PROGETTO DI RICERCA

Titolo: Metodi computazionali e modelli per analisi di dati metagenomici ambientali e biomedici (Progetto TRIGGER)

Title: Computational methods and modelling for the analysis of environmental and biomedical metagenomic data (TRIGGER Project)

**ITA**

Moderne tecniche sperimentali (shotgun next-generation sequencing) permettono di raccogliere informazioni relative alla popolazione (eucariotica, batterica, virale, fungina) all’interno di specifici ecosistemi, sia ambientali (es acque reflue, aerosol, terreni) che biologici (microbiota intestinale e respiratorio). Questi dati possono fornire informazioni preziose per la descrizione di patologie e la loro evoluzione, oppure per identificare marker diagnostici o indicatori di variazioni nell’ambiente, ad esempio dovute ad eventi anomali o estremi legati al cambiamento climatico.

Per analizzare questo tipo di dati è necessaria 1) una competenza in ambito di programmazione, per poter gestire complesse pipeline di calcolo ad alte prestazioni (es. calcolo parallelo, calcolo su GPU per algoritmi di Intelligenza Artificiale); 2) competenze in ambito biostatistico, per potere elaborare i dati e svolgere analisi descrittive, di stratificazione supervisionata o non supervisionata dei dati, riduzione di dimensionalitá; 3) competenze teoriche, sia relative alla tipologia dei dati utilizzati (dati composizionali, per i quali servono metriche ed elaborazioni ad hoc) che alla loro integrazione (ad esempio mediante la teoria dei network per identificare comunitá di organismi e la loro evoluzione nel tempo).

Il progetto TRIGGER si propone di campionare soggetti ed ambienti con tecniche di genomica e metagenomica con lo scopo di comprendere meglio gli effeti del cambiamento climatico sulla salute: aumento del rischio per soggetti con patologici o caratteristiche genetiche specifiche, variazione della popolazione di patogeni in funzione di fattori ambientali quali inquinamento, temperatura e uminiditá.

Gli obiettivi di questo progetto di ricerca sono quindi:

1) Lo sviluppo di modelli matematici basati sulla teoria dei dati composizionali per analizzare i dati metagenomici generati all’interno del progetto o presenti in letteratura (es. ENA database);

2) l’implementazione e l’applicazione di pipeline bioinformatiche per l’analisi di dati metagenomici (genomica di patogeni e dell’ambiente) e genomici (profilazione del paziente a livello mitocondriale).

3) applicazione di metodi derivati dalla teoria dei network per caratterizzare comunitá di organismi, e modelli matematici per studiare l’evoluzione di questi network o di loro caratteristiche (es. diffusione).

Il progetto è altamente interdisciplinare, per cui sará necessaria una interazione con esperti in ambito virologico/batteriologico, epidemiologico, medico/clinico e di fisica ambientale.

**ENG**

Modern experimental techniques (shotgun next-generation sequencing) allow the collection of information relating to the population (eukaryotic, bacterial, viral, fungal) within specific ecosystems, both environmental (e.g. wastewater, aerosols, soil) and biological (e.g. intestinal microbiota and respiratory). These data can provide valuable information for the description of pathologies and their evolution, or to identify diagnostic markers or indicators of changes in the environment, for example due to anomalous or extreme events linked to climate change.

To analyze this type of data, you need 1) programming expertise to be able to manage complex high-performance calculation pipelines (e.g. parallel calculation, GPU calculation for Artificial Intelligence algorithms); 2) skills in the biostatistical field, to process data and carry out descriptive analyses, supervised or unsupervised stratification of data, dimensionality reduction; 3) theoretical skills, both relating to the type of data used (compositional data, for which ad hoc metrics and processing are needed) and to their integration (for example through network theory to identify communities of organisms and their evolution over time) .

The TRIGGER project aims to sample subjects and environments with genomic and metagenomics techniques with the aim of better understanding the effects of climate change on health: increased risk for subjects with specific pathological or genetic characteristics, variation in the pathogen population as a function of environmental factors such as pollution, temperature and humidity.

The objectives of this research project are therefore:

1) The development of mathematical models based on compositional data theory to analyze the metagenomic data generated within the project or present in the literature (e.g. ENA database);

2) the implementation and application of bioinformatics pipelines for the analysis of metagenomic (pathogen and environmental genomics) and genomic (patient profiling at the mitochondrial level) data.

3) application of methods derived from network theory to characterize communities of organisms, and mathematical models to study the evolution of these networks or their characteristics (e.g. diffusion).

The project is highly interdisciplinary, therefore interaction with experts in the virological/bacteriological, epidemiological, medical/clinical and environmental physics fields will be necessary.

**PIANO DI ATTIVITÀ**

ITA

Il progetto si propone di formare un ricercatore con competenze modellistiche e di analisi di dati multiomici con specifiche finalitá:

1) sviluppo e applicazione di algoritmi bioinformatici per analisi di dati genomici e metagenomici

2) sviluppo di modelli a network per caratterizzare sistemi ecologici in senso lato e studiare la loro evoluzione

3) applicazione della teoria dei dati composizionale per l’elaborazione dei dati e la loro successiva analisi statistica

4) interazione interdisciplinare in ambito biomedico e fisico

Il ricercatore sarà coinvolto direttamente nei progetti, e sarà coinvolto in tutte le fasi che vanno dalla raccolta dei dati alla loro analisi e modellizzazione.

Il ricercatore sarà invitato a elaborare i dati ottenuti sotto forma di report e articoli scientifici in lingua inglese.

Inoltre, il ricercatore sarà invitato a partecipare a lab meeting, congressi nazionali e internazionali per imparare nuove tecniche e presentare alla comunità scientifica i dati ottenuti.

ENG

The project aims to train a researcher with modeling and multi-omics data analysis skills with specific purposes:

1) development and application of bioinformatic algorithms for analysis of genomic and metagenomic data

2) development of network models to characterize ecological systems in a broad sense and study their evolution

3) application of compositional data theory for data processing and their subsequent statistical analysis

4) interdisciplinary interaction in the biomedical and physical fields

The researcher will be directly involved in the projects, and will be involved in all phases ranging from data collection to their analysis and modeling.

The researcher will be invited to process the data obtained in the form of reports and scientific articles in English.

Furthermore, the researcher will be invited to participate in lab meetings, national and international conferences to learn new techniques and present the data obtained to the scientific community.