



# Corso di Bioinformatica e Statistica A.A. 2016-2017

## parte V (Biocatalisi) e parte VI (Bioinformatica per le Biotecnologie Ambientali)

---



### Metagenomica e microbiomi

La classificazione microbiologica classica o i metodi molecolari basati sul sequenziamento Sanger non descrivono adeguatamente tutta la **biodiversità** delle comunità microbiche (**microbiomi**) in un **determinato ambiente** (tratto digerente, pelle, radici di una pianta, un terreno, uno stagno, un depuratore, acque reflue ecc...).

L'avvento dei **metodi NGS** ha consentito invece di ottenere contemporaneamente un numero enorme di sequenze, adeguatamente rappresentative dei differenti batteri presenti nel campione (approccio di **metagenomica**).

Dopo il sequenziamento NGS e l'assemblaggio dei frammenti, le sequenze sono confrontate *in silico* per **allineamento** con quelle depositate in database (o di campioni di riferimento). I dati sono analizzati sia dal punto di vista **qualitativo** (numero e tipo di differenti sequenze e appartenenza alle varie specie) che **quantitativo** (maggiore o minore rappresentatività nel campione di ceppi, specie, sottofamiglie e famiglie di batteri).



# Corso di Bioinformatica e Statistica A.A. 2016-2017

## parte V (Biocatalisi) e parte VI (Bioinformatica per le Biotecnologie Ambientali)

---



La tipizzazione metagenomica di microbiomi trova applicazione in numerosi campi:

- **scienze naturali**: il confronto dei profili metagenomici in vari ambienti descrive e predice cambiamenti e sostenibilità
- **biomonitoraggio e bioremediation**: la variazione metagenomica dei microbiomi in risposta a particolari sostanze inquinanti, misurata in termini di **comparsa/amplificazione/riduzione/scomparsa** di sottoprofili associati a ceppi/specie/famiglie di batteri, è un sensibile **bioindicatore**; allo stesso tempo, consente la selezione di ceppi utili alla bioremediation
- **agricoltura e allevamento**: i microbiomi simbiotici di piante e animali contribuiscono a **resa** e **qualità** di raccolti e derivati animali, alla **peculiarità di maturazione** di alcuni di essi (ad es. formaggi) ed i profili metagenomici degli ospiti possono anche **monitorare lo stato di salute** di piante e animali e prevenire il trasferimento di **patogeni** o loro prodotti (**tossine**) pericolosi per la salute umana
- **energia** attraverso l'identificazione o analisi comparata di microorganismi utilizzati o utilizzabili per biofuel
- **farmacologia**: la variazione di profili metagenomici in risposta ad antibiotici può rivelare dove nascono le **resistenze**; al contrario, variazioni in comunità microbiche prossimali ad organismi filtratori o comunque stanziali possono rivelare interessanti **attività antimicrobiche**.
- **medicina forense**, ad un livello più completo delle indagini realizzabili con approcci microbiologici classici.

Infine, profili metagenomici includenti batteri che producono metaboliti di interesse possono identificare specie correlate potenzialmente utilizzabili in progetti di selezione e/o ingegnerizzazione per la **biocatalisi**



# Corso di Bioinformatica e Statistica A.A. 2016-2017

## parte V (Biocatalisi) e parte VI (Bioinformatica per le Biotecnologie Ambientali)

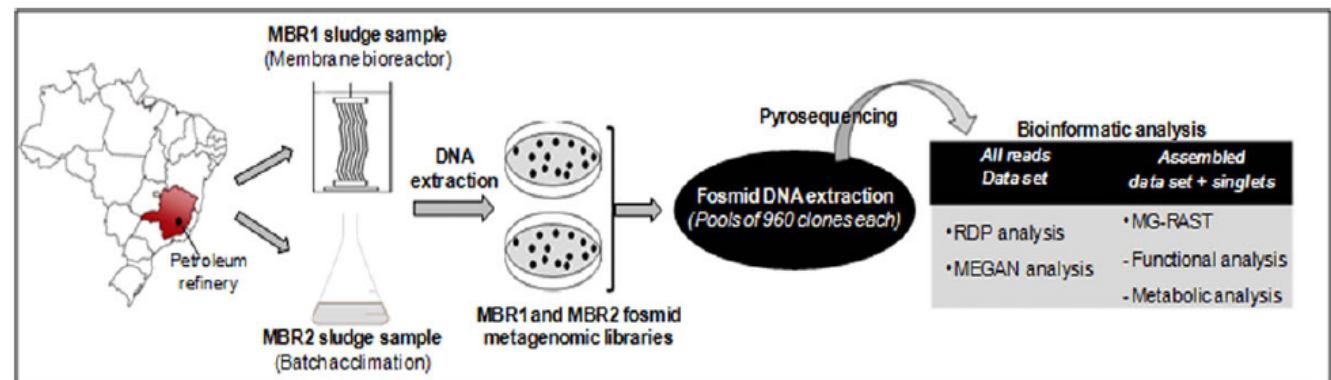
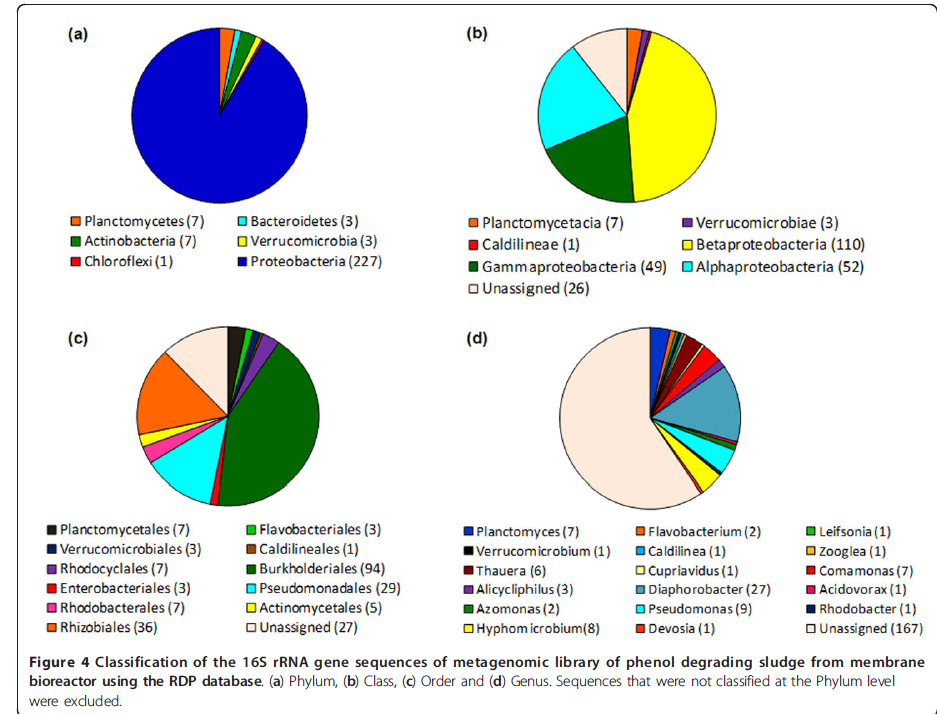


### Metagenomica, microbiomi e raffinerie

Le **acque reflue** di una raffineria (o di un allevamento, gli scarti industriali e fognari urbani) sono un ambiente a gradiente di concentrazioni e ad elevata biodiversità.

Profiling metagenomici e **analisi filogenetiche** ai vari stadi del percorso degli inquinanti (seguendo la concentrazione variabile degli stessi) permettono di seguire la riduzione in specie sensibili e l'arricchimento in specie potenzialmente dotate di **pathway e enzimi utilizzabili per bioremediation** (capaci cioè di degradare/trasformare le sostanze tossiche).

Per la classificazione tassonomica si confrontano le sequenze rRNA 16S con quelle depositate ad es. nell'European Ribosomal RNA database; a partire dalle altre sequenze si cercano i prodotti proteici per ORF finding e BLAST search.





# Corso di Bioinformatica e Statistica A.A. 2016-2017

## parte V (Biocatalisi) e parte VI (Bioinformatica per le Biotecnologie Ambientali)

---



### Metagenomica, metatrascrittomica e DB:

Negli anni recenti sono stati compiuti molti studi metagenomici e metatrascrittomici che hanno prodotto una grande quantità di dati. Questi ultimi sono quindi fruibili grazie all'interrogazione di database generali e speciali.

Un esempio di database specializzato - ed interrogabile via interfaccia web - è **AromaDeg**, un database relazionale curato manualmente (circa 4000 entries) e dedicato alla filogenomica dei batteri aerobi coinvolti nella degradazione di composti aromatici (in particolare, idrocarburi).

L'accurata annotazione manuale di alcune migliaia di schede permette attraverso ricerche per similarità di predire capacità metaboliche sulla base della vicinanza filogenetica.

Un altro esempio di database specializzato e web-based è **BioSurfDB**, dedicato alla contaminazione da olii grezzi, ai biosurfactanti ed alla biodegradazione. BioSurfDB integra e confronta:

- dati metagenomici da campioni a vari livelli di contaminazione con olii grezzi
- dati su geni, proteine e pathway coinvolti nella bioremediation
- risultati di esperimenti di bioremediation
- un elenco di biosurfactanti



# Corso di Bioinformatica e Statistica A.A. 2016-2017

## parte V (Biocatalisi) e parte VI (Bioinformatica per le Biotecnologie Ambientali)

---



### Computational toxicology

Campo interdisciplinare definito dall'EPA (Environment Protection Agency) "integrazione di biologia computazionale e molecolare per valutare il rischio chimico ambientale".

Consente un forte ampliamento delle sostanze sotto monitoraggio rispetto alla tossicologia tradizionale. Riduce i costi e la necessità di sperimentazione animale.

Alcuni tools riguardano la predizione delle caratteristiche chimiche. Ad esempio:

**EPI suite** (Estimation Program Interface) dell'EPA stima **proprietà chimiche di base** necessarie per il **percorso e l'assorbimento** dei reflui di lavorazione (coefficienti di partizione, idrosolubilità, velocità di idrolisi)

**SPARC** calcola varie **proprietà termodinamiche**.

Le limitazioni di EPI Suite e SPARC sono superate da programmi più complessi come **CATABOL**, un **simulatore di degradazione** che comprende una libreria gerarchica di trasformazioni abiotiche o enzimatiche.

Una grande risorsa è **PubChem** di NCBI, un database e sistema di ricerca pubblico che contiene più di 10 milioni di strutture chimiche e più di 500 bioassays. PubChem include anche dati di tossicità o attività antitumorale di circa 100000 sostanze saggate su 60 linee cellulari, poi integrati nel database relazionale **CellMiner**

Questi dati sono stati recentemente integrati con quelli provenienti da analisi molecolari (HTS), consentendo di tracciare profili abbastanza completi per migliaia di composti.



# Corso di Bioinformatica e Statistica A.A. 2016-2017

## parte V (Biocatalisi) e parte VI (Bioinformatica per le Biotecnologie Ambientali)

---



### Computational toxicology

Quando per una determinata sostanza non sono disponibili dati sperimentali rilevanti si ricorre al **modeling** dei pathway o di specifici step nel processo di tossicità cellulare.

I progetti di modeling che includono una **proteina target ed una serie di ligandi noti** permettono di estrapolare la possibile **compatibilità di legame** di un ligando putativo, oppure di predire possibili ulteriori ligandi.

In questi studi è fondamentale ottimizzare finemente i modelli (per consentire lo studio in dettaglio dei motif di legame) mediante cicli di **quality assessment** e **model refinement**.

I concetti di **flexibility** e **motion** sono cruciali per valutare correttamente possibilità e dinamiche di binding al target. In particolare, flessibilità e movimento locale sono determinanti per le cinetiche di binding. Si integrano pertanto approcci di molecular **docking** e **QSAR (Quantitative Structure Activity Relationship)** multidimensionali per predire come i **global modes** di protein motion influenzano interazioni tra subunità e ligandi.

L'automazione del processo iterativo di simulazioni e refinements può essere affidato al **machine learning**.



# Corso di Bioinformatica e Statistica A.A. 2016-2017 parte V (Biocatalisi) e parte VI (Bioinformatica per le Biotecnologie Ambientali)

---



## Computational toxicology

Gli approcci **QSAR (Quantitative Structure Activity Relationship)** e **QSPR (Quantitative Structure Property Relationship)** consentono di modellizzare dati sperimentali mancanti sulla base del **Similar Property Principle** *“molecole strutturalmente simili avranno proprietà chimico-fisiche e attività biologiche simili”*.

Tuttavia, **dov'è il confine tra tra simile e non simile?** Proteine con fold conservato (molto simile) possono avere funzioni simili o completamente diverse. Anche molecole piccole molto “simili” possono variare nelle proprietà. Dunque, tali strumenti predittivi restano utili ma le deduzioni sono sempre da “prendere con le pinze”!

Approcci **QSPR** sono stati utilizzati per studi del potenziale di dispersione e biodegradation associato a numerosi agenti surfactanti ed al loro impatto ambientale.

Il software **CODESSA** (Comprehensive Descriptors for Structural and Statistical Analysis) produce modelli QSAR e più di 400 descrittori per i surfactanti e consente di predire la cmc (critical micelle concentration) per la maggior parte dei surfactanti anionici e non ionici.

Predizioni più accurate sono state ottenute integrando QSPR e **algoritmi ANN (Artificial Neural Network)**



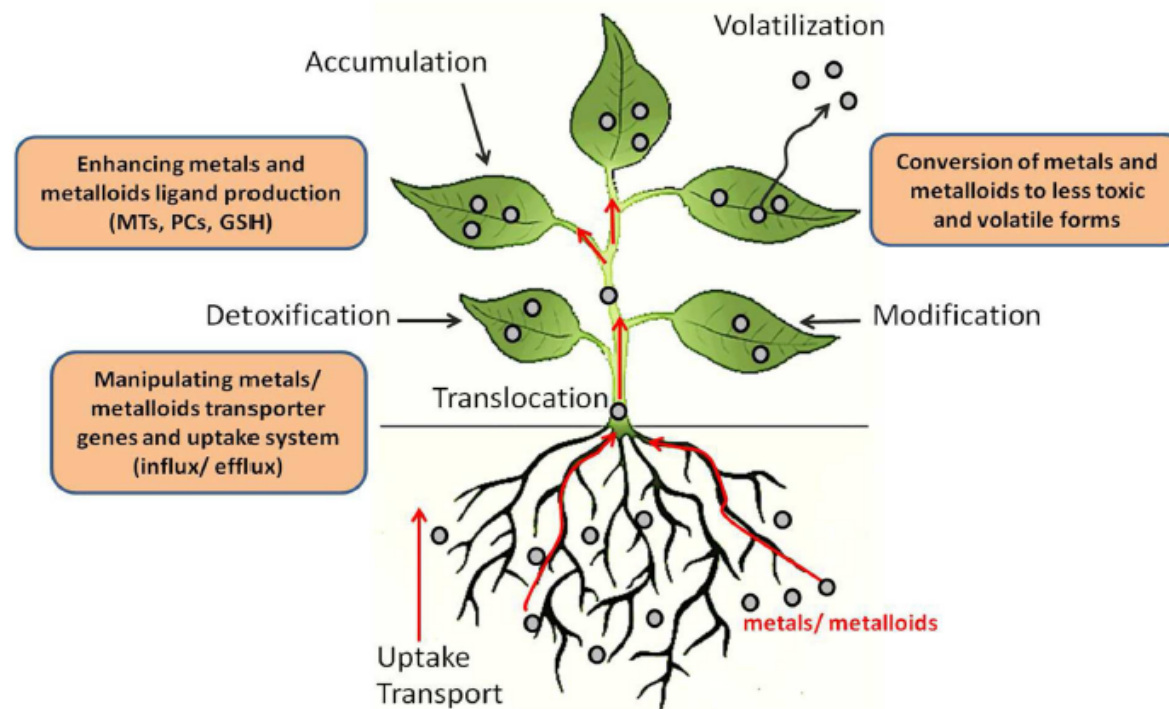
# Corso di Bioinformatica e Statistica A.A. 2016-2017 parte V (Biocatalisi) e parte VI (Bioinformatica per le Biotecnologie Ambientali)



## Strategie di Bioremediation:

- screening metagenomico di microbiomi e selezione di specie/ceppi capaci di metabolizzare inquinanti
- ricerca di omologi di geni/proteine già utilizzati per bioremediation
- directed evolution di geni/proteine per dotarli di capacità catalitiche volute
- rational design per bioremediation genes/proteins

## Strategie di Phytoremediation:







# Corso di Bioinformatica e Statistica A.A. 2016-2017 parte V (Biocatalisi) e parte VI (Bioinformatica per le Biotecnologie Ambientali)



## Strategie di Phytoremediation

