

LifeWatch

e-Science European Infrastructure
for Biodiversity and Ecosystem Research



Piano Strategico 2016-2017

CENTRO TEMATICO BIOMOLECOLARE



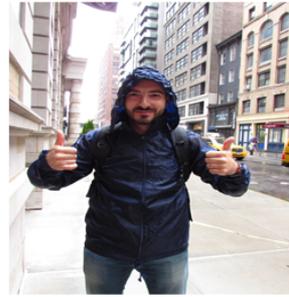
Adriana Zingone & Graziano Pesole

Presentazione Centri Tematici &
Piano delle attività 2016/2017
15-17 Febbraio 2016 ROMA





**UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO**



Presentazione Centri Tematici &
Piano delle attività 2016/2017
15-17 Febbraio 2016 ROMA





In the **Biomolecular Thematic Centre (CTB)**, skills and advanced facilities for molecular and bioinformatics analyses are integrated to provide the scientific community with **services** and **counselling** for Molecular Biodiversity studies.



Main Applications Whole genome sequencing; RNA sequencing; Exome sequencing; and more..	 NextSeq 500	
Output 120 Gb	Read Number 400 M	Read Length 2x150 bp
Main Applications 16S Metagenomics; Targeted RNA; Custom amplicon; De novo sequencing, and more..	 MiSeq	
Output 15 Gb	Read Number 25 M	Read Length 2x300 bp

The activities of the Thematic Centre are supported by Molecular Biodiversity Laboratory (**MoBiLab**), with its fully operative platforms based on the most innovative NGS technologies and powerful resources of data storage and computational analysis. In addition to their support to LifeWatch, the services hosted by MoBiLab contribute to the Italian node of the European infrastructure **ELIXIR**.





1. Costruzione delle componenti dell'infrastruttura di LifeWatch-ITA e del Service Centre

1.a) Sviluppo Thesauri. Lo sviluppo Thesauri è in progress per integrare la componente molecolare della biodiversità relativa a dati di Genomica, Metagenomica e DNA-barcoding, validi per qualunque specie vivente o comunità.

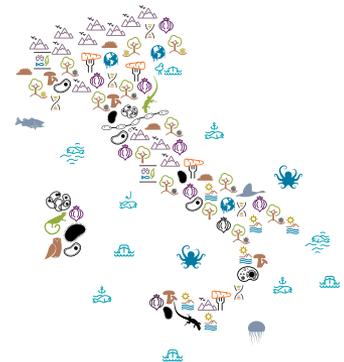
1.b) Sviluppo ontologie. Sulla base del punto 1.a completare la formalizzazione di un'ontologia dei tratti molecolari da utilizzare nello sviluppo del sistema di query avanzato "DataRetrieval" che permetta l'estrazione e l'integrazione di dati da risorse di sequenze nucleotidiche disponibili in rete.





1c. Popolamento del Catalogo Risorse Dati

- **ITSoneDB**: collezione completa di sequenze dell'ITS1 (Internal Transcribed Spacer 1) del cluster codificante per gli RNA ribosomali volta a supportare analisi metagenomiche di comunità (già disponibile come web service: <http://itsonedb.ba.itb.cnr.it>), con ulteriore estensione a tutte le specie microbiche eucariotiche.
- **HMaDB**: collezione di sequenze corrispondenti al 18S rRNA di microalghe tossiche e specie congeneriche non tossiche, manualmente curata e arricchita di metadati, e corredata da una valutazione del potere risolutivo di frammenti V4 e V9 nella identificazione delle specie di interesse. Popolamento del catalogo strumenti/servizi.





1d. Popolamento del catalogo strumenti/servizi:

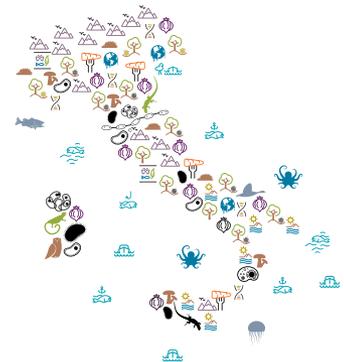
- **BioMaS:** pipeline per l'analisi bioinformatica degli Ampliconi Metagenomici (già disponibile come web service interfacciato con i database di riferimento RDP, GreenGenes, Silva (eucarioti) e PR2/HMaDB): <https://recasgateway.ba.infn.it/> , [http:// galaxy.cloud.ba.infn.it:8080](http://galaxy.cloud.ba.infn.it:8080), ed eventualmente espandibile includendo collezioni di riferimento relative ad altri marcatori di specie eucariotiche, tra i quali COX1.
- **MetaShot:** pipeline per l'analisi bioinformatica di sequenze metagenomiche shotgun (in fase di sviluppo, funziona al momento come sistema standalone)
- **MSA-PAD:** pipeline per il multiallineamento di sequenze nucleotidiche sulla base di domini PFAM presenti nelle regioni codificanti (già disponibile come web service, al momento funzionante solo sulla risorsa BOLD: <https://recasgateway.ba.infn.it/>; <http://www.myexperiment.org/workflows/4549.html>; <http://www.myexperiment.org/workflows/4551.html>).
- **DataRetrieval:** Sistema di interrogazione per l'estrazione da banche dati di sequenze di DNA barcode e dei metadati ad esse associati (già disponibile come web service su farm INFN – manca interfaccia grafica, funzionante su qualsiasi tipo di organismo, alieno o no, al momento estrae soltanto da BOLD)





2a. Rafforzamento networking con la comunità scientifica internazionale: possibili temi ICT chiave

- Collegamento tra l'infrastruttura ICT LW e la risorsa di calcolo Recas (INFN) dove risiedono gli strumenti di analisi bioinformatica (citati sopra) e dove sarà implementato uno spazio di storage temporaneo dei dati intermedi e finali dall'analisi bioinformatica (in collaborazione con INFN e Centro Servizi).
- Sviluppo presso il portale LW del sistema di autenticazione per l'accesso alle risorse di dati biomolecolari e agli strumenti di analisi (in collaborazione con INFN e Centro Servizi).
- Integrazione ed esposizione del sistema di query "DataRetrieval" sul portale LW (in collaborazione con INFN e Centro Servizi)





2b. Rafforzamento networking con la comunità scientifica internazionale: componenti ICT e collegamenti internazionali dei casi di studio / attività già avviati/e

- Collegamenti con le principali iniziative internazionali, particolarmente quelle in fase di sviluppo nell'ambito dell'infrastruttura ELIXIR, attive nel campo di sviluppo di **Reference Database** per indagini molecolari della composizione microbica.
- Sviluppo di componenti ICT, nell'ambito di progettualità europee già in corso quali INDIGO, EXCELERATE e EMBRIC, per la portabilità in ambiente "cloud" di risorse e strumenti per analisi biomolecolari.



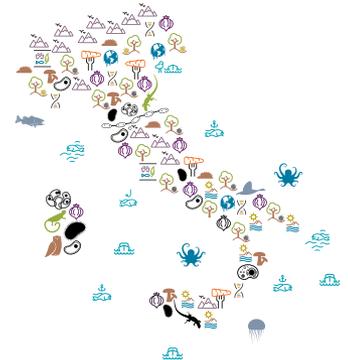
2c. Rafforzamento networking con la comunità scientifica internazionale: Temi scientifici prioritari

L'ambiente di ricerca virtuale sviluppato all'interno del Centro Tematico Biomolecolare e del relativo Laboratorio di Biodiversità Molecolare (MoBiLab) intende contribuire a rendere accessibili, in maniera rapida, comprensiva e accurata, l'enorme mole di dati molecolari giornalmente prodotti a supporto della ricerca sulla Biodiversità, offrendo anche la possibilità di analizzarli attraverso i più innovativi e potenti workflow computazionali. A tal fine prioritario è lo sviluppo di database e strumenti di analisi per la caratterizzazione molecolare di specie e popolazioni e della loro dinamica spazio-temporale.



2d. Rafforzamento networking con la comunità scientifica internazionale: Networking

- Partecipazione all' EMBO /EMBL Symposium: A New Age of Discovery for Aquatic Microeukaryotes, Heidelberg, 26-29/01/16
- Partecipazione al 1st Workshop of Globis-B a Lipsia, 29 febbraio – 2 marzo 2016
- Partecipazione alle sessioni speciali "Alien species ecological impacts: from genomics to macroecology" e "High-throughput sequencing in Limnology" del Congresso organizzato dalla International Society of Limnology (SIL), Torino 31 luglio – 5 agosto 2016
- Partecipazione al 17th International Conference on Harmful Algae, Florianopolis (Brasile), 09–14 October 2016, per diffondere l'uso dei tools messi a punto sulla ricerca delle alghe tossiche con approcci di meta-barcoding.





3. Collegamento con altri progetti / iniziative

GLOBIS-B. Il progetto GLOBIS-B “GLOBal Infrastructures for Supporting Biodiversity research”, Horizon 2020, si propone essenzialmente di riunire gli scienziati con operatori delle infrastrutture di ricerca globali ed esperti di interoperabilità legale per definire i requisiti di ricerca e i servizi infrastrutturali per il calcolo delle Variabili di Biodiversità Essenziali (EBVs).

OSD. Con il supporto del MoBiLab, il CTV Biomolecolare partecipa all’OSD fornendo expertise sul sequenziamento HTS della componente eucariotica planctonica dei siti marini di interesse.

EXCELERATE. Il progetto EXCELERATE (Fast-track ELIXIR implementation and drive early user exploitation across the life sciences) ha l’obiettivo di supportare l’implementazione dell’infrastruttura ELIXIR e comprende attività nell’ambito del WP sulla Metagenomica marina volte alla ottimizzazione di risorse e strumenti per lo studio della Biodiversità marina. Risorse proposte per l’analisi: BioMaS, ITSoneDB e MetaShot

EMBRIC. La collaborazione nell’ambito del progetto EMBRIC (European Marine Biological Research Infrastructure Cluster to promote the Blue Bioeconomy), coordinato dall’infrastruttura di ricerca EMBRIC, comprende lo sviluppo di strumenti per l’analisi di dati metagenomici.

INDIGO. La collaborazione nell’ambito del progetto INDIGO (INtegrating Distributed data Infrastructures for Global ExpLOitation), coordinato da INFN, comprende lo studio di fattibilità per la portabilità in ambiente “cloud” di risorse e strumenti di analisi per la Metagenomica.





4. Sviluppo servizi specifici attraverso casi di studio in collegamento a progetti attivi o in fase di presentazione

Prossima **CTB call** per progetti basati su produzione ed analisi di dati NGS da effettuare con il supporto delle infrastrutture/ competenze del MoBiLaB.





5. Rafforzamento dello staff di LifeWatch-ITA con cofinanziamento da parte di progettualità spin-off

Lo staff attualmente afferente al CTB, oltre ai coordinatori e **Monica Santamaria**, comprende 3 assegnisti di ricerca: **Caterina Manzari** (scad. 21.1.2017), **Marinella Marzano** (scad. 31.8.2016), **Maria Paola Tomasino** (scad. 2.9.2016, presso SZN di Napoli). Sono inoltre direttamente coinvolte nelle attività LW anche altre unità di personale a valere sulle progettualità spin-off sopra indicate: **Bachir Balech, Bruno Fosso, Marco Tangaro e Roberta Piredda**.



Priorità



17/02/2016



BIOMOLECULAR

10.00- 10.30 | Il Piano strategico del CT Biomolecolare: obiettivi e necessità – Graziano Pesole & Adriana Zingone
10.30-11.00 | Strumenti e servizi molecolari – Monica Santamaria
11.00-11.30 | Ricerca di alghe tossiche con approcci meta-DNA barcoding-Maria Paola Tomasino e Adriana Zingone
11.30-12.00 | Thesauri attinenti a dati di genomica e DNA-barcoding - Bachir Balech
12.00-12.30 | Discussione



Presentazione Centri Tematici &
Piano delle attività 2016/2017
15-17 Febbraio 2016 ROMA



