

Il Centro Tematico Biomolecolare - Strumenti e Servizi Molecolari -



Monica Santamaria & Bachir Balech

IBBE-CNR, Bari



Nel **Centro Tematico Biomolecolare (CTB)** competenze e facilities avanzate per analisi molecolari e bioinformatiche sono integrate al fine di fornire alla comunità scientifica servizi e consulenza per studi di Biodiversità Molecolare.



Le attività del CTB sono supportate dal **Laboratorio di Biodiversità Molecolare (MoBiLab)**, con le sue piattaforme oggi pienamente operative basate sulle più innovative tecnologie NGS e su potenti risorse per lo storage e l'analisi computazionale dei dati. Oltre al supporto a LifeWatch, i servizi ospitati dal MoBiLab contribuiscono alle attività del nodo italiano dell'infrastruttura europea **ELIXIR**.

Infrastruttura molecolare e bioinformatica avanzata volta a supportare l'analisi di dati genomici, metagenomici, filogenetici, filogeografici e di DNA barcoding in studi di Biodiversità Molecolare.

MoBiLab attualmente comprende:

Main Applications
Whole genome sequencing;
RNA sequencing;
Exome sequencing;
and more..



NextSeq 500

Output	Read Number	Read Length
120 Gb	400 M	2x150 bp

Main Applications
16S Metagenomics;
Targeted RNA;
Custom amplicon;
De novo sequencing;
and more..



MiSeq

Output	Read Number	Read Length
15 Gb	25 M	2x300 bp

1. **Protocolli Meta-barcoding and shotgun basati sulle tecnologie high-throughput sequencing (HTS);**
2. **Strumenti per l'analisi molecolare e bioinformatica di comunità** procariotiche ed eucariotiche, a vita libera o simbiotici.
3. **Strumenti per l'analisi molecolare e bioinformatica di genomi e della loro espressione** in procarioti ed eucarioti.
4. **Strumenti per l'accesso a dati di Biodiversità Molecolare** distribuiti su risorse World-Wide.

Strumenti & servizi bioinformatici per la Genomica e la Metagenomica



Multiallineamento

- **MSA-PAD**

Analisi di dati
Meta-barcoding

- **BioMaS**



Ricerca di
biosequenze e
metadati

- **DataRetrieval**

Analisi di dati di
Metagenomica Shotgun

- **MetaShot**

Banche dati
specializzate

- **ITSoneDB**

METAGENOMICA

un nuovo sguardo sul mondo invisibile

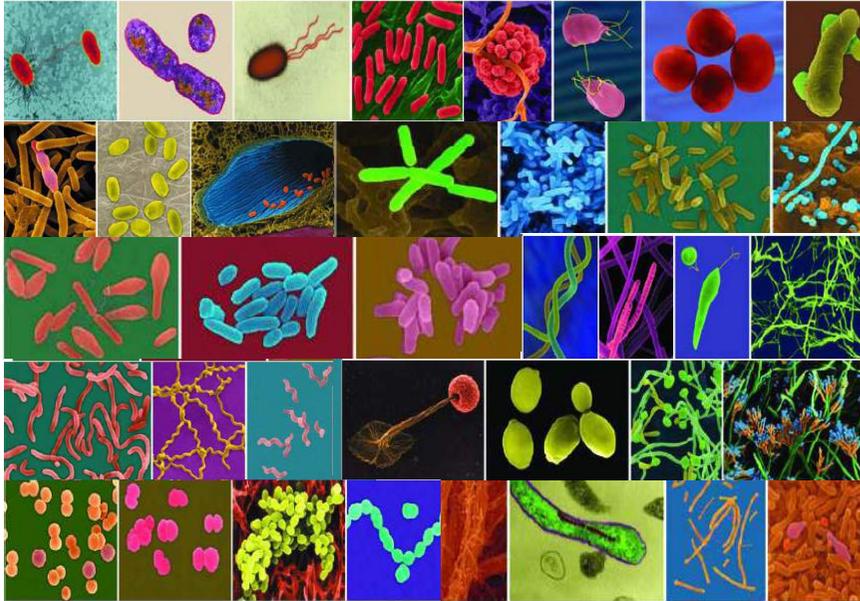
Ogni ambiente può essere considerato come un grande singolo organismo. Il suo patrimonio genetico, al quale contribuiscono tutti gli organismi residenti in esso, viene chiamato **Metagenoma**.



Un Genoma

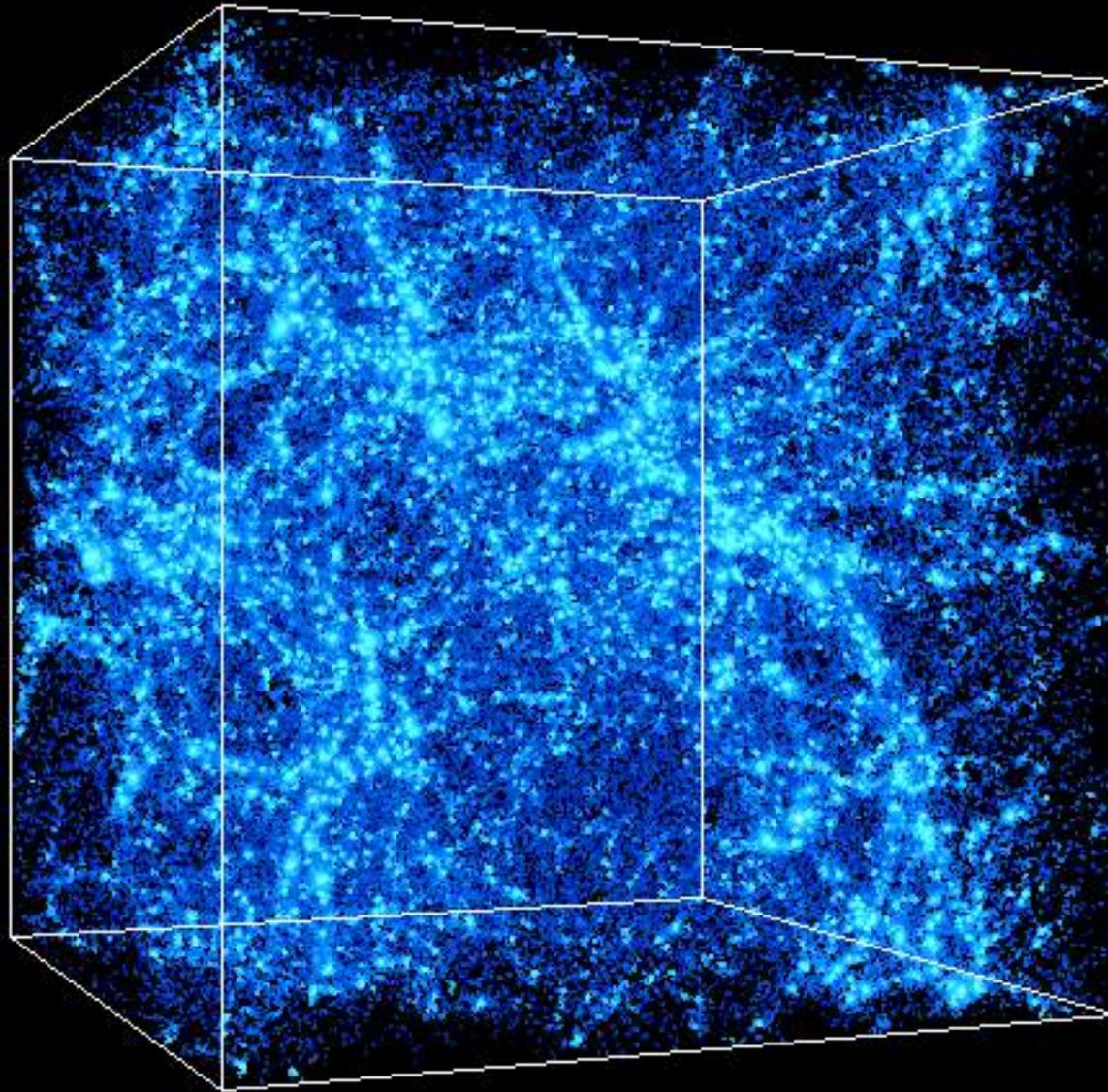


Un Metagenoma



...in particolare questo nuovo approccio è stato applicato con successo in molti studi pilota per rivelare la diversità genetica e tassonomica delle comunità microbiche, senza la necessità di isolare e mettere in coltura i singoli organismi.

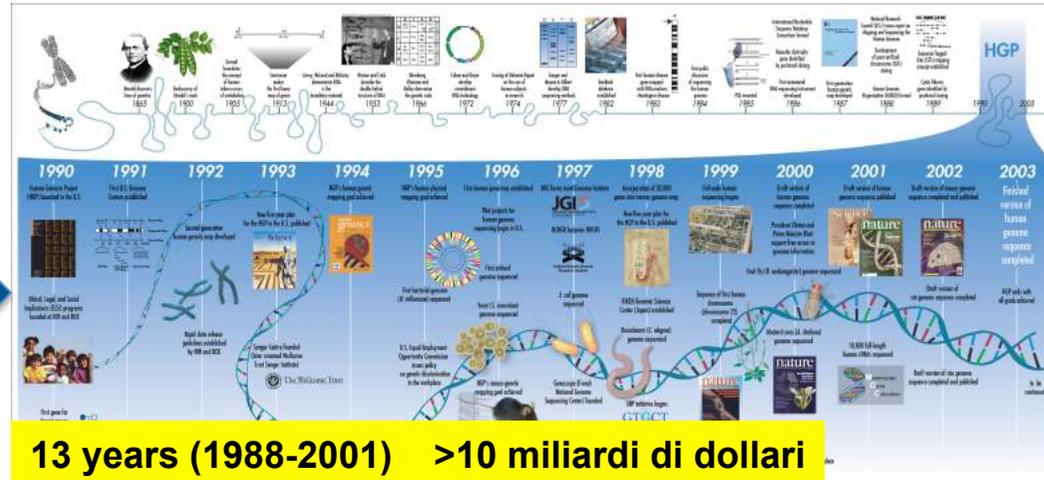
Poichè più del 99% degli organismi viventi in qualsiasi ambiente non può essere coltivato mediante tecniche standard, il metodo metagenomico, essendo coltura-indipendente, potrebbe determinare un ampliamento senza precedenti della biodiversità indagabile. Infatti gli studi metagenomici stanno diventando fortemente informativi riguardo alla diversità genetica, alla struttura delle popolazioni, alle relazioni e al ruolo ecologico di intere comunità microbiche.



I will shed
light on the
Dark
Matter !!



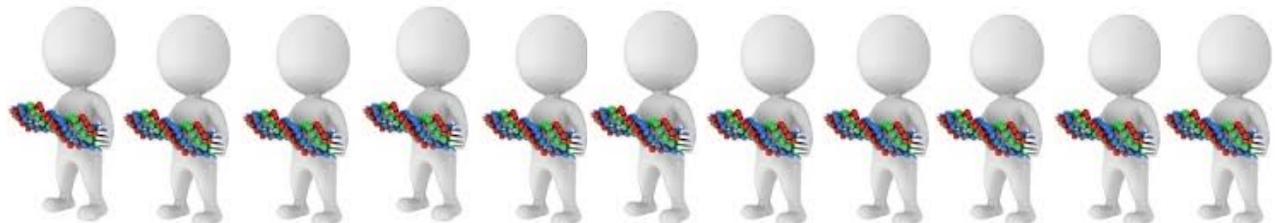
La straordinaria ascesa delle tecnologie di indagine genetica!



1953 – struttura del DNA



Oggi – pochi giorni < 1000 dollari

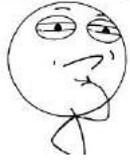




Bioinformatica di nuova generazione!



CHALLENGE CONSIDERED



Metagenomica: alcuni problemi pratici dell'analisi bioinformatica

- La disponibilità di risorse di calcolo adeguate per la veloce e globale analisi dei dataset metagenomici rappresenta ancora il più stretto collo di bottiglia.
- Filtering, denoising, clustering e assegnazione tassonomica dei dati metagenomici prodotti da HTS sono tutti passaggi fondamentali per una classificazione tassonomica profonda e corretta. Spesso alcuni di questi passaggi sono sottovalutati dai ricercatori meno esperti.
- Il considerevole eccesso di sequenze appartenenti all'ospite prodotto in un tipico esperimento HTS su comunità simbiotiche può rappresentare un notevole rumore di fondo, che deve essere rimosso, nell'identificazione bioinformatica della controparte microbica.
- Se una o più fasi precedenti sono trascurate possono essere prodotte inferenze errate.
- Collegare le diverse fasi di analisi bioinformatica può non essere banale: per esempio convertire il formato di output di uno step di analisi nel formato diverso opportuno per lo step successivo.
- Quando si considera l'enorme mole di dati HTS che devono fluire attraverso l'intero processo di analisi, la capacità computazionale diventa il fattore limitante più importante.
- Necessità assoluta di ricche e ben annotate risorse di riferimento

CHALLENGE ACCEPTED



Analisi di dati di Meta-barcoding

■ BioMaS



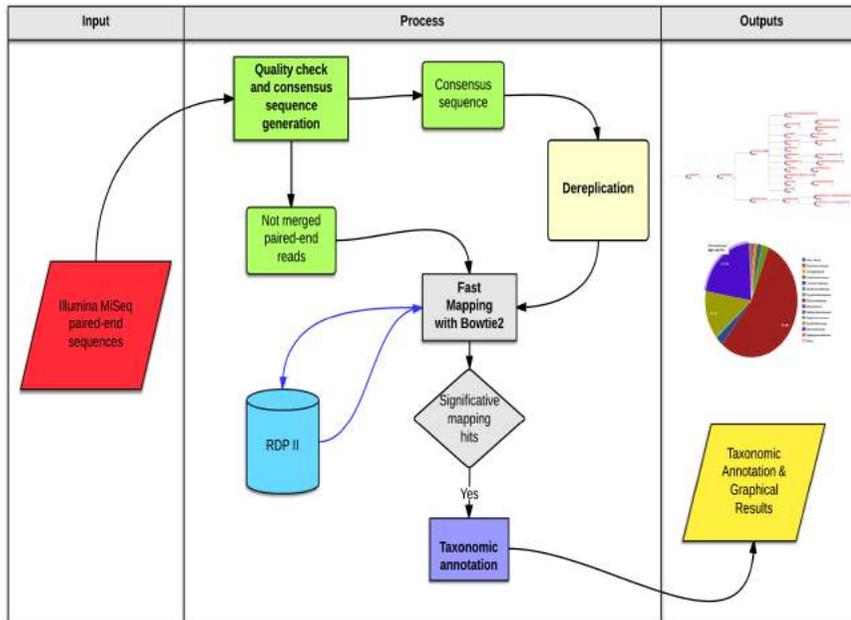
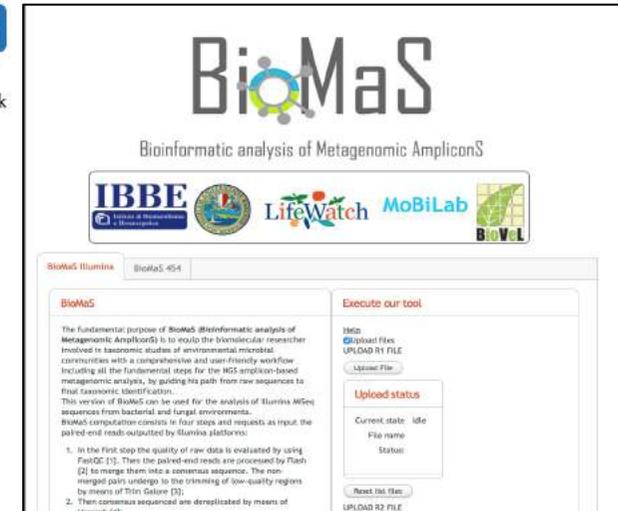
METHODOLOGY

Open Access

BioMaS: a modular pipeline for Bioinformatic analysis of Metagenomic Amplicons



Bruno Fosso¹, Monica Santamaria¹, Marinella Marzano¹, Daniel Alonso-Aleman², Gabriel Valiente², Giacinto Donvito³, Alfonso Monaco³, Pasquale Notarangelo³ and Graziano Pesole^{1,4,5*}



BioMaS è una pipeline volta alla rapida conversione dei dati grezzi di DNA Meta-barcoding prodotti dalle piattaforme HTS Illumina and Roche 454 in una dettagliata descrizione tassonomica di comunità microbiche ambientali. Sviluppi futuri comprendono l'estensione del range tassonomico indagabile ai Metazoi (marcatore barcode COI)

Disponibile come web service all'indirizzo <https://recasgateway.ba.infn.it/>

In collaborazione con:



Metagenomica Target-oriented (*Meta-barcoding*)

campionamento



Estrazione del metagenoma



Amplificazione del meta-barcode



HTS



Analisi bioinformatica

Il barcode ottimale

- ◆ Ubiquità nel range tassonomico considerato (es: tutti i batteri, tutti i funghi, etc.);
- ◆ Alto potere di discriminazione alivello di specie;
- ◆ Struttura con domini a diverso grado di variabilità (regioni ipervariabili fiancheggiate da regioni conservate);
- ◆ Dimensione compatibile con la lunghezza delle read HTS.

Identificazione di specie batteriche



una o alcune regioni ipervariabili del gene 16S

rRNA

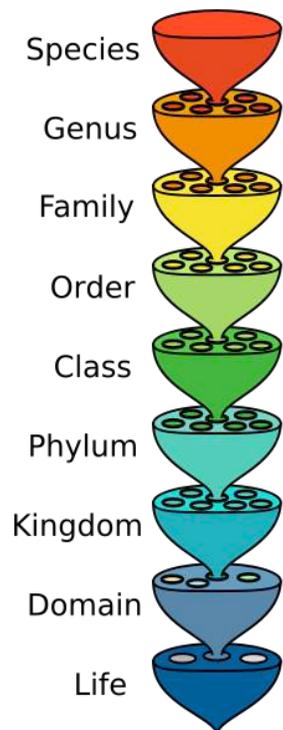
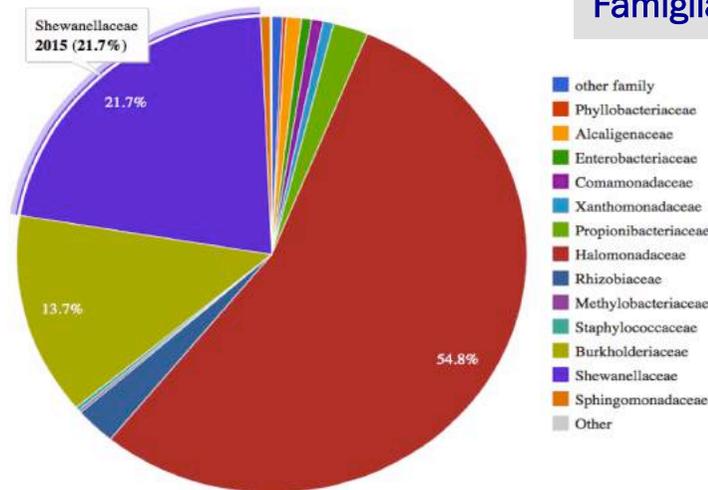
Un marker molto sensibile nei Funghi



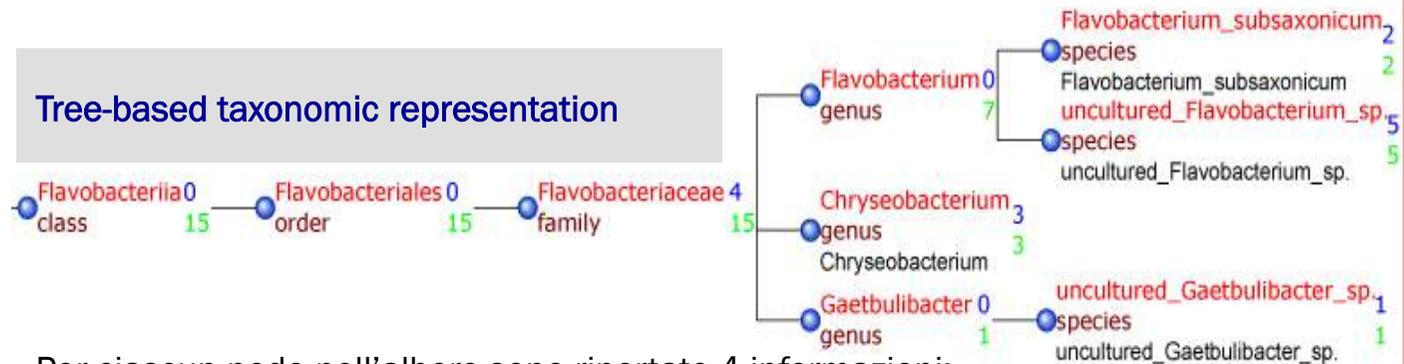
sequenze ITS

BioMaS report file

Pie chart:distribuzione tassonomica a livello di Famiglia



Tree-based taxonomic representation



Per ciascun nodo nell'albero sono riportate 4 informazioni:

- Nome scientifico (rosso)
- Livello tassonomico (marrone)
- Numero di reads direttamente assegnate al nodo (blu)
- Numero di reads assegnate al nodo e ai suoi discendenti (verde)

BioMaS web services

http://testjst.ba.infn.it/openaccess/BIOMAS_HELP/BiomassPostProcessingTools.zip

Nucleic Acids Research Advance Access published May 29, 2012

Nucleic Acids Research, 2012, 1–8
doi:10.1093/nar/gks497

METAGENassist: a comprehensive web server for comparative metagenomics

David Arndt¹, Jianguo Xia², Yifeng Liu¹, You Zhou², An Chi Guo¹, Joseph A. Cruz¹, Igor Sinelnikov¹, Karen Budwill³, Camilla L. Nesbø^{2,4} and David S. Wishart^{1,2,5,*}

¹Department of Computing Science, ²Department of Biological Sciences, University of Alberta, Edmonton, ³Environment & Carbon Management Division, Alberta Innovates—Technology Futures, Edmonton, Alberta, Canada, ⁴CEES, Department of Biology, University of Oslo, Oslo, Norway and ⁵National Research Council, National Institute for Nanotechnology (NINT), Edmonton, Alberta, Canada T6G 2E8

Received March 6, 2012; Revised May 2, 2012; Accepted May 3, 2012

METAGENassist
statistical tools for comparative metagenomics

Welcome [click here to start](#)

News & Updates

- Improvements to dendrogram and heatmap for viewing many samples/features. (08/13/13) **NEW**
- Added cross-validation, variable importance, and permutation test to PLS-DA. Improved stability of PLS-T (08/13/13) **NEW**

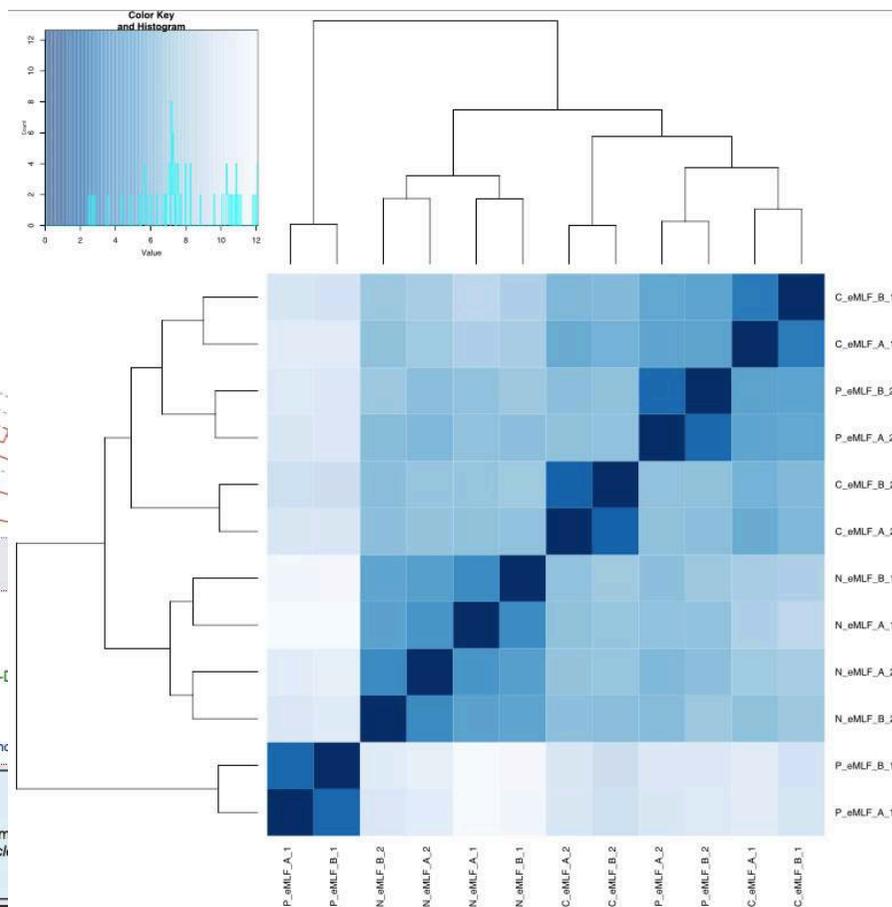
[Read more](#)

Please Cite:

David Arndt, Jianguo Xia, Yifeng Liu, You Zhou, An Chi Guo, Joseph A. Cruz, Igor Sinelnikov, Karen Budwill, Carr Nesbø and David S. Wishart [METAGENassist: A comprehensive web server for comparative metagenomics](#) *Nucleic Acids Research* 2012 Jul;40(Web Server issue):W88-95. Epub 2012 May 29.

Project objective: To provide a user-friendly, web-based analytical pipeline for comparative metagenomic studies. In particular, METAGENassist allows users to take bacterial census data from different environment sites or different biological hosts, and perform comprehensive multivariate statistical analyses on the data. These multivariate analyses can be done using either taxonomic or automatically generated phenotypic labels and visualized using a variety of high quality graphical tools. The bacterial census data can be derived from 16S rRNA data, NextGen shotgun sequencing or even classical microbial culturing techniques.

Data input formats: Users upload a taxonomic profile in one of several supported formats ([details](#)). A taxonomic profile file contains the names of microbial species (or other taxonomic classes) and their relative abundance in at least 2 samples; accepted formats include CSV format, or data generated by *mothur*, *QIIME*, *MG-RAST*, *MEGAN*, and *STAMP*. Users will also benefit by uploading an optional file containing metadata for each sample (CSV format).



Applicazione di BioMaS

BioMas è stato applicato all'analisi di diversi ambienti utilizzando vari marcatori batterici.

<i>Comunità microbica</i>	<i>Marker</i>
Microbioma della filiera vinicola	V5-V6 16S rDNA
Microbioma sedimenti polari	V1-V2 16S rDNA
Microbioma terreno di coltura Cardoncello	V5-V6 16S rDNA
Microbioma intestinale	V1-V3 16S rDNA V5-V6 16S rDNA
Microbioma del respiro	V5-V6 16S rDNA
Comunità di nematodi	V4 18S rDNA
Microbioma di ambienti acquatici	V5-V6 16S rDNA V3-V4 16S rDNA

Banche dati specializzate:

■ ITSoneDB



BRIEFINGS IN BIOINFORMATICS, VOL 13, NO 6, 682–695
Advance Access published on 10 July 2012

doi:10.1093/bib/bbs036

Reference databases for taxonomic assignment in metagenomics

Monica Santamaria*, Bruno Fosso*, Arianna Consiglio, Giorgio De Caro, Giorgio Grillo, Flavio Licciulli, Sabino Liuni, Marinella Marzano, Daniel Alonso-Aleman, Gabriel Valiente and Graziano Pesole

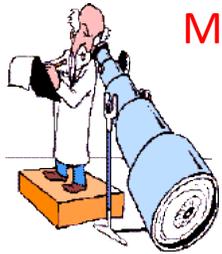
Submitted: 14th March 2012; Received (in revised form): 2nd June 2012

The screenshot shows the ITSoneDB website interface. At the top, there is a navigation bar with buttons for HOME, SIMPLE SEARCH, TREE SEARCH, and ADVANCED SEARCH. Below this is a header for the 'Fungal Ribosomal Internal Transcribed Spacer 1 Database'. A descriptive paragraph explains that ITSoneDB is a comprehensive collection of fungal ribosomal RNA ITS1 sequences. Below the text is a table titled 'ITSoneDB Content' with five rows of statistics. At the bottom, there is a footer with the text 'CNR Istituto di Tecnologie Biomediche, Bari - CNR Istituto di Biomembrane e Bioenergetica, Bari - Università degli studi di Bari'.

ITSoneDB Content	
Number of the nucleotide sequences	491,919
Total number of species (according to the NCBI taxonomy)	40,855
Number of ITS1 nucleotide sequences with start and end positions inferred by both GB information and HMM profiles	34,316
Number of ITS1 nucleotide sequences with start and end position inferred only by HMM profiles mapping	101,253
Number of ITS1 nucleotide sequences with start and end position inferred only by GB information	356,350

ITSoneDB è una collezione curata di sequenze ITS1 volta allo studio delle comunità microbiche fungine. Sviluppi futuri comprendono aggiornamento e ampliamento ad altre specie microbiche eucariotiche.

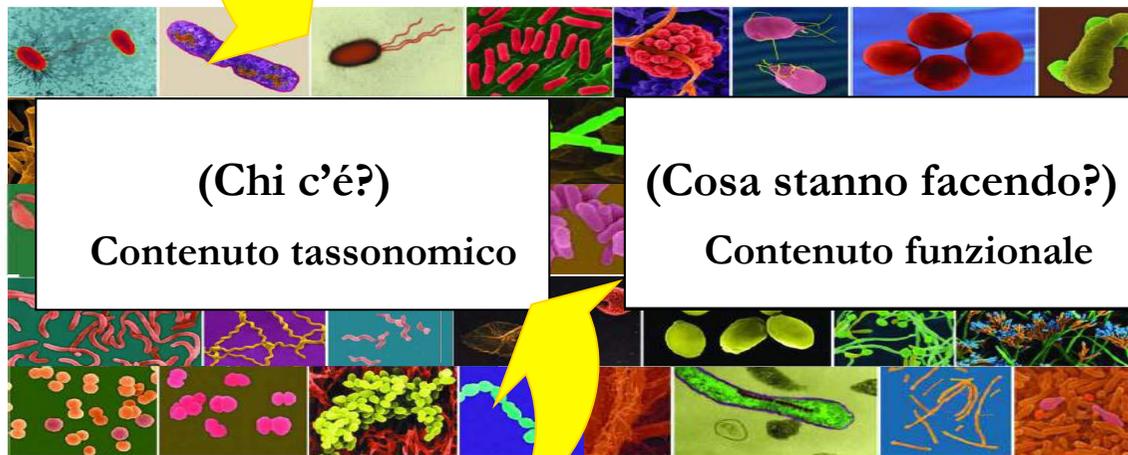
Già disponibile come web service all'indirizzo <http://itsonedb.ba.itb.cnr.it:8080/ITS1>



METAGENOMICA

Random shotgun NGS

- Identifica specie, geni e attività funzionali;
- non permette sufficiente copertura per rilevare le specie rare;
- molto più costoso in termini di sequenziamento ed analisi computazionale.



Barcode-based

- Alta sensibilità nella discriminazione ed identificazione di specie;
- Meno costoso in termini di sequenziamento e analisi computazionale;
- Condizioni universali di PCR;
- Database di riferimento specializzati (ad esempio RDP for 16S, ITSoneDB for ITS1)
- può essere influenzato dalla diversa efficienza di amplificazione del marker in diverse specie;
- Nessuna informazione funzionale.

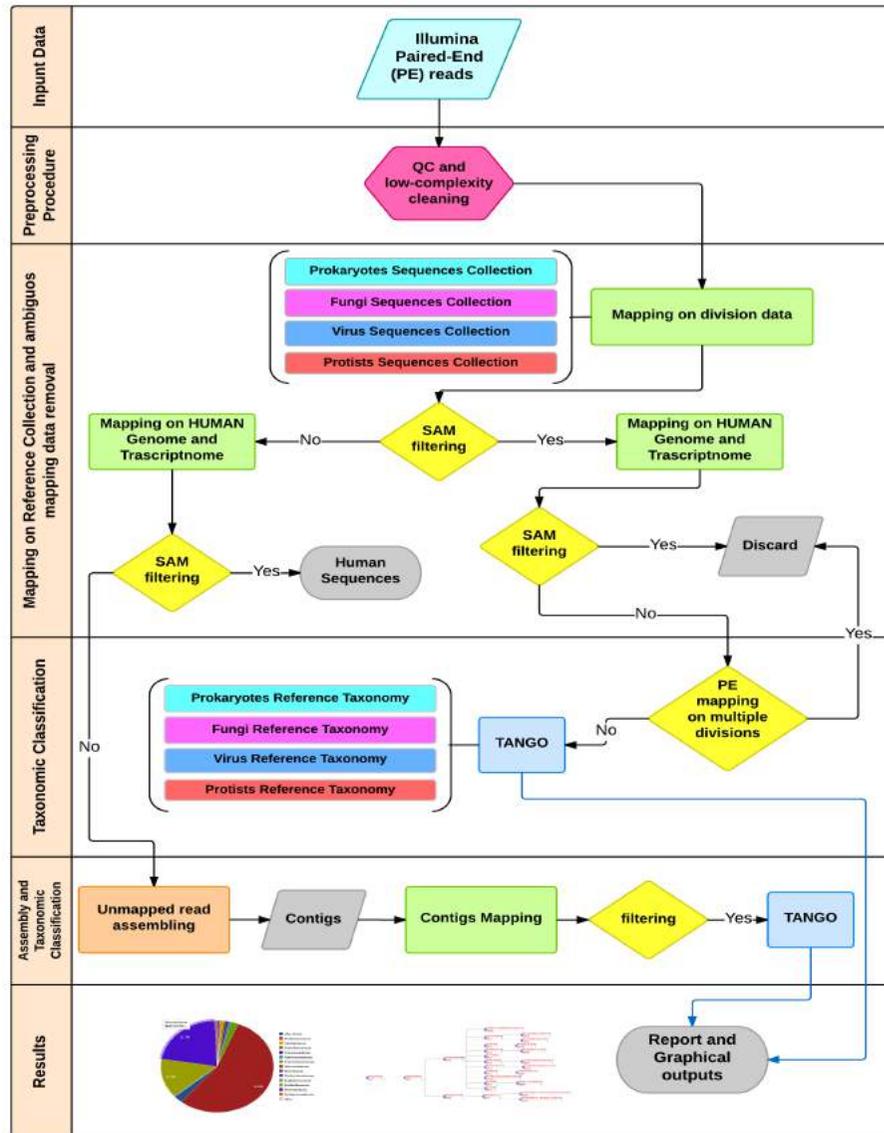
Analisi di dati di Metagenomica Shotgun:

■ MetaShot

MetaShot è una pipeline per la caratterizzazione tassonomica di dati HTS di Metagenomica shotgun orientata in particolare allo studio del microbioma dell'uomo o di altri ospiti (Fosso et al., in preparation).



Funziona al momento come sistema stand alone



Alcune pubblicazioni su riviste internazionali...

Biol Invasions (2015) 17:923–940
DOI 10.1007/s10530-014-0810-2

MOLECULAR TOOLS

The influence of invasive jellyfish blooms on the aquatic microbiome in a coastal lagoon (Varano, SE Italy) detected by an Illumina-based deep sequencing strategy

Caterina Manzari · Bruno Fosso · Marinella Marzano · Anita Annese · Rosa Caprioli · Anna Maria D'Erchia · Carmela Gissi · Marianna Intranuovo · Ernesto Picardi · Monica Santamaria · Simonetta Scorrano · Giuseppe Sgararella · Loredana Stabili · Stefano Piraino · Graziano Pesole

Received: 14 November 2013 / Accepted: 31 October 2014 / Published online: 19 November 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Bioinformatics, 31(15), 2015, 2571–2573
doi: 10.1093/bioinformatics/btv141
Advance Access Publication Date: 26 March 2015
Applications Note

OXFORD

Sequence analysis

MSA-PAD: DNA multiple sequence alignment framework based on PFAM accessed domain information

Bachir Balech¹, Saverio Vicario², Giacinto Donvito³, Alfonso Monaco³, Pasquale Notarangelo³ and Graziano Pesole^{1,4,*}

¹Istituto di Biomembrane e Bioenergetica and ²Istituto Tecnologie Biomediche, Consiglio Nazionale delle Ricerche, 70126 Bari, Italy, ³Istituto Nazionale di Fisica Nucleare, 70126 Bari, Italy and ⁴Dipartimento Bioscienze, Biotecnologie e Biofarmaceutica, University of Bari, 70126 Bari, Italy

*To whom correspondence should be addressed.
Associate Editor: David Posada

Received on August 8, 2014; revised on March 6, 2015; accepted on March 9, 2015



BRIEFINGS IN BIOINFORMATICS, VOL 13, NO 6, 682–695
Advance Access published on 10 July 2012

doi:10.1093/bib/bbs062

Reference databases for taxonomic assignment in metagenomics

Monica Santamaria*, Bruno Fosso*, Arianna Consiglio, Giorgio De Caro, Giorgio Grillo, Flavio Licciulli, Sabino Liuni, Marinella Marzano, Daniel Alonso-Aleman, Gabriel Valiente and Graziano Pesole

F
D

Submitted: 14th March 2012; Received (in revised form): 2nd June 2012

CS

METHODOLOGY

Open Access

BioMaS: a modular pipeline for Bioinformatic analysis of Metagenomic AmpliconS



Bruno Fosso¹, Monica Santamaria¹, Marinella Marzano¹, Daniel Alonso-Aleman², Gabriel Valiente², Giacinto Donvito³, Alfonso Monaco³, Pasquale Notarangelo³ and Graziano Pesole^{1,4,5*}



ANNOUNCEMENT

Draft genome sequence of *Sphingobium* sp. strain ba1, resistant to kanamycin and nickel ions

Caterina Manzari¹, Matteo Chiara², Alessandra Costanza¹, Claudia Leoni¹, Mariateresa Volpicella¹, Ernesto Picardi^{1,3,4}, Anna Maria D'Erchia¹, Antonio Placido³, Massimo Trotta⁵, David S. Horner², Graziano Pesole^{1,3,4,6} & Luigi R. Ceci³



**CNR-IBBE e Università di
Bari**

Prof. Graziano Pesole

Dott. Bruno fosso

INFN

Prof. Giorgio Pietro Maggi

Dott. Giacinto Donvito

Dott. Pasquale Notarangelo

Dott. Alfonso Monaco

**Universitat Politècnica de
Catalunya**

Prof. Gabriel Valiente

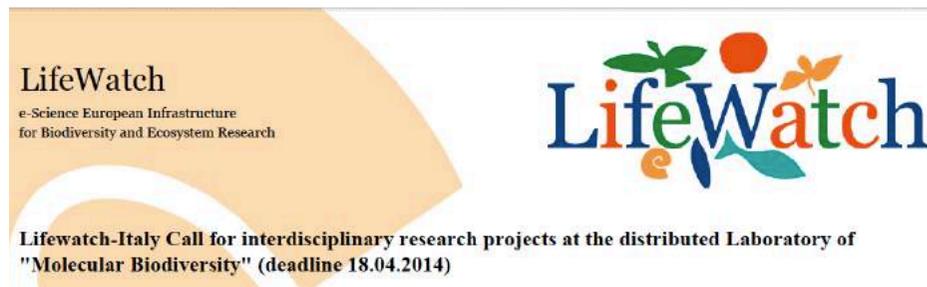
Dr. Daniel Alonso Alemany

CINECA, Roma

Dott. Mattia D'antonio

CTB CALL

Per l'investigazione di specie aliene o invasive attraverso la produzione e l'analisi di dati HTS



AFFILIATION	PI	RESEARCH TITLE	APPLICATIONS
SZN, Napoli	Casotti	Isolation and genetic characterization of bacterial strains isolated in the presence of polyunsaturated aldehydes from sea water samples of diatoms bloom.	Genomics
CNR-ISMAR, Venezia	Quero	MALWARe: MetAgenomic study of microbial invasive species introduced by Ballast Water.	Metagenomics
University of Milan Bicocca	Casiraghi	Design of an innovative pipeline for DNA metabarcoding of eukaryotes to monitor potential invasive species in atmospheric particulate during Milan EXPO 2015.	Metagenomics
SZN, Napoli	Ferrante	An investigation of the genetic basis of toxin production in the diatom species <i>Pseudo-nitzschia</i> .	Genomics
University of Salento	Basset	Biotic Interactions in Detritus food Webs: metagenomics analysis of the microfungi-detritivores and detritivore-detritivore interactions in lagoon ecosystems (BID_Webs).	Metagenomics
CNR-ISE, Verbania	Eckert	Daphnia as Vectors for Potential Pathogens and Antibiotic Resistant Bacteria in Freshwater Ecosystems.	Metagenomics
CNR-ISE, Verbania	Fontaneto	MEIOfauna DIVERsity (MEIODIVE): Testing ecogenomics of meiofauna as a tool for biological monitoring of marine sediments.	Metagenomics
University of Trieste	Pallavicini	Detection of alien species from gut contents of commercially important fish in the Adriatic Sea, possible effect on feeding behaviour and food webs.	Metagenomics
CNR-ISMAR, Venezia	Bongiorni	Responses of benthic microbial diversity and functioning to the recent invasion of <i>Gracilaria vermiculophylla</i> in the lagoon of Venice.	Metagenomics
University of Milan	Gissi	Comparing fresh with formalin-preserved samples: application of the NGS technology to the molecular characterization of old museum specimens	Genomics

Dieci progetti sono stati finanziati e sono attualmente in progress

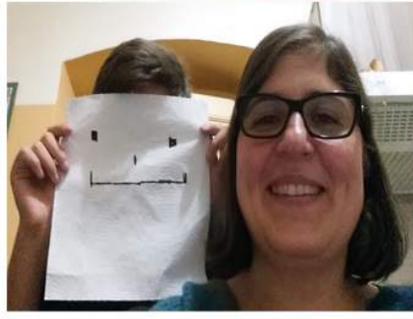




Staff



Personnel & contact details



Collaborations

