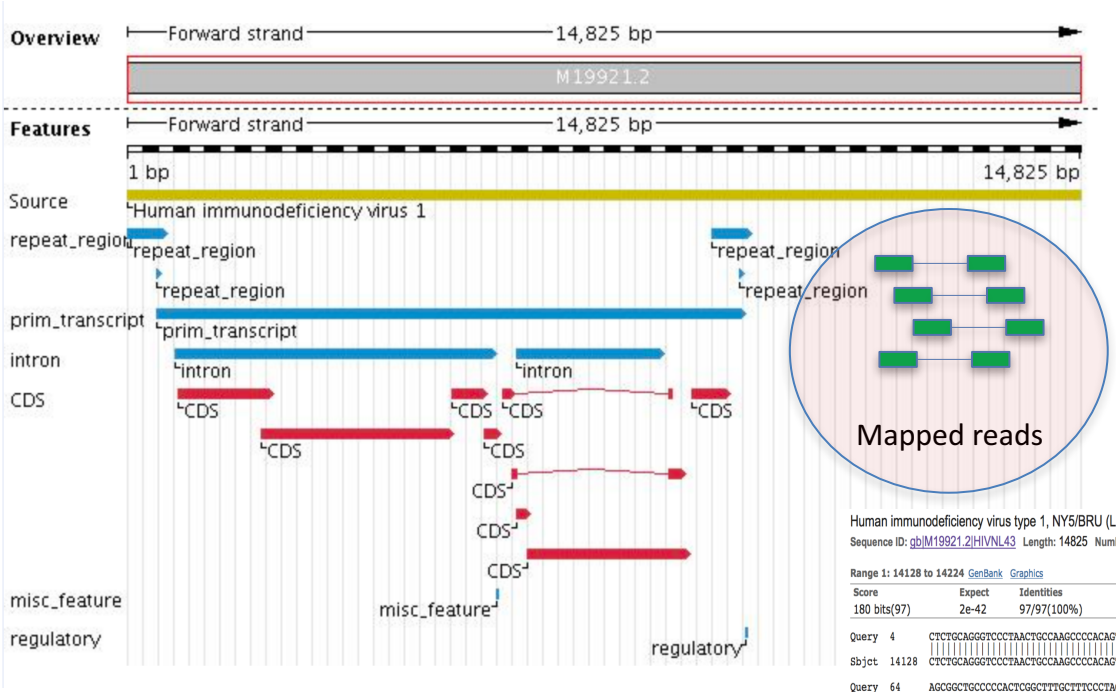
Alignments Download GenBank Gra

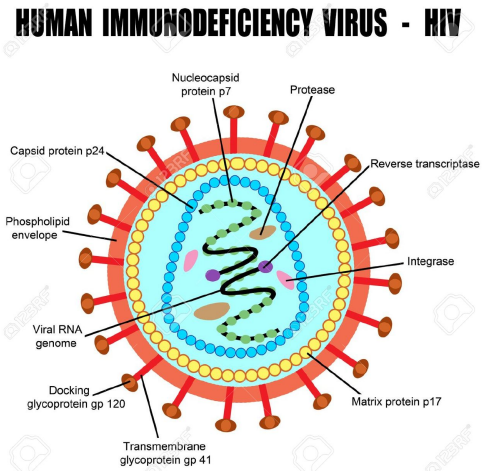
Description	score	score	uery cover	E value	Ident	Accession
Sphingomonas sanxanigenens strain NX02 23S ribosomal RNA gene, complete sequence	2392	2392	99%	0.0	95%	NR_122018.1
Sphingomonas wittichii strain RW1 23S ribosomal RNA gene, complete sequence	2381	2381	99%	0.0	95%	NR_076508.1
Sphingopyxis alaskensis strain RB2256 23S ribosomal RNA gene, complete sequence	2322	2322	99%	0.0	95%	NR_076422.1

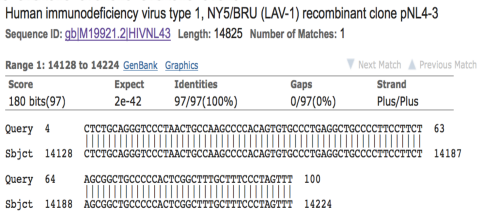
HIV INFECTION???

In the analysis of DNAseq shotgun data (500 M reads, PE 100x2 bp) from a cervical cancer carcinoma sample we detected several reads assigned to HIV. Is then the patient HIV-positive???

We realized that the relevant ENA matching entry (M19912) was indeed **chimeric** as its 3' portion mapped to the human genome. Any such information was present in the Feature Table.







TAXONOMY INCONSISTENCIES

The entry GAXI01005457 is associated to a *Chryseobacterium* in SILVA database, but to the giant springtail (*Tetrodontophora bielanensis*) in EMBL database.

silva
high quality ribosomal RNA databases

Home SILVAngs Browser Search Aligner Download Documentation Projects FISH & Probes Contact

Database SSU 128 Show Cart
Taxonomy SILVA

Cart: 1
Show
Clear
Download

SILVA ▸ Bacteria ▸ Bacteroidetes ▸ Flavobacteriia ▸ Flavobacteriales ▸ Flavobacteriaceae ▸ Chryseobacterium ▸ GAXI01005455

Flavobacteriales (0%)
(7) (13)
Blattabacteriaceae
Cryomorphaceae
Flavobacteriaceae (0%)
NS7 marine group
NS9 marine group
R103-B20
Schleiferiaceae

Flavobacteriaceae (0%)
(144) (336)
Actibacter
Aequorivita
Aestuariibaculum
Aestuariivivens
Algibacter
Algitalea
Antarcticimonas
Apibacter
Aquibacter
Aquimarina
Arenibacter
Arenitalea
Aurantiacicella

Chryseobacterium (0.02%)
(5566)
(1/112) **next**
- *Oryza sativa Indica Group (long-grained rice)*
- *Oryza sativa Indica Group (long-grained rice)*
- *metagenome*
- *Flavobacteriaceae bacterium 2*
- *Flavobacterium sp. B17*
- *Haloanella gallinarum*
- *Chryseobacterium proteolyticum*
- *uncultured Bacteroidetes bacterium*
- *uncultured Bacteroidetes bacterium*
- *uncultured bacterium*

GAXI01005455

Accession Nr	GAXI01005455
Description	TSA: Tetrodontophora bielanensis s5462_L_15759_0 transcribed RNA sequence.
Regions	1
Length	1233
Quality	
Sequence	
Alignment	
Pintail	

Taxonomy

SILVA	Bacteria ▸ Bacteroidetes ▸ Flavobacteriia ▸ Flavobacteriales ▸ Flavobacteriaceae ▸ Chryseobacterium ▸
SILVA Ref	Bacteria ▸ Bacteroidetes ▸ Flavobacteriia ▸ Flavobacteriales ▸ Flavobacteriaceae ▸ Chryseobacterium ▸
SILVA Ref NR	Bacteria ▸ Bacteroidetes ▸ Flavobacteriia ▸ Flavobacteriales ▸ Flavobacteriaceae ▸ Chryseobacterium ▸
LTP	Unclassified ▸
EMBL	Eukaryota ▸ Metazoa ▸ Ecdysozoa ▸ Arthropoda ▸ Hexapoda ▸ Collembola ▸ Collembola ▸ Poduromorpha ▸ Poduroidea ▸ Onychiuridae ▸ Tetrodontophorinae ▸ Tetrodontophora ▸
greengenes	Unclassified ▸
RDP	Unclassified ▸

Conclusions and Perspectives

1. Great room for improving computational tools both in term of accuracy and computational efficiency

2. Establish international standard benchmark datasets

3. Clean Reference nucleotide databases for both metabarcoding and shotgun analysis

4. Need for unified and accurate reference taxonomies

Acknowledgments

Bruno Fosso
Monica Santamaria
CNR-IBIOM, Bari

Marco Crescenzi & Co.
ISS, Roma

Gabriel Valiente & Co.
University of Catalunia, Barcelona