

Corso introduttivo di bioinformatica per docenti

Lezioni di Bioinformatica

Istituto:

Docente: Elisa Corteggiani
Carpinelli

Anno scolastico: 2019-2020

Alla base della bioinformatica

Un'idea molto semplice

Le macromolecole biologiche sono polimeri costituiti da catene di monomeri simili:

- DNA - catene di nucleotidi
- RNA - catene di ribonucleoidi
- Proteine - catene di aminoacidi
- Polisaccaridi - catene di monosaccaridi

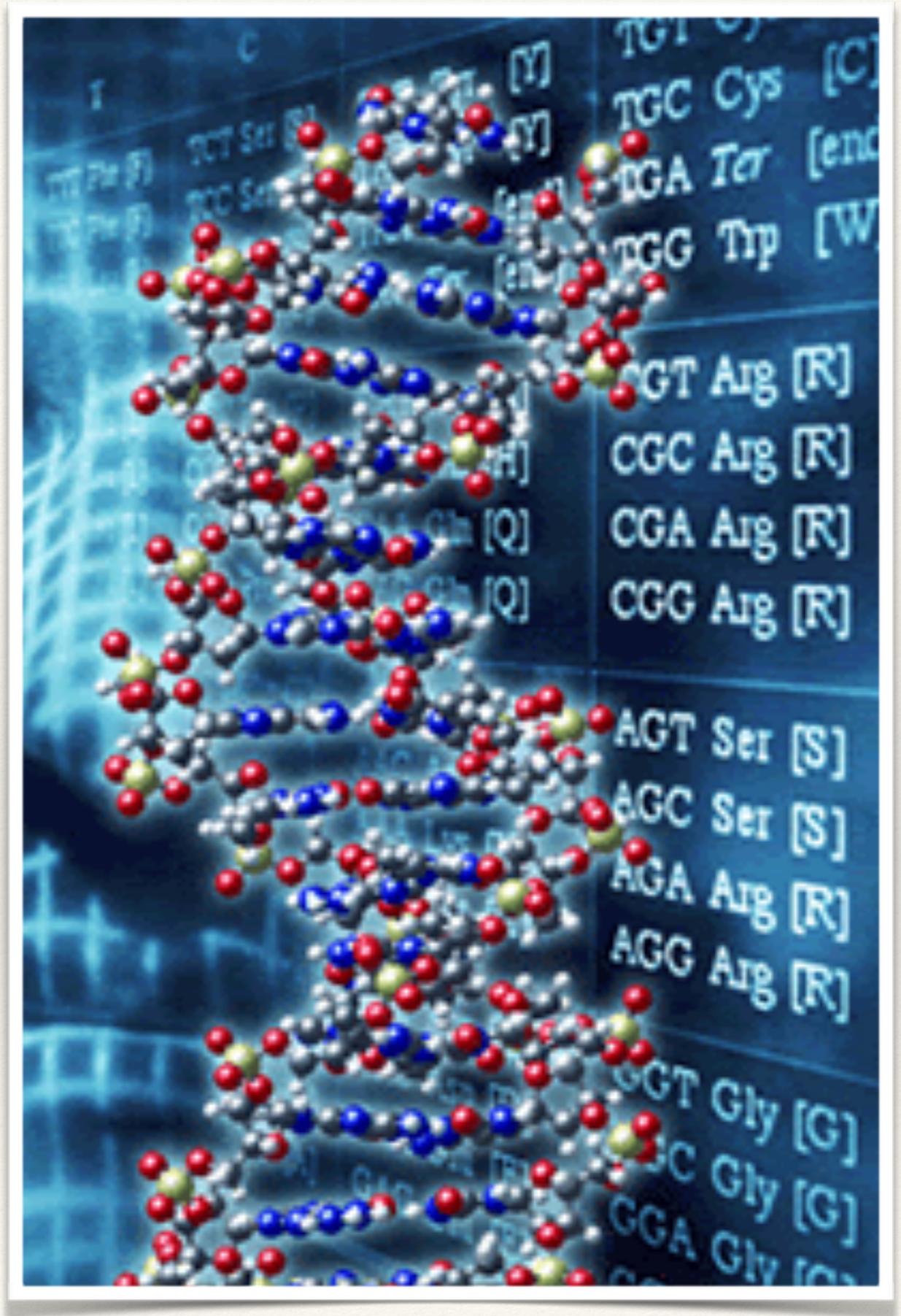


Alla base della bioinformatica

Ogni monomero è una lettera

I monomeri che si uniscono in sequenza a formare le macromolecole hanno alcune proprietà chimiche in comune ma differiscono uno dall'altro:

- le basi azotate sono 4 (A, C, T, G)
- gli aminoacidi sono 20 ...

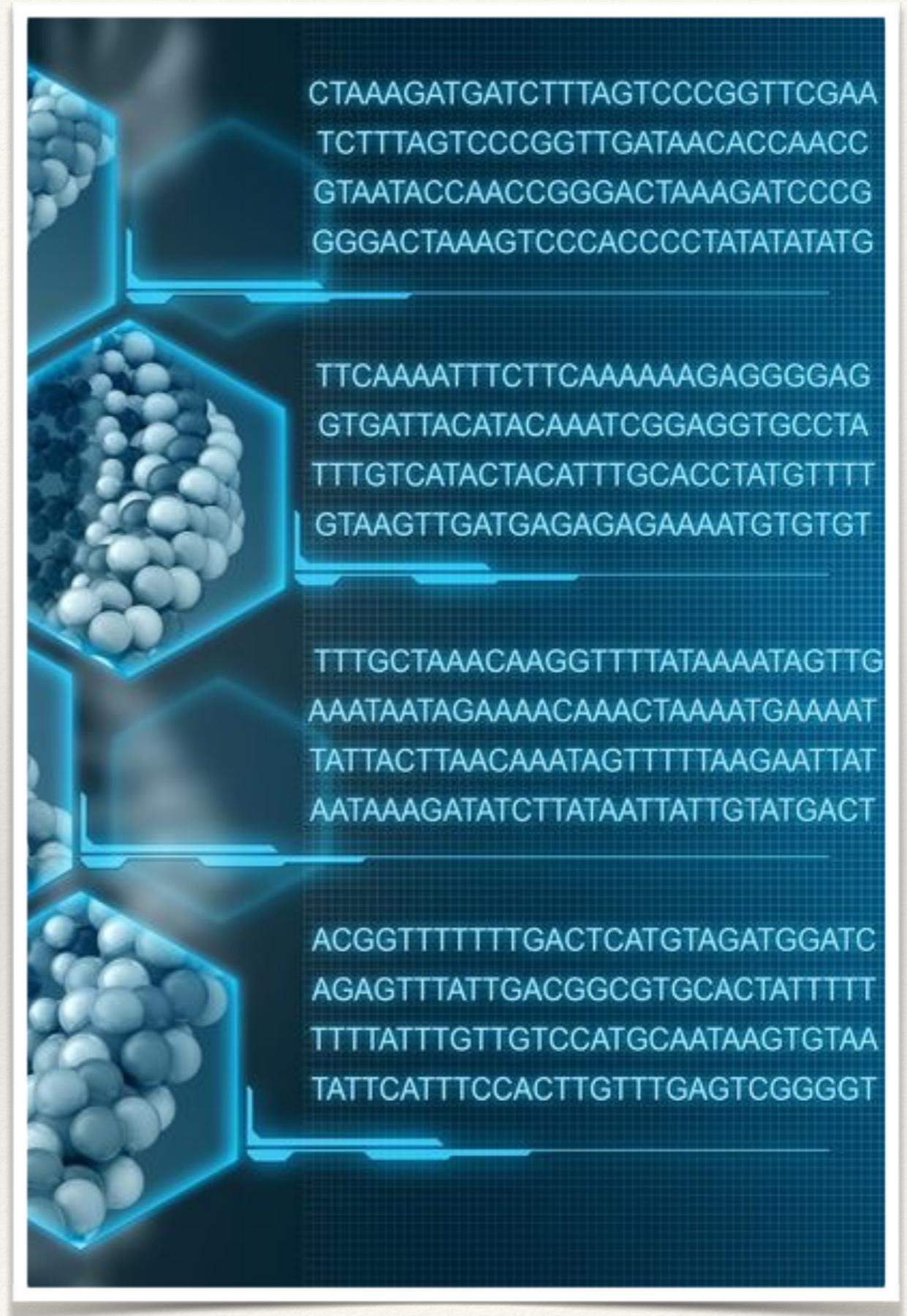


Alla base della bioinformatica

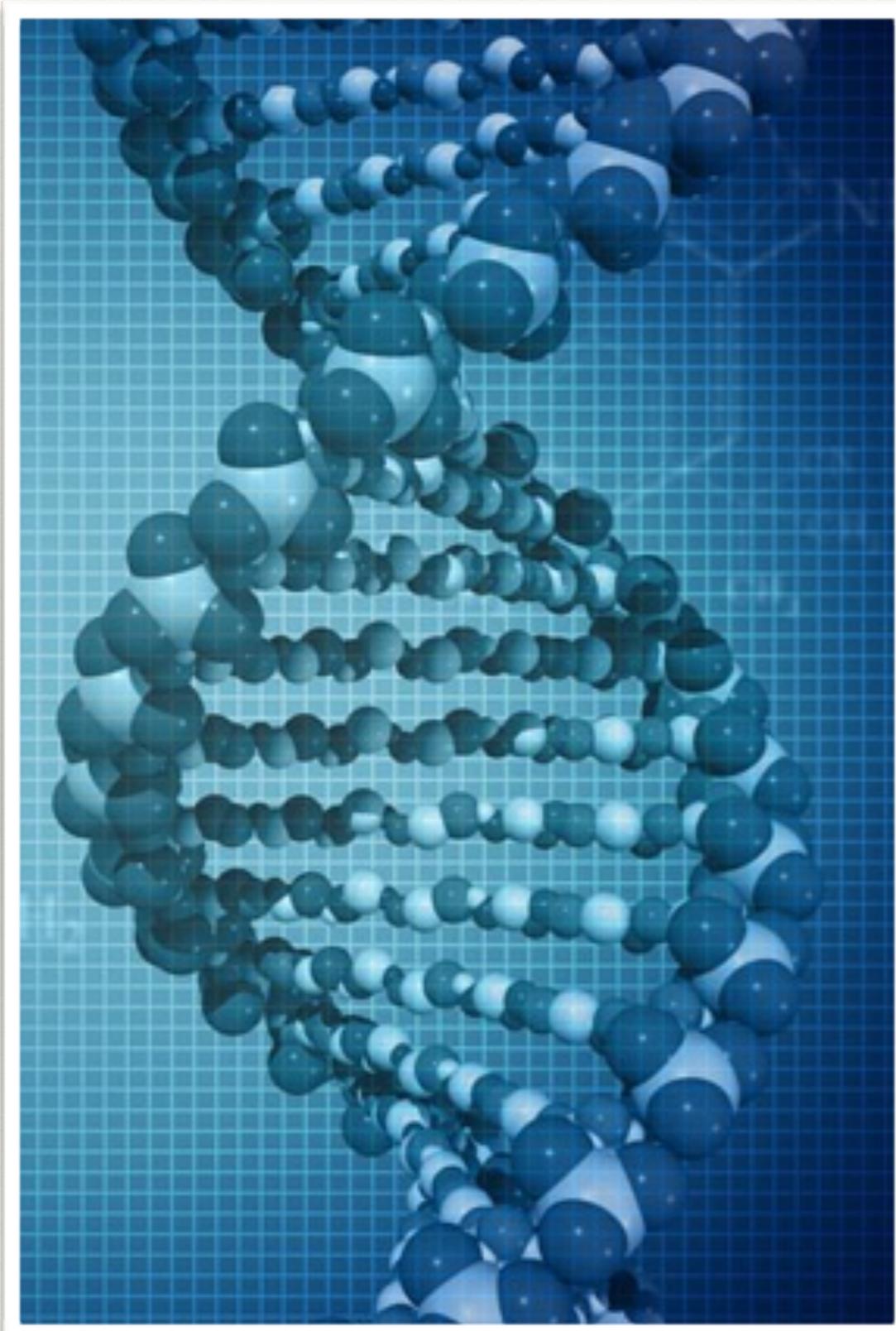
Un polimero è una sequenza di lettere

Un frammento di DNA costituito da una sequenza adenina-timina-guanina (ATG) ha una funzione biologica molto diversa da un frammento costituito da timina-adenina-adenina (TAA)

Una porzione di proteina costituita dalla sequenza lisina-lisina-lisina (KKK) ha proprietà chimiche diverse da ..



Il sequenziamento negli ultimi anni: sequenze prodotte e costi



numero di basi presenti in genebank e relative a progetti di sequenziamento di interi genomi (WGS)

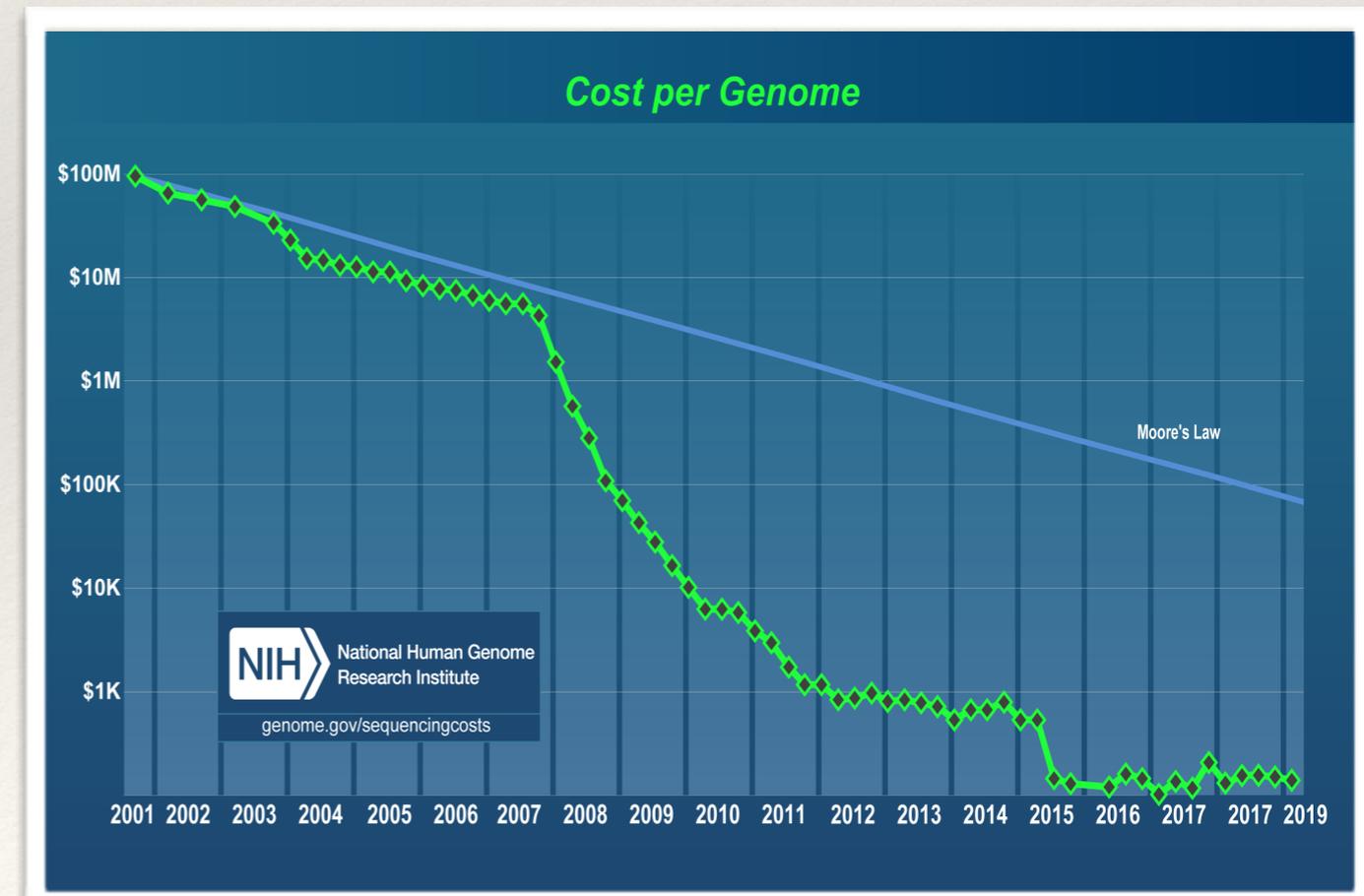
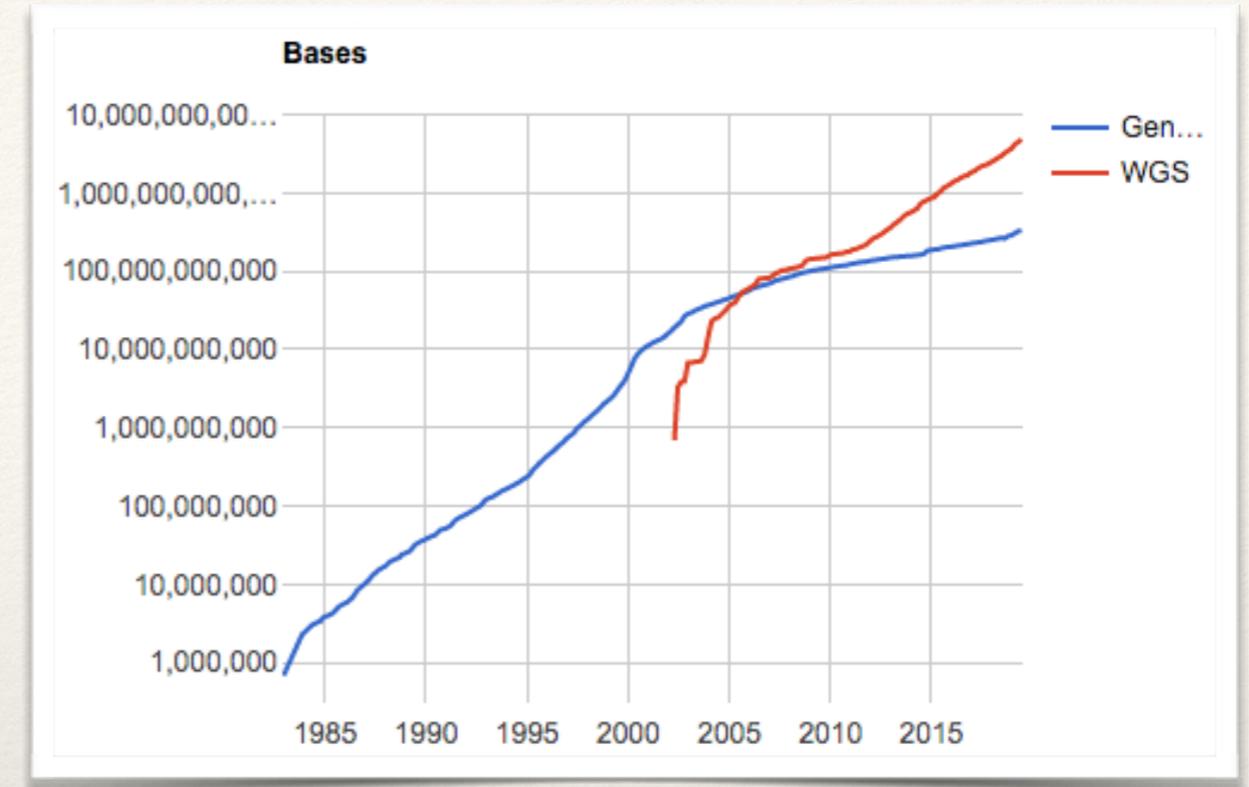
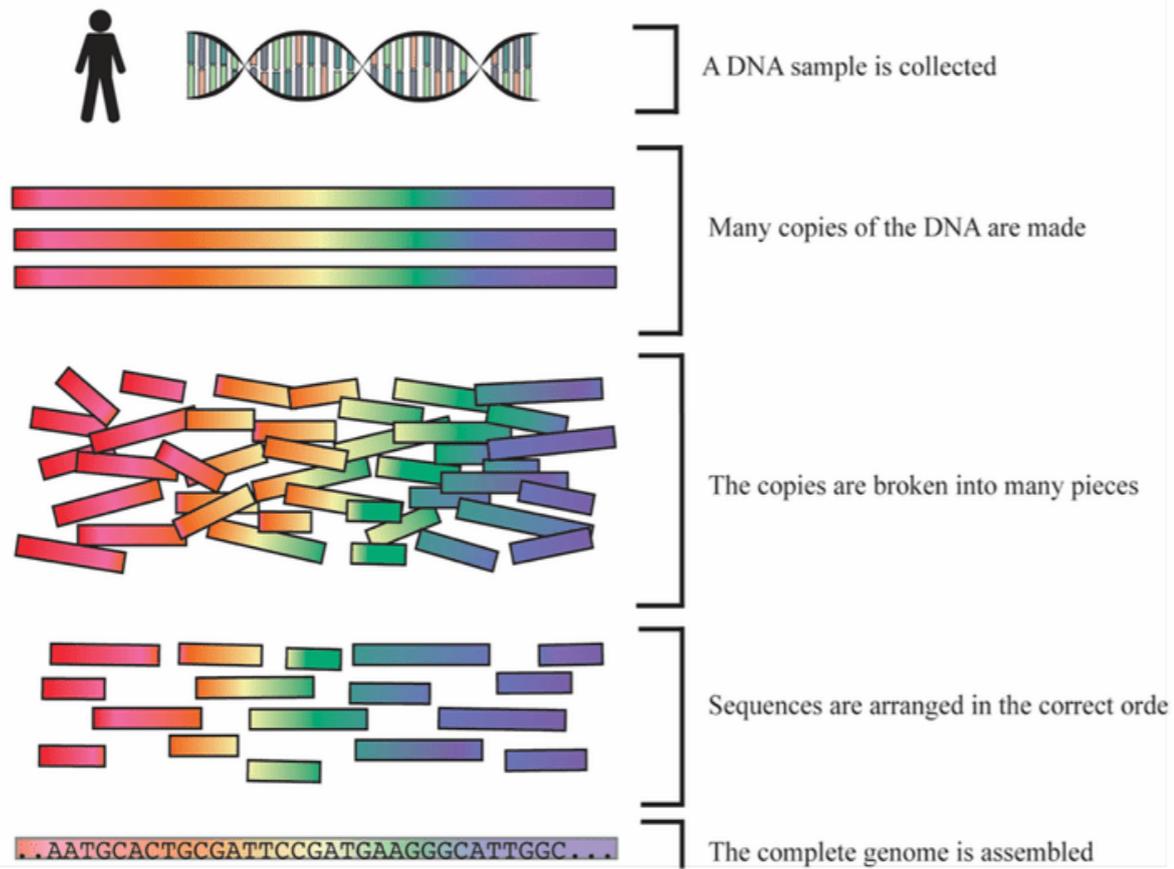
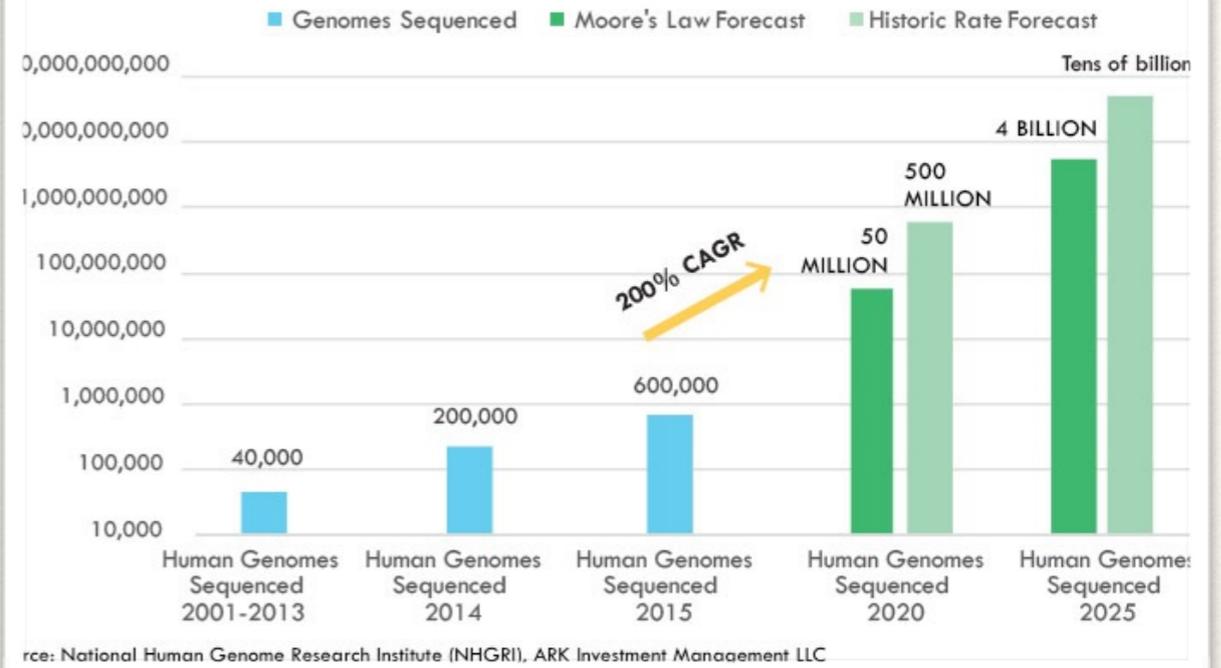


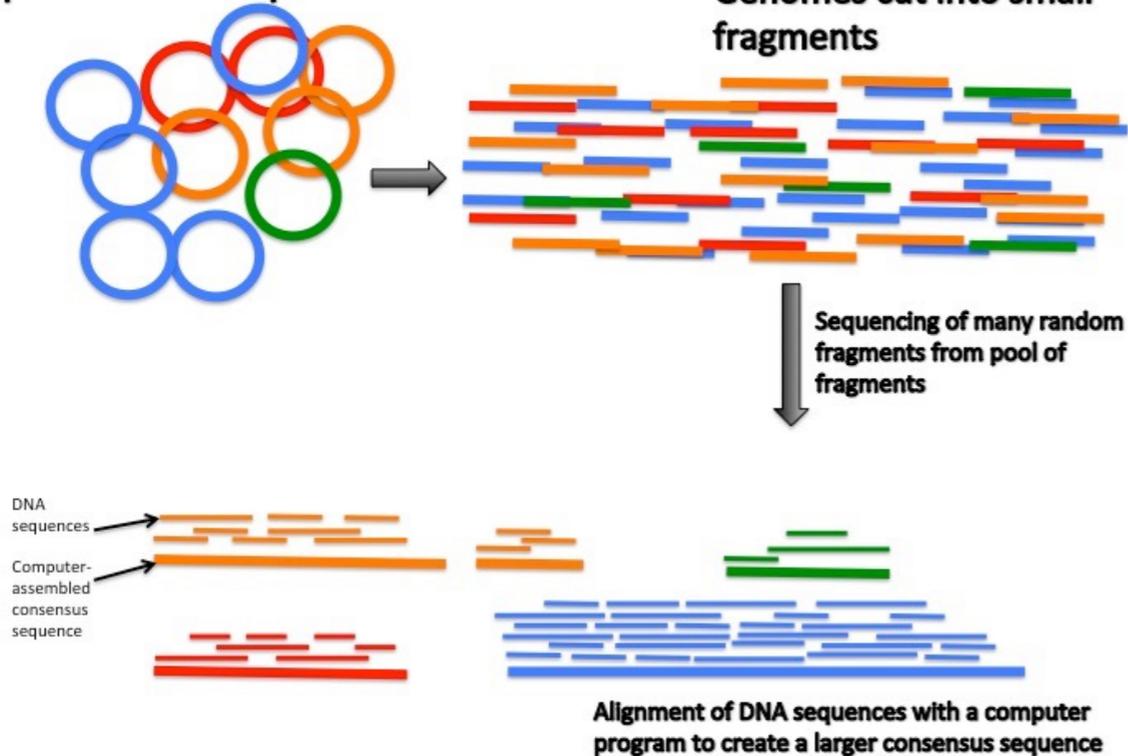
Figure 2: Shotgun Whole-Genome Sequencing



The Number of Human Genomes Sequenced (log scale)

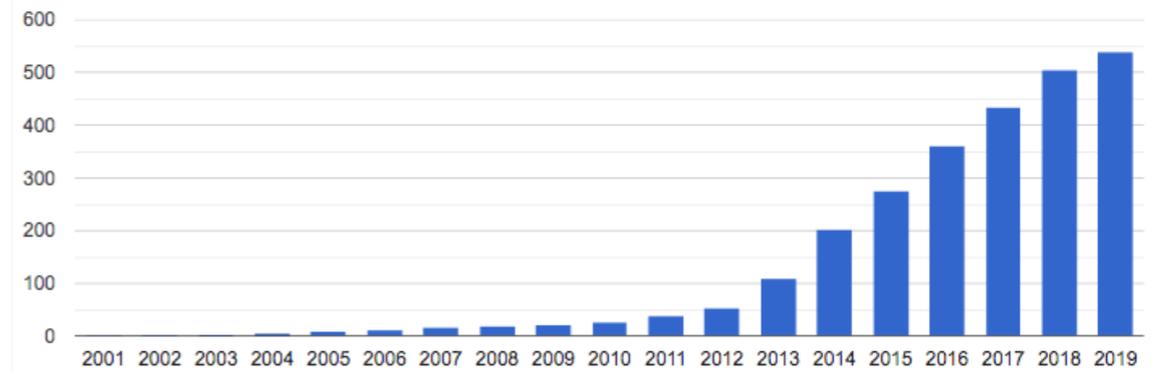


Bacterial genomes present in a sample

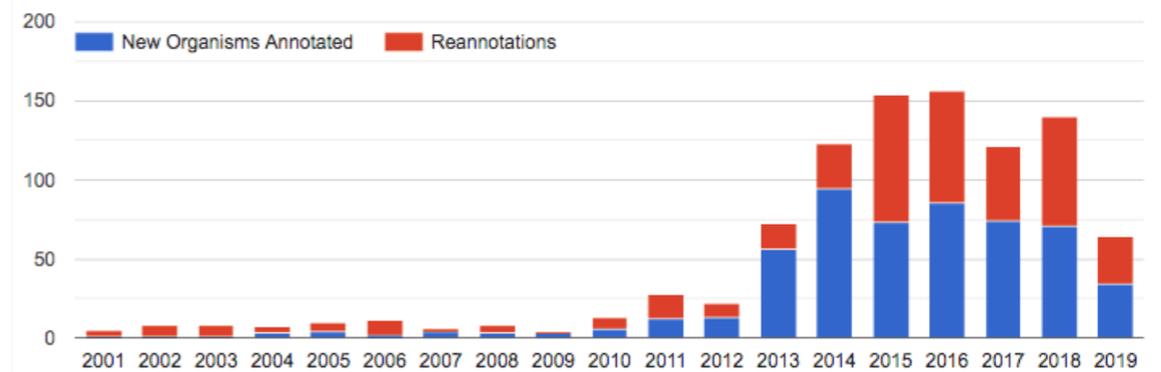


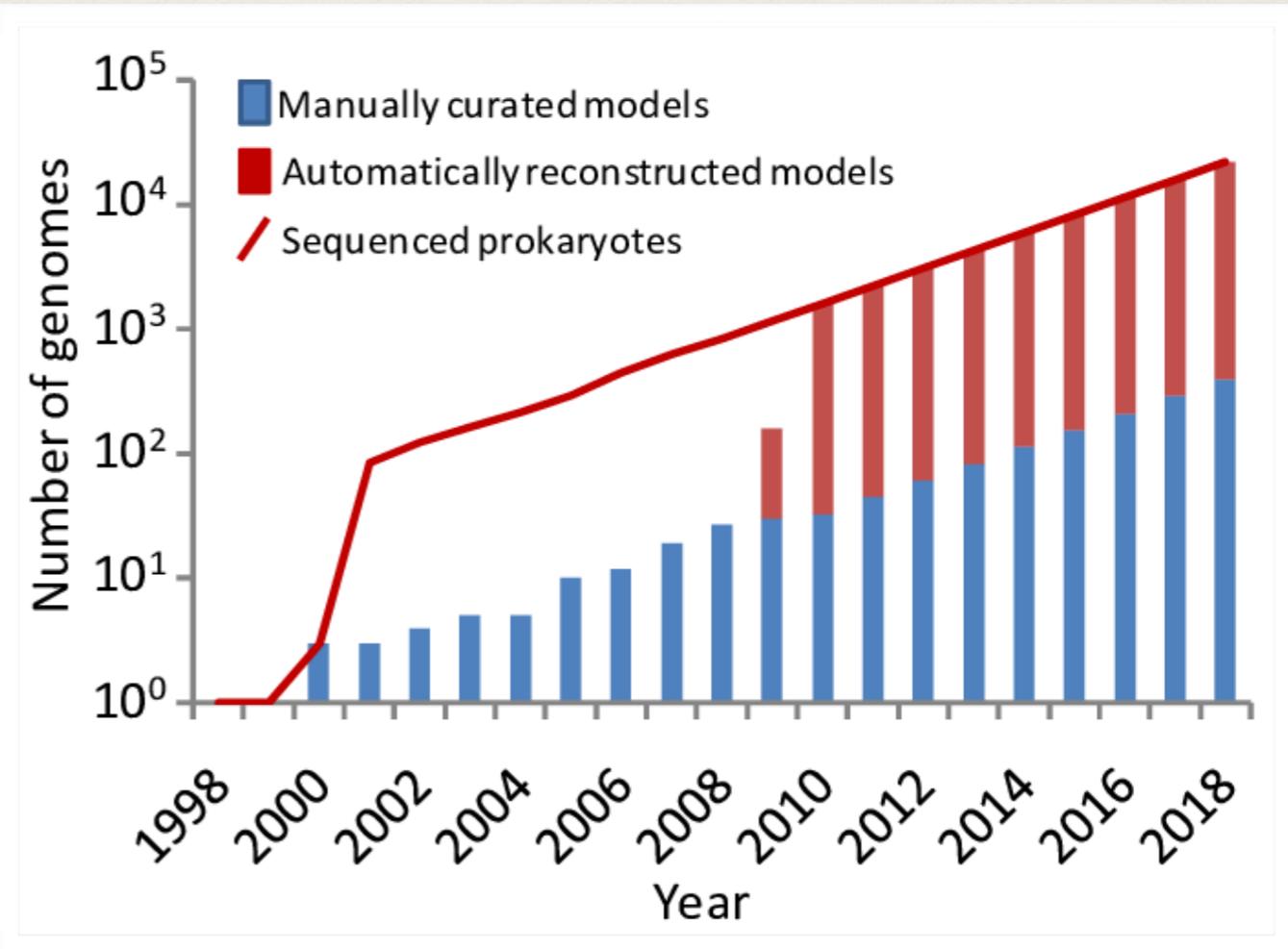


Cumulative Number Of Different Eukaryotic Genomes Annotated By NCBI



Eukaryotic Genomes Annotated By NCBI Per Year





Cosa ce ne facciamo di questo potenziale?

tecnologia sperimentale	applicazione	metodologie di analisi bioinformatica
sequenziamento <i>de novo</i>	sequenza del genoma di organismi interessanti o di costrutti per le biotecnologie (plasmidi e cromosomi artificiali)	assemblaggio e annotazione
ri-sequenziamento	Varianti fra genomi di singoli gruppi o singoli individui; riarrangiamenti cromosomici; varianti e geni malattia; mutazioni somatiche	comparazione con genomi di riferimento e ricerca di varianti
trascrittomica	caratterizzazione dei trascritti, livelli di espressione genica dei vari trascritti	assemblaggio dei trascritti; conta e analisi statistica dei livelli di espressione
studi di regolazione dell'espressione genica	studio degli RNA con funzione regolatoria; studio dei siti di legame dei fattori di trascrizione	identificazione, assemblaggio e conta dei trascritti; mappe dei livelli di espressione
analisi delle modificazioni epigenetiche	studi sulla metilazione del DNA e sulla modificazione strutturale dei nucleosomi	assemblaggio, comparazione e mappatura
metagenomica e diversità microbica	sequenziamento e classificazione dei campioni di DNA microbici provenienti da vari ambienti; studio della presenza e dell'attività metabolica dei vari microbi e comunità microbiche	assemblaggio, classificazione, conta e analisi statistica delle sequenze, raggruppamento dei profili
paleogenomica	sequenziamento del DNA antico	assemblaggio e comparazione

Come capire qualcosa da tutti questi dati?

- ❖ il futuro della biologia è nell'informatica!
- ❖ La grande sfida sta nell'interpretazione di una enorme quantità di informazioni che sono state raccolte per noi ormai facilmente e che aspettano solo di essere interrogate!

La bioinformatica come strumento per insegnare la biologia

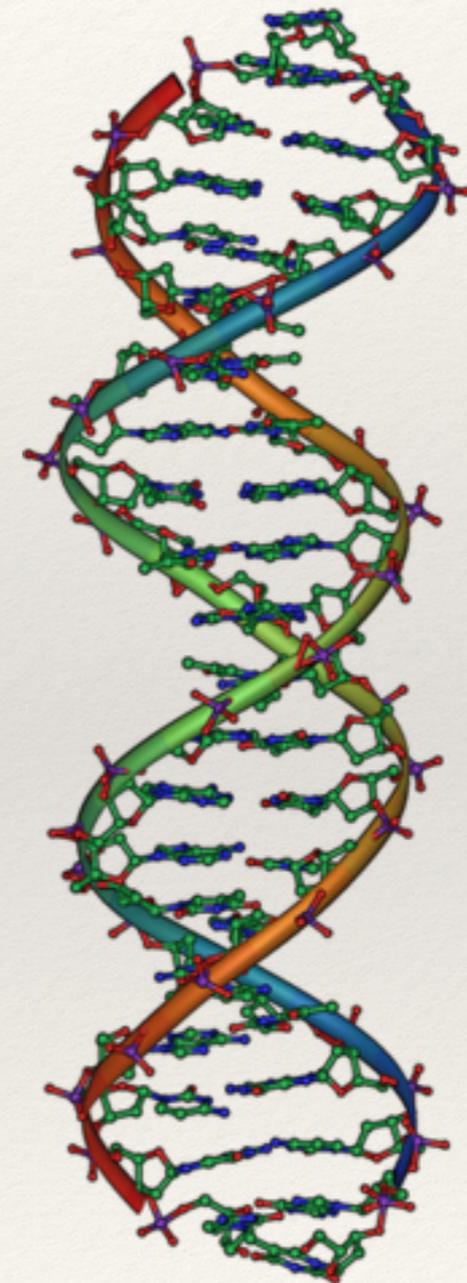
Portiamo la Bioinformatica nei licei

- ❖ Perché è importante
 - ❖ Quali ambiti della bioinformatica introdurre
-

Qualche buona ragione

❖ **Contenuto**

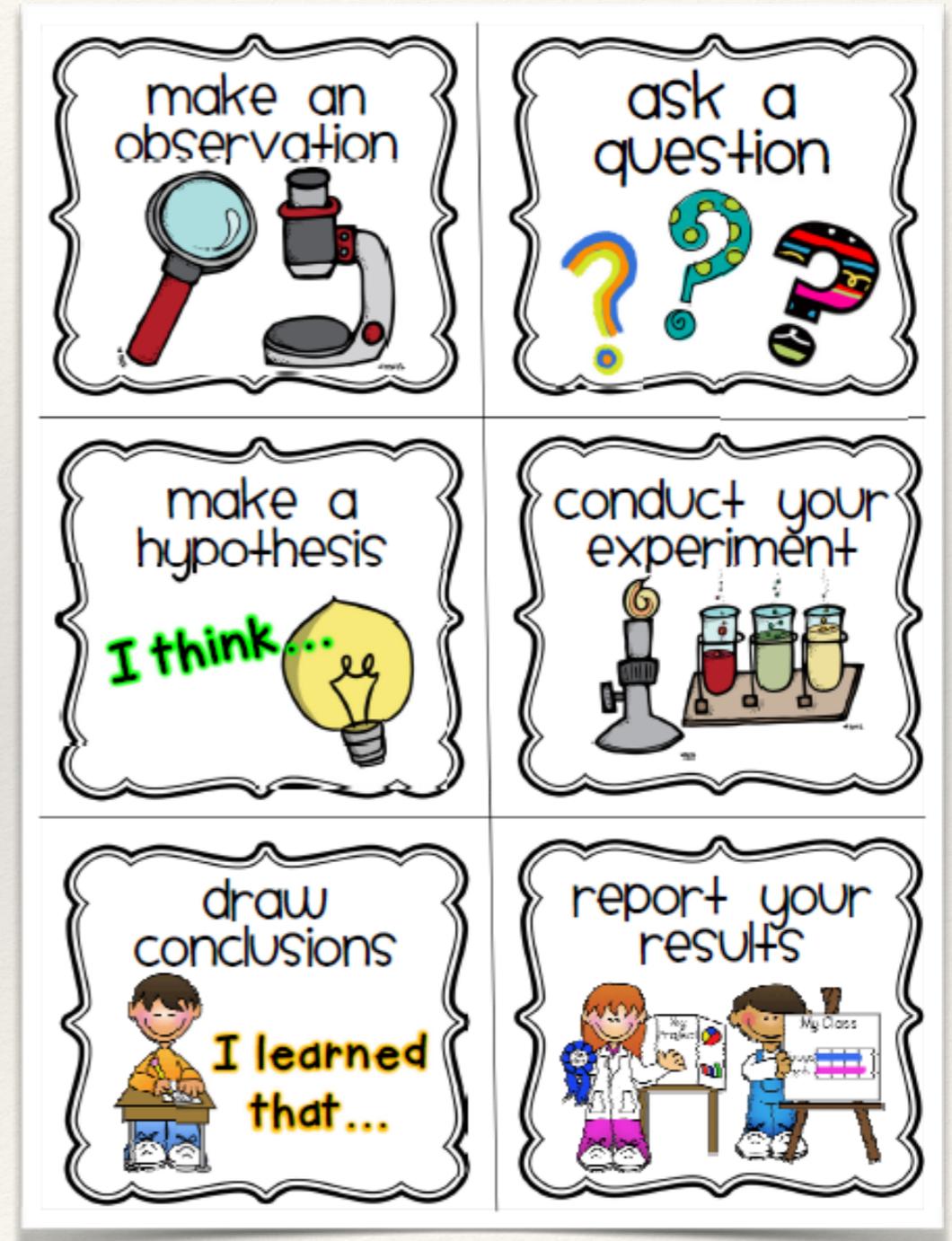
- ❖ la bioinformatica si presta molto bene per affrontare molti concetti di biologia molecolare, biologia umana, genetica ed evoluzione
- ❖ iniziare ad usare le banche dati e molti software di analisi bioinformatica dà agli studenti un insieme di strumenti che saranno utilissimi durante gli studi universitari scientifici.
- ❖ I contenuti sono diversi nei vari curriculum di studi:
 - ❖ insegnare a sviluppare programmi per analizzare dati o
 - ❖ insegnare ad utilizzare programmi preesistenti per analizzare un particolare gruppo di dati.



Qualche buona ragione

❖ Metodo

- ❖ la bioinformatica è un ottimo metodo per far esercitare gli studenti con il metodo sperimentale
- ❖ si può applicare a domande molto concrete e attuali
- ❖ utilizza strumenti moderni che gli studenti conoscono bene
- ❖ molte applicazioni sono estremamente interattive
- ❖ si presta per assegnare agli studenti percorsi differenziati e compiti individuali



Argomenti

- ❖ Filogenesi
- ❖ Geni
- ❖ caratteristiche dei trascritti
- ❖ Genetica umana: varianti geniche e geni
- ❖ Epigenetica
- ❖ Enzimi e proteine
- ❖ Mappe del metabolismo cellulare



Guida alla realizzazione di lezioni utili e contestualizzate

10 regole d'oro

Cosa tenere a mente per organizzare una buona esperienza pratica di bioinformatica

da: David Form and Fran Lewitter (2011) "Ten Simple Rules for Teaching Bioinformatics at the High School Level" Plos Computational Biology

1. Esercizi semplici

- ❖ Impostare uno o pochissimi obiettivi per ogni esercitazione
- ❖ Gli obiettivi devono essere tecnicamente facili da raggiungere, in modo tale che gli studenti non perdano di vista perché stanno facendo le analisi
- ❖ Esempio pratico: in un file GenBank trovare la sequenza codificante e salvarla in formato FASTA
- ❖ Una lezione successiva spiegherà il significato biologico e l'utilità applicativa di questa semplice azione

2. Familiarità con l'argomento

- ❖ Ci saranno inevitabilmente molti concetti nuovi durante una esercitazione di bioinformatica, pertanto è importante cercare di usare come esempi geni o malattie che gli studenti conoscano bene.
- ❖ Esempio pratico: se si analizza un gene si può scegliere un gene che codifichi per una proteina che gli studenti conoscono bene come l'insulina o ci si può occupare di una variante genica che abbia a che fare col colore degli occhi.

3. Applicare la bioinformatica a concetti già affrontati in biologia

- ❖ Usare la lezione di bioinformatica solo per spiegare in un altro modo concetti del programma, non per fare cose in più
- ❖ Esempio pratico: se a lezione è stata spiegata la proteina emoglobina, durante l'esercitazione di bioinformatics si può visualizzare con gli student il modello 3-D della proteina ai diversi pH
- ❖ oppure si potrebbero visualizzare i modelli 3-D di un enzima che conoscono e di una proteina strutturale che conoscono (es. emoglobina e collagene) per mettere in luce le differenze

4. Costruire un progetto

- ❖ Si possono costruire piccoli progetti con gli studenti organizzando una serie di esercitazioni successive:
- ❖ ciascuna esercitazione insegna una sola abilità alla volta
- ❖ i concetti vengono elaborati in modo tale che il progetto progredisca dalla prima all'ultima esercitazione.
- ❖ Esempio pratico: mettiamoci a studiare le opsine

5. Domande e scoperte

- ❖ Gli studenti imparano meglio quando hanno una loro curiosità da inseguire e quando loro notano o scoprono o osservano cose interessanti. Quindi perchè non procedere assegnandogli delle sfide e poi guidandoli nella risoluzione dei problemi?
- ❖ Esempio pratico: ad uno studente che è interessato ad una certa malattia (perché ne è affetto, perché se ne parla nei telegiornali..) si assegna il compito di cercare la sequenza dell'RNA. Lo studente scoprirà che non trova una ma molte sequenze di RNA.. Perché? Dopo avergli spiegato che l'RNA subisce delle modifiche post-trascrizionali gli sarà più chiaro quello che ha scoperto e sarà capace di spiegare ai suoi compagni di classe l'RNA ending e lo splicing alternativo, quando racconta i risultati delle sue ricerche su quella malattia.

6. A ciascuno il suo progetto

- ❖ Gli studenti sono più interessati se hanno un loro progetto di cui sono responsabili.
- ❖ Le attività di bioinformatica non devono necessariamente essere svolte in aula tutti insieme, si possono anche assegnare durante l'anno a ciascuno studente dei piccoli progetti su cui riferire in classe.
- ❖ applicando la bioinformatica di volta in volta ad argomenti diversi del programma

7. Tanti stili per imparare tutti

- ❖ Le esercitazioni di bioinformatica possono focalizzarsi su molti stili e aspetti interattivi diversi, ciascuno dei quali può essere più congeniale a studenti diversi
- ❖ Esempio pratico: si può studiare leggendo le informazioni di una proteina; si può utilizzare la rappresentazione grafica del risultato di un blast; si possono esaminare o produrre grafici e statistiche.

8. Mettiamoli alla prova

- ❖ Per fissare le competenze acquisite, proponiamo agli studenti un problema interessante a cui trovare risposta
- ❖ Per farlo dovranno fare appello alle competenze acquisite durante una delle esercitazioni di bioinformatica e così faranno veramente propria quella competenza
- ❖ Questa è un grande insegnamento, perché la bioinformatica è proprio questo: una sfida continua nella quale ci sono domande, ci sono dati e dobbiamo aprire una cassetta degli attrezzi per trovare risposte in quei dati
- ❖ Esempio pratico: assegnare come compito per casa a fine lezione una curiosità a cui trovare risposta in letteratura o in un dato di sequenza o in una banca dati di fossili

9. Progettare prima di realizzare

- ❖ Una analisi bioinformatica è un esperimento vero e proprio e in quanto tale, anche se non richiede l'uso di consumabili va progettato prima di essere eseguito
- ❖ E' una ottima abitudine scrivere prima il progetto su carta, magari con una mappa concettuale e discuterlo
- ❖ Facendo così al momento della realizzazione gli studenti avranno più chiaro il significato di ogni passaggio e saranno più capaci di individuare eventuali errori

10. Realizzare un prodotto

- ❖ Se usando le loro competenze gli studenti arrivano a produrre qualcosa allora le conoscenze, le competenze ed il livello di entusiasmo aumentano e l'impegno con il quale svolgono il loro compito anche
- ❖ Esempi pratici: se stanno studiando la struttura e la funzione di una proteina possono progettare un farmaco che la blocchi; se stanno reperendo delle informazioni sul livello di similarità di un gene in diverse specie possono realizzare un grafico e pubblicarlo on line per rendere l'informazione disponibile anche ad altri

I risultati sperimentali sono sempre il miglior punto di partenza

La raccolta di dati

- ❖ 3 banche dati che non possiamo non conoscere
 - ❖ anche la biologia più tradizionale si ammoderna..
 - ❖ il genoma umano e i microbiomi
-

Banche dati di sequenze

- ❖ NCBI (<http://www.ncbi.nlm.nih.gov/>)
- ❖ EBI (<http://www.ebi.ac.uk/>)
- ❖ PDB (<http://www.rcsb.org/pdb/home/home.do>)



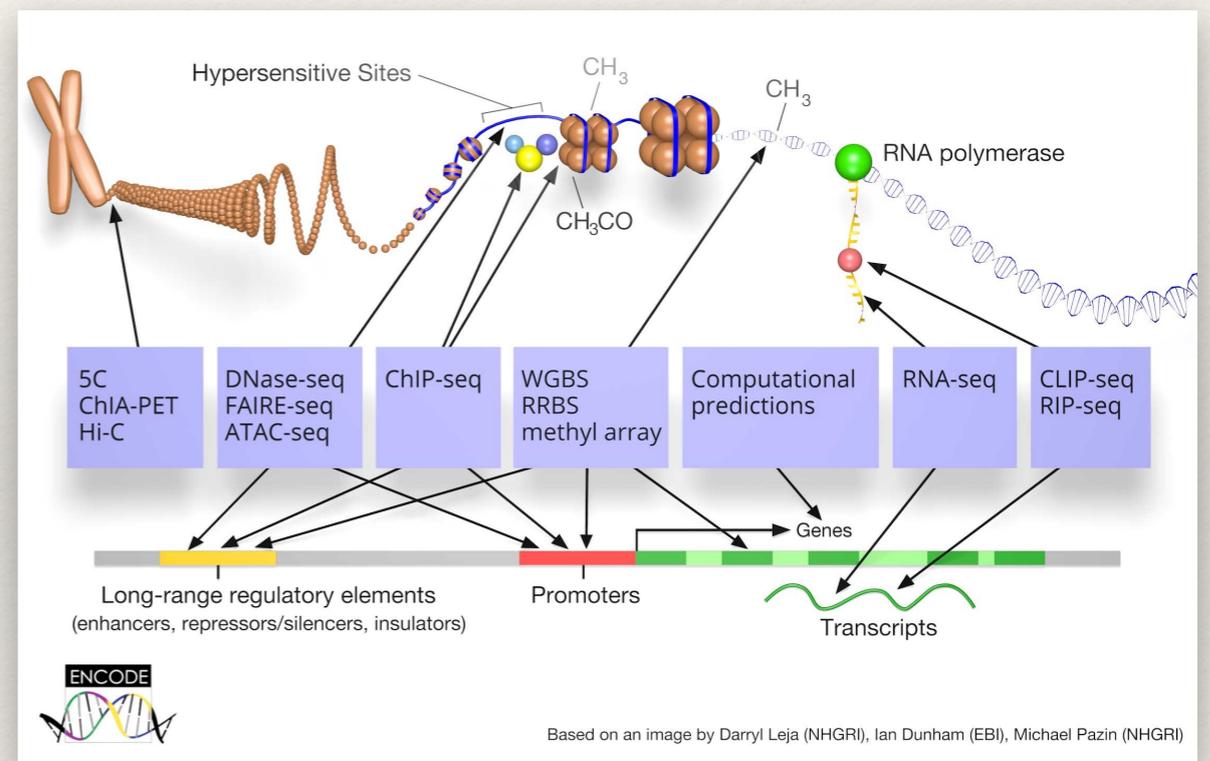
Banche dati di reperti fossili

- ❖ Paleobiology Database (<http://fossilworks.org/?a=home>)



L'enciclopedia del genoma umano

- ❖ The Encyclopedia of DNA Elements (ENCODE) (<https://www.encodeproject.org/>)
- ❖ Risorsa sui principali genomi eucariotici (e!Ensembl) (<https://www.ensembl.org/index.html>)



metagenomica e microbiomi

- ❖ il portale della genomica microbica (<https://www.ebi.ac.uk/metagenomics/>)

The screenshot shows the EBI Metagenomics website interface. At the top, there is a navigation bar with the EBI logo and the text "EBI Metagenomics". To the right, there is a "secure login" button. Below the navigation bar, there is a "navigation menu" with a "SUBMIT DATA" button. The main content area is divided into several sections:

- resource stats:** A table showing various statistics:

8316 data sets	4066 metagenomics	5848 runs	2468 runs
	847 metatranscriptomics	5079 samples	2539 samples
	3336 amplicons	143 projects	106 projects
	67 assemblies		
- Browse projects:** A section titled "public project browsing options" with a sub-section "By selected biomes". It features icons for different biomes: Soil (18), Marine (30), Forest (4), Non-human host (38), Engineered (12), Freshwater (9), Grassland (5), Human gut (22), Air (1), and Wastewater (2). There is a "View all biomes" link.
- Latest projects:** A list of recent projects with titles and brief descriptions, such as "Effect and mode of action of chitosan and ivy fruit saponins on the microbiota, fermentation and methanogenesis in the rumen simulation technique".
- Spotlight project highlight:** A section titled "TARA ocean project" featuring a map of the TARA ocean project routes and the text "Planitron ecosystems contain a phenomenal reservoir of life: more than 10 billion".
- Tools analysis tools:** A section titled "Functional sample comparison" featuring a bar chart comparing sequencing runs within a project.

Proviamo a fare noi un esercizio che potremmo proporre agli studenti

Prima esercitazione

Iniziamo da una proposta
introduttiva sulle sequenze
di DNA

Sequenze di DNA e resistenza agli antibiotici

❖ <http://www.sperimentando.com/?p=69>

La ricerca in letteratura: dati affidabili e bufale on-line

La letteratura scientifica

a portata di click!



- ❖ c'è un tutorial per ogni cosa (a patto di capire l'inglese) : <http://www.nlm.nih.gov/bsd/disted/pubmed.html>
- ❖ **Proviamo noi a fare qualche esercizio:**
 1. Risalite al nome completo della rivista: PNAS
 - ❖ Inoltre verificate se è accessibile gratuitamente da rete;
 - ❖ Accedete alla home page della rivista. Cercate tramite il sito un articolo su un argomento di vostro interesse. Salvate in un file di testo la query effettuata ed il titolo dell'articolo che, dalla lista dei risultati, vi sembra più interessante.
 2. Accedete con PubMed alla lista di pubblicazioni del professor John Walker, insignito del premio Nobel per le sue scoperte sull'ATP sintasi. Provate a cercare
 1. Walker JE[Author]
 2. Walker J[Author]
 3. Walker J
 4. Walker John
- ❖ Aprite un file di testo e iniziate a riportare i risultati dei vari esercizi che faremo. Riportate i risultati ottenuti per questa ricerca su una tabella dati, provate a fare qualche osservazione e proporre interpretazioni dei risultati.



3. Ricercate un articolo in PUBMED rispettando le fasi sotto elencate:

- ❖ cercate un articolo inerente al comportamento sociale delle api usando le seguenti parole “*Apis mellifera social behaviour*”
- ❖ scorrete con attenzione i risultati e annotate quanti articoli trovate, la data di pubblicazione e gli argomenti a cui i titoli fanno più frequentemente riferimento
- ❖ ora ripetete la ricerca considerando, accanto alla ricerca iniziale “*Apis mellifera social behaviour*” i termini “methylation” e “epigenetics”

Durante la ricerca ricordate di utilizzare gli operatori booleani (AND, OR...) e i limiti.

- ❖ Scrivete in un file di testo le query che avete formulato.
- ❖ Quanti articoli vengono trovati? Qual è l’anno di pubblicazione? Ce ne sono di accessibili liberamente? Cosa avete imparato da questa ricerca?

potete inventarvi questo esercizio a partire da qualsiasi notizia di scoperta scientifica che compare su un giornale e far verificare agli studenti quale sia la fonte di questa notizia, se ce n’è una..



4. Verificate le informazioni riportate nel seguente articolo uscito qualche anno fa su “la Repubblica” http://www.repubblica.it/salute/ricerca/2014/12/16/news/con_un_farmaco_lautismo_si_potr_battere-103045571/
 - ❖ cercate su Pubmed l’articolo scientifico a cui si fa riferimento nel post di “La Repubblica”
 - ❖ quali sono i risultati della ricerca sperimentale? Sa avete un congiunto affetto da autismo, pensate che potrà presto prendere un farmaco sperimentale?

5. Verificate le informazioni riportate nei seguente articolo usciti qualche tempo fa http://www.repubblica.it/scienze/2014/05/07/news/dna_espanso_basi_artificiali-85497745/ ; <http://www.liberoquotidiano.it/news/11608390/Dna--verso-la-vita-artificiale.html>
 - ❖ cercate su Pubmed l’articolo scientifico a cui si fa riferimento negli articoli
 - ❖ quali sono i risultati della ricerca sperimentale? Che utilità potete immaginare che trarremo in medicina da questi risultati?

La potenza del peer-review sul lungo periodo

6. A giugno 2019 i giornali hanno riportato la notizia di una pubblicazione scientifica che sosteneva l'efficacia terapeutica dell'omeopatia per il controllo del dolore, ritirata perché i dati che presentava erano “falsati” (questi i link alla [notizia dell'ansa](#) ripresa poi ad esempio da “[Il messaggero](#)” e “[il fatto quotidiano](#)”). La rivista scientifica ha immediatamente ritirato l'articolo [pubblicando un comunicato](#).
- ❖ siete in grado di ricercare informazioni su un analogo caso di ritiro di un articolo per la presenza di dati falsati questa volta riguardo alle presunte controindicazioni dei vaccini? La rivista è “The lancet” e l'anno di pubblicazione è 1998. Usate Pubmed e riportate sul vostro file le parole chiave usate per la ricerca e i risultati ottenuti.
 - ❖ Cosa dicono i medici e i ricercatori oggi riguardo alla sicurezza dei vaccini? Puoi commentare l'impatto sulla comunità scientifica e sulla società di questa frode scientifica?



La rete è una fonte di informazioni ricchissima
sono i nostri strumenti culturali che ci aiutano a verificare
l'attendibilità delle notizie e dei dati che troviamo nella
rete

E' vero anche quando usiamo google!

sulle notizie scientifiche abbiamo due marce in più: la
capacità di consultare la letteratura scientifica e il nostro
intelletto critico



- ❖ a questo link è disponibile un esempio di esercitazione che potete fare con i vostri studenti e che viene periodicamente aggiornata: <http://www.sperimentando.com/?p=43>
- ❖ ovviamente potete costruirne una simile su notizie che hanno incuriosito i vostri studenti o su esperienze storiche che è importante non ripetere

Tante esercitazioni già pronte on line da proporre agli studenti

Risorse per gli insegnanti

- ❖ siti in italiano e in inglese in continuo cambiamento e aggiornamento

Dove trovare esempi di esercitazioni

- ❖ Sperimentando (<http://www.sperimentando.com/?tag=bioinformatica>)
- ❖ CusMIbio (<http://www.cusmibio.unimi.it/sperimenta.html>)



i tutorial sul progetto di sequenziamento del genoma umano

- ❖ <https://www.genome.gov/25019885/online-education-kit-how-to-sequence-a-human-genome>



Esercitazione 2

- ❖ “Caccia al gene” (disponibile al seguente url: <http://www.cusmibio.unimi.it/scaricare/caccia.pdf>)

Esercitazione 3

- ❖ “Navigare sul DNA” (disponibile al seguente url:
[http://www.cusmibio.unimi.it/scaricare/
Navigare09.pdf](http://www.cusmibio.unimi.it/scaricare/Navigare09.pdf))