



BIG DATA

Il supporto dei satelliti all'epidemiologia spaziale

Paolo Calistri



Definizione di Big Data

Caratteristiche Fondamentali

I Big Data si caratterizzano per il volume, la varietà e la velocità dei dati, richiedendo tecnologie avanzate per la loro gestione e analisi.

Origine dei Dati

I dati provengono da fonti diverse, come sensori, social media e dispositivi genomici, contribuendo a una comprensione più profonda dei fenomeni biologici.

Impatto sulla Ricerca

L'analisi dei Big Data permette di scoprire pattern e correlazioni significative, rivoluzionando il modo in cui vengono condotte le ricerche nelle scienze «omiche».



BIG DATA

Biologia dei Sistemi e Scienze «Omiche»

Fabrizio Vitale



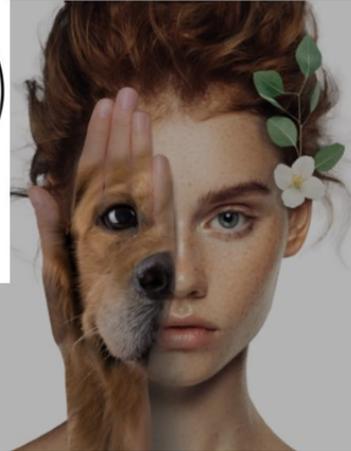
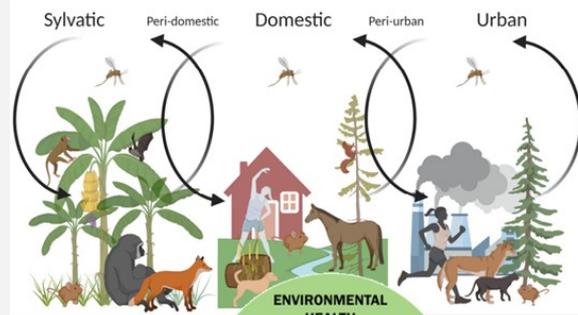
BIG DATA

Situazione epidemiologiche, dig data e sistemi informativi nella pratica quotidiana

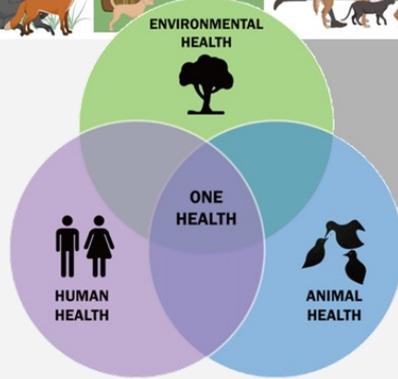
Stefano Cappai



One Health



è una strategia globale per la promozione di collaborazioni multisettoriali e transdisciplinari. Interconnessione e collaborazione tra salute umana, animale e ambientale.



L'approccio One Health richiede un atteggiamento **proattivo** in cui “ognuno fa la sua parte”



Seconda Giornata dei Centri e dei Laboratori di Referenza Nazionali degli Istituti Zooprofilattici Sperimentali nell'ottica One Health

*«Il mondo è nelle mani di coloro che hanno il coraggio di sognare e di correre il rischio di vivere i propri sogni»
(Paulo Coelho)*

Big Data: Biologia dei Sistemi e Scienze «Omiche»

...ovvero come l'integrazione dei Big Data nella biologia dei sistemi e nelle scienze «omiche» possa rivoluzionare la ricerca scientifica e migliorarne le applicazioni...

Roma 05 Novembre 2024

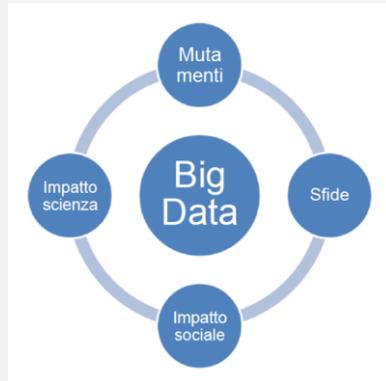


SIMeVeP
Società Italiana di Medicina
Veterinaria Preventiva

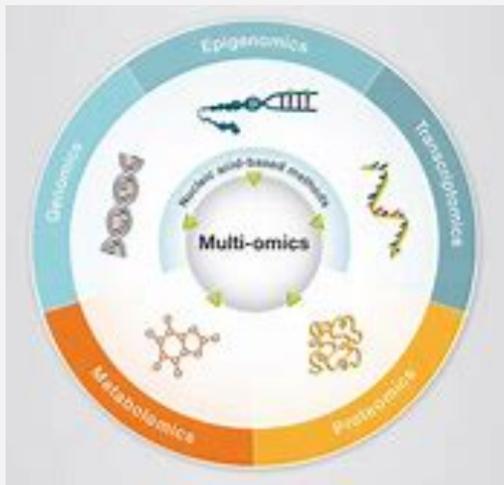


PANORAMICA

1. Gli strumenti di **analisi predittiva** sono i nuovi protagonisti della scienza medica che utilizza i Big Data
2. **Biologia dei sistemi** e scienze “omiche”: un nuovo paradigma nello studio degli organismi viventi
3. **Sistemi Complessi**
4. Nuove frontiere nella prevenzione e nella gestione delle **Malattie Infettive**.



La rivoluzione omica: dalle Biotecnologie alla Metagenomica



Integrazione delle Scienze Omiche Sinergia tra Discipline

L'integrazione delle scienze omiche, come genomica, proteomica e metabolomica, consente di ottenere una visione complessiva e interconnessa dei processi biologici, facilitando scoperte innovative e applicazioni cliniche più efficaci.

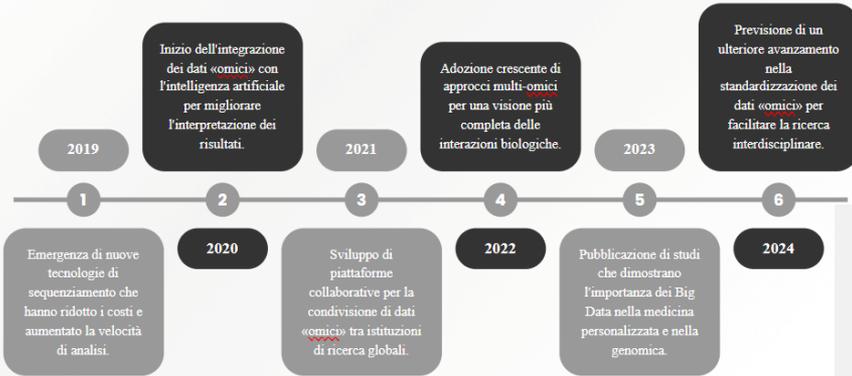


- ✓ **GENOMICA**: i geni contenuti nel DNA (genomica) e le loro molteplici funzioni (genomica funzionale);
- ✓ **EPIGENOMICA**: le modifiche epigenetiche presenti nel materiale genetico di una cellula (l'epigenoma);
- ✓ **TRASCRIPTOMICA**: un gran numero o di tutti i trascritti (trascrittoma);
- ✓ **PROTEOMICA**: il complesso delle proteine, in particolare delle loro strutture e funzioni;
- ✓ **METABOLOMICA**: tutti i metaboliti di un organismo biologico, che sono i prodotti finali del suo metabolismo.

Questo cambiamento è dovuto alla maggiore diffusione dei sequenziamenti di acidi nucleici NGS (Next Generation Sequencing) ad alta processività e direttamente da un campione ambientale senza alcuna conoscenza preliminare delle specie presenti.



Evoluzione delle Scienze Omiche



Esempi di Applicazioni Pratiche

01

Analisi Genomica Avanzata

L'uso dei Big Data nella genomica ha permesso di identificare varianti genetiche associate a malattie, migliorando la diagnosi e il trattamento personalizzato dei pazienti.

02

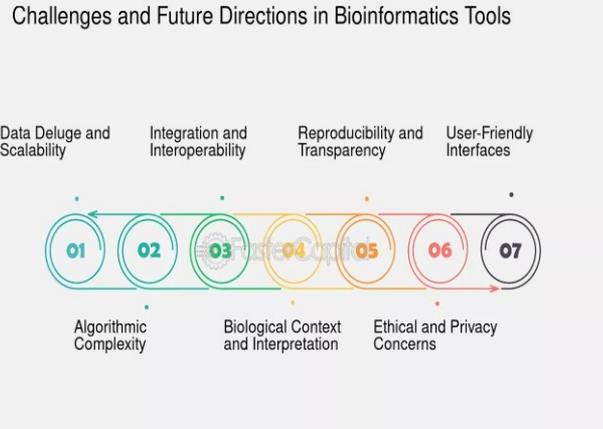
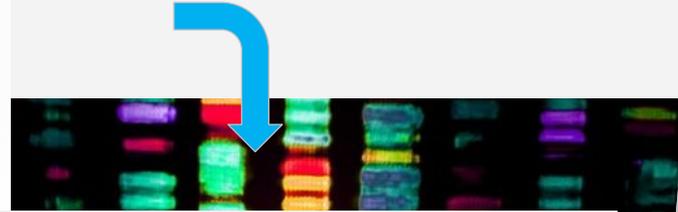
Scoperta Di Farmaci

Le tecnologie «omiche» integrate con l'analisi dei Big Data hanno accelerato il processo di scoperta di nuovi farmaci, consentendo la simulazione di interazioni molecolari su larga scala.

03

Monitoraggio Della Salute Pubblica

L'analisi dei dati provenienti da fonti diverse, come social media e registri sanitari, ha facilitato il monitoraggio delle epidemie e la risposta rapida a crisi sanitarie globali.



TOWARDS THE FUTURE WE WANT

Cambiare punto di vista: perché ne abbiamo bisogno?

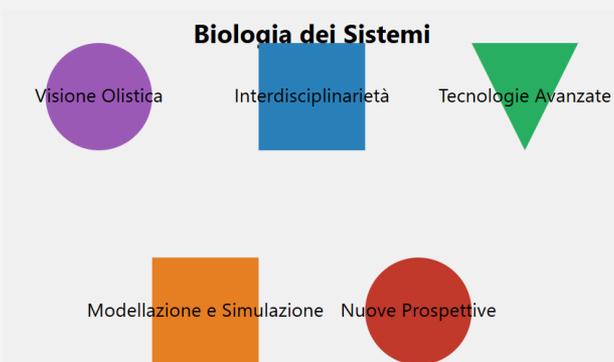
La ricerca biomedica ha avuto come obiettivo principale quello di identificare i meccanismi molecolari alla base delle patologie concentrata sui dettagli dell'interazione tra un enzima e il suo substrato oppure sull'azione di una proteina in una specifica via di segnalazione.

...non esistono geni "buoni" o geni "cattivi", ma solo reti che esistono a diversi livelli, sono connesse in vario modo, e rispondono in modo diverso alle perturbazioni...

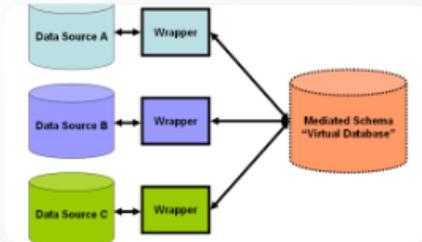
cit. Craig Venter

La biologia dei sistemi: una rivoluzione nei paradigmi e nei metodi della ricerca biologica

- 1. Visione olistica:** Invece di scomporre gli organismi nei loro singoli elementi (geni, proteine, etc.), la biologia dei sistemi li studia nella loro interezza, analizzando le interazioni e le dinamiche emergenti a livello sistemico.
- 2. Approccio interdisciplinare:** La biologia dei sistemi richiede l'integrazione di diverse discipline, come biologia, informatica, ingegneria, matematica e fisica, per comprendere la complessità dei sistemi biologici.
- 3. Tecnologie avanzate:** L'avvento di tecnologie come l'analisi high-throughput, la bioinformatica e la modellazione computazionale ha permesso di raccogliere ed elaborare enormi quantità di dati, essenziali per l'analisi sistemica.
- 4. Modellazione e simulazione:** Anziché basarsi solo su esperimenti e osservazioni, la biologia dei sistemi utilizza modelli matematici e simulazioni computerizzate per prevedere e testare il comportamento dei sistemi biologici.
- 5. Nuove prospettive:** Questo approccio olistico ha portato a nuove prospettive, come la comprensione di meccanismi di regolazione genica, di segnalazione cellulare e di malattie complesse.

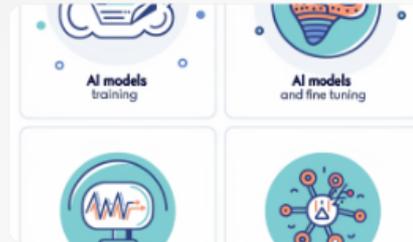


Importanza della Biologia dei Sistemi



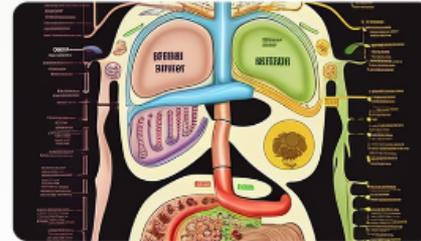
Integrazione dei Dati

La biologia dei sistemi consente di integrare dati provenienti da diverse fonti «omiche», facilitando una comprensione olistica delle interazioni biologiche e dei meccanismi cellulari.



Modelli Predittivi

Utilizzando approcci computazionali, la biologia dei sistemi sviluppa modelli predittivi che possono anticipare risposte biologiche a trattamenti, migliorando l'efficacia della medicina personalizzata.



Innovazione nella Ricerca

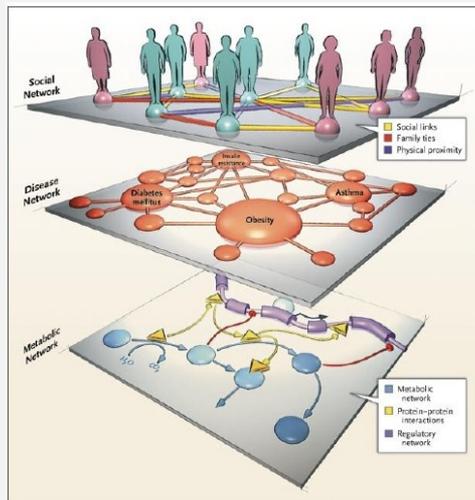
La biologia dei sistemi stimola l'innovazione nella ricerca scientifica, promuovendo nuove scoperte e applicazioni pratiche attraverso l'analisi avanzata dei Big Data e delle scienze «omiche».



La suddivisione tra diversi tipi di rete è utile per analizzare uno specifico gruppo di biomolecole, ma **l'organismo vivente** è un sistema complesso, le cui condizioni fisiologiche o patologiche possono essere descritte solo tenendo conto di **molte reti che si influenzano a vicenda**. Lo stesso vale per lo studio delle malattie complesse, la cui evoluzione dipende da reti che si trovano a diversi livelli di organizzazione e i cui effetti si stratificano nel tempo.

Alla base di questo **modello a "millefoglie"** troviamo la rete la rete di **regolazione dell'espressione genica** che connette geni e RNA messaggeri; gli effetti di questa rete si intersecano con la **rete di interazioni proteiche** e con la **rete metabolica** che regola le reazioni biochimiche della cellula.

Fuoriuscendo dai confini cellulari, approdiamo alle **reti di comunicazione tra cellule e di trasduzione del segnale**, cui afferiscono per esempio i meccanismi di segnalazione endocrina tra tessuti e organi diversi. Se ci spingiamo oltre il singolo individuo, troviamo la **rete delle esposizioni ambientali** e quella delle **relazioni sociali**, che svolgono un ruolo chiave



(Fonte:
Albert-László Barabási, *N Engl J Med* 2007;
357:404-407; DOI: 10.1056/NEJMe078114)

Le **scienze omiche** sono il principale bacino da cui la *network medicine* attinge i dati da analizzare. Grazie all'enorme mole di dati raccolti tramite esperimenti *high throughput*, oggi la ricerca biomedica può fotografare lo stato molecolare di una cellula o di un tessuto catalogandone l'insieme di geni (**genomica**), RNA (**trascrittomica**), proteine (**proteomica**) e sostanze metaboliche (**metabolomica**).

A questi si aggiungono anche altri set di dati, come quelli dell'**esposoma**, che mappa l'effetto dell'esposizione a diversi fattori ambientali (alimentazione, radiazioni ecc.).

La medicina delle reti fornisce gli strumenti per studiare l'intera impalcatura molecolare su cui si regge il meccanismo patogenetico.



Strumenti di Analisi dei Dati

01

Software di Analisi

Strumenti come R e Python offrono librerie avanzate per l'analisi statistica e la visualizzazione dei dati, essenziali per interpretare grandi volumi di informazioni biologiche.

02

Piattaforme di Big Data

Tecnologie come Hadoop e Spark consentono l'elaborazione distribuita dei dati, facilitando l'analisi di dataset complessi provenienti da studi «omici».

03

Strumenti di Machine Learning

Algoritmi di apprendimento automatico sono utilizzati per identificare pattern nei dati biologici, migliorando la predizione e la classificazione in ambito biomedico.

L'applicazione di algoritmi di apprendimento automatico, come le reti neurali e gli alberi decisionali, consente di estrarre informazioni utili dai Big Data, migliorando la capacità di previsione e classificazione in ambito biologico.



Impatti dei Big Data sulla Ricerca Biologica

Trasformazione della Ricerca Biomedica

Rivoluzione dei Metodi di Ricerca

L'integrazione dei Big Data nella ricerca biomedica ha trasformato i metodi tradizionali, consentendo analisi più rapide e approfondite, migliorando la scoperta di biomarcatori e l'efficacia dei trattamenti personalizzati.



Integrazione dei Dati Clinici

L'uso dei Big Data consente di integrare informazioni cliniche, genomiche e ambientali, migliorando la personalizzazione dei trattamenti e l'efficacia delle terapie per i pazienti.



Analisi Predittiva

Le tecniche di analisi predittiva applicate ai Big Data permettono di anticipare le risposte ai farmaci, ottimizzando le strategie terapeutiche e riducendo gli effetti collaterali.



Sviluppo di Biomarcatori

I Big Data facilitano l'identificazione di biomarcatori specifici per malattie, contribuendo a diagnosi più accurate e a trattamenti mirati basati sul profilo genetico del paziente.

L'analisi dei Big Data ha portato a scoperte significative nel campo della biologia dei sistemi, come l'identificazione di nuovi biomarcatori per malattie complesse e la comprensione delle interazioni tra geni e ambiente, contribuendo a sviluppare terapie personalizzate più efficaci.

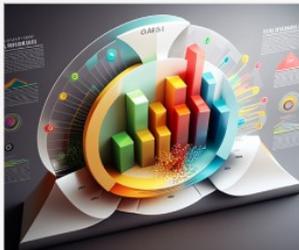


Progetti di Ricerca Rilevanti



Progetto Genomica Integrata

Studio che utilizza Big Data per analizzare interazioni tra geni e ambiente, migliorando la comprensione delle malattie complesse e sviluppando terapie personalizzate.



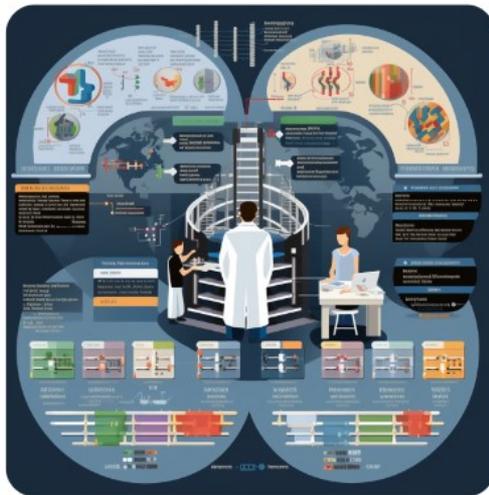
Analisi Multi-Omica

Iniziativa che integra dati genomici, proteomici e metabolomici per ottenere una visione olistica delle malattie, facilitando scoperte innovative nel campo della biomedicina.



Piattaforma di Condivisione Dati

Progetto di collaborazione...

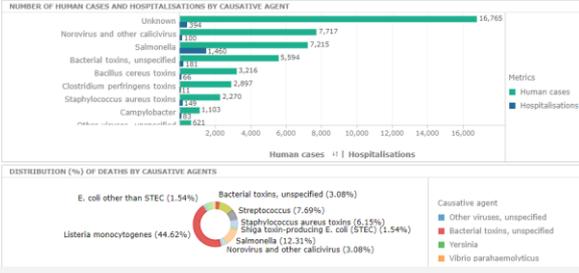
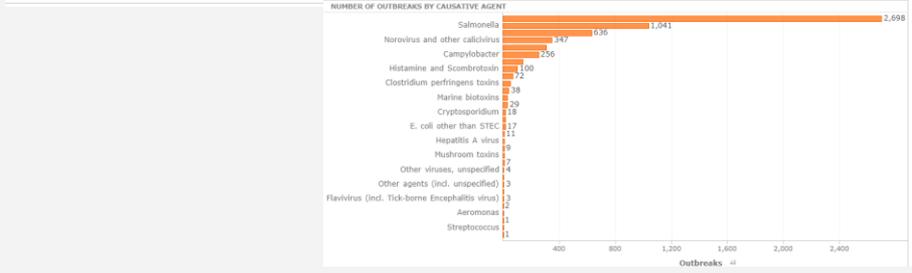
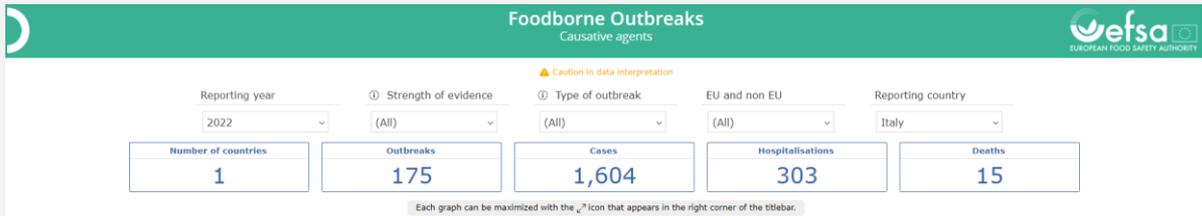


Rete di Ricerca Globale

Le collaborazioni internazionali nel campo dei Big Data e delle scienze «omiche» favoriscono la condivisione di risorse, competenze e dati, accelerando l'innovazione scientifica e migliorando la comprensione delle malattie a livello globale.

Collaborazioni Internazionali

Il carattere transnazionale dei focolai di infezione alimentari, la sorveglianza molecolare e le banche dati

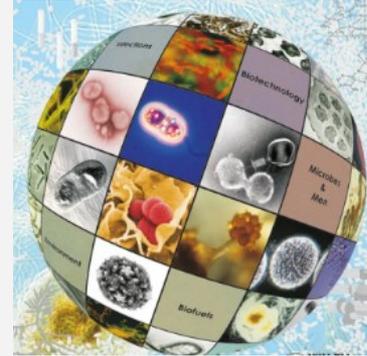


Il carattere transnazionale dei focolai di infezione alimentari, la sorveglianza molecolare e le banche dati

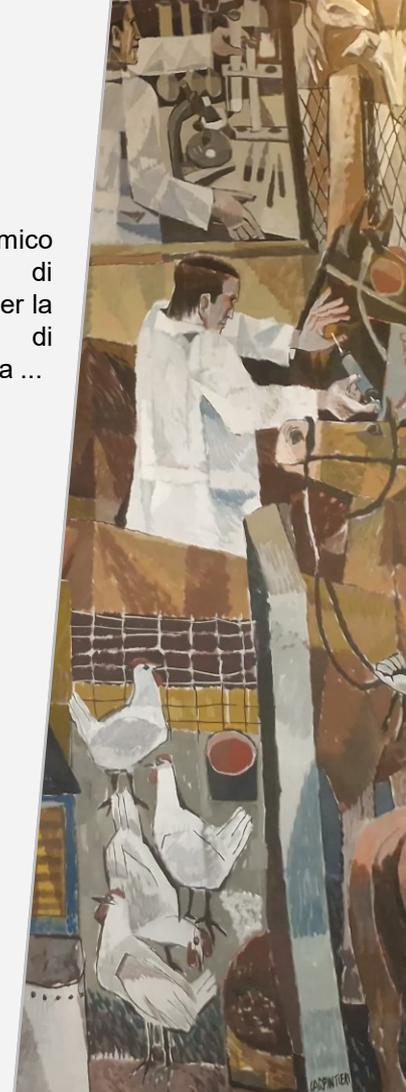
La complessità delle reti transnazionali di produzione e distribuzione determinano la possibilità che i patogeni possano diffondersi sfruttando le molteplici vie di propagazione offerte dalle reti dei flussi, presentandosi come «clusters» di uno stesso patogeno con ampia distribuzione geografica e temporale e dunque interessare località distanti tra loro anche migliaia di chilometri.



Il database epidemiologico genomico per l'identificazione globale di microrganismi è una piattaforma per la memorizzazione di dati di sequenziamento dell'intero genoma ...



Soluzioni avanzate per la gestione dei sistemi di tracciabilità e l'attivazione dei sistemi di sorveglianza epidemiologica che sfruttando le tecnologie molecolari (epidemiologia molecolare) riescono a fornire risposte per la rapida identificazione dei patogeni e loro localizzazione spazio-temporale soprattutto nelle fasi iniziali degli episodi di infezione



La Comunicazione Globale

La capacità di scambiare e gestire grandi quantità di dati del DNA utilizzando sistemi basati sul Web è aumentata in modo sorprendente negli ultimi anni. Ciò ha consentito la creazione di banche dati globali del DNA dei diversi ceppi batterici.

Global Microbial Identifier GMI

Sequenziamento e Condivisione dei genomi batterici e virali.

Principale obiettivo è la creazione di una piattaforma informatica (database) all'interno della quale vengono memorizzate le sequenze genomiche (WGS) di microrganismi patogeni (libreria genomica) utili per l'identificazione di geni rilevanti e per il confronto dei patogeni emergenti o responsabili di focolai infettivi..



La Veterinaria Pubblica

Cambiamenti climatici, globalizzazione, aumento demografico, incremento della produzione alimentare ecc. possono essere ritenuti, in modo più o meno diretto e con un diverso peso, importanti **driver** che condizionano l'emergenza e la diffusione dei contaminanti nelle catene alimentari. Si stima che un terzo di tutte le infezioni umane riconoscono come serbatoio il mondo animale e che tale associazione sia aumentata costantemente negli ultimi 40 anni. I dati epidemiologici ci dicono che negli ultimi due decenni il 75% delle malattie emergenti sono state zoonosi e il rischio di infezioni zoonosiche aumenterà in futuro.



Al mondo veterinario viene riconosciuto un **ruolo chiave** in sanità pubblica per la tutela sanitaria delle popolazioni animali e quindi delle produzioni alimentari e per la sicurezza dei consumatori attraverso le attività di controllo ufficiale e prevenzione del rischio delle relative filiere zootecniche. I nuovi rischi sanitari legati agli alimenti e quelli legati ai processi biotecnologici devono spingere a definire nuove competenze e criteri gestionali attraverso una impostazione più scientifica dei piani dei controlli e la formulazione di strategie sanitarie basate sui risultati della valutazione del rischio e sulla conoscenza della distribuzione dei fattori di rischio.

I rischi emergenti o i rischi sistemici derivanti dal processo di globalizzazione possono essere gestiti in modo efficace solo adottando una metodologia ispettiva *science-based* e con alto valore predittivo, in combinazione con i nuovi sistemi informatici per lo scambio rapido di informazioni secondo un modello di integrazione interprofessionale.

La Sanità pubblica veterinaria dovrà dotarsi di nuove competenze, di nuovi modelli organizzativo-gestionali, di nuove modalità di integrazione che superino l'attuale organizzazione all'interno del Servizio sanitario nazionale secondo una logica di innovazione del sistema *future-oriented*.



L'era del Data Mining

One Health promuove la **conoscenza** e la **responsabilità**.

La **conoscenza** si espande attraverso lo scambio di dati e informazioni.

La **responsabilità** dipende dalla trasparenza delle informazioni; la mancanza di trasparenza porta a pregiudizi.

La trasparenza dei dati richiede la **sinergia** dei sistemi e la **condivisione** tra sistemi.

Big data per una salute in grado di mettere in sinergia diversi sistemi:

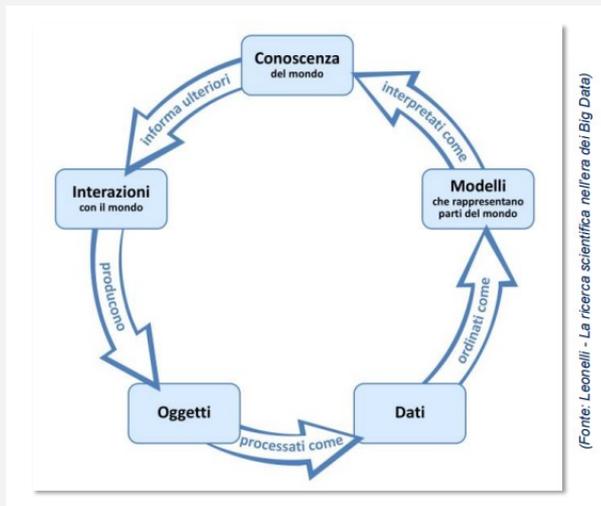
- *Google (Maps, Trends, Translate);*
- *immagini satellitari;*
- *applicazione della geostatistica;*
- *habitat e migrazione degli uccelli, degli animali e delle specie marine;*
- *agricoltura;*
- *cambiamento climatico;*
- *gestione del rischio climatico;*
- *temperatura; pioggia; inondazioni; cicloni; uso del suolo; copertura forestale; raccolta differenziata; qualità dell'aria; gestione dei rifiuti, ecc.;*
- *l'integrazione di WAHID-WOAH, EMPRES-FAO e WHO.*



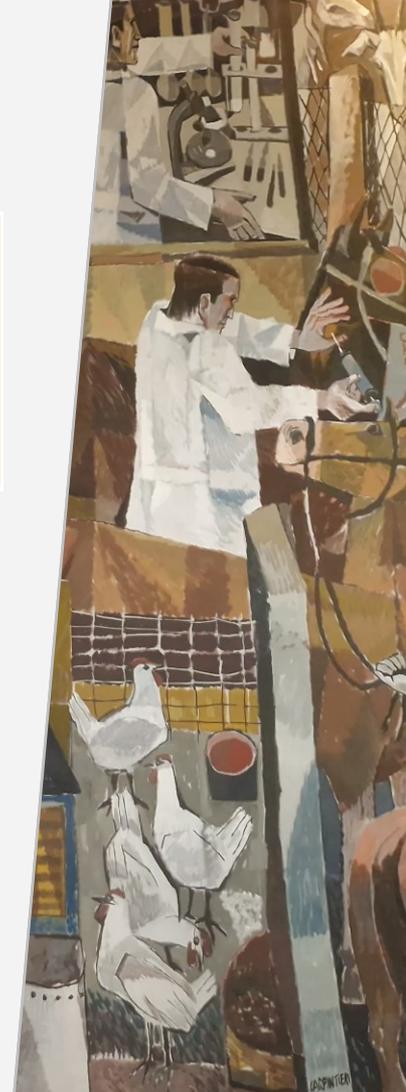
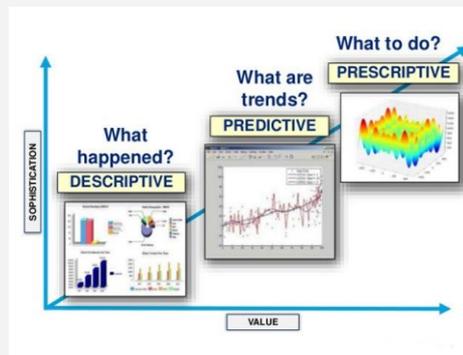
L'organizzazione Data Driven

L'enfasi sui big data è aumentata. Con il basso costo della generazione dei dati, la gestione intelligente dei dati può contenere informazioni più utili.

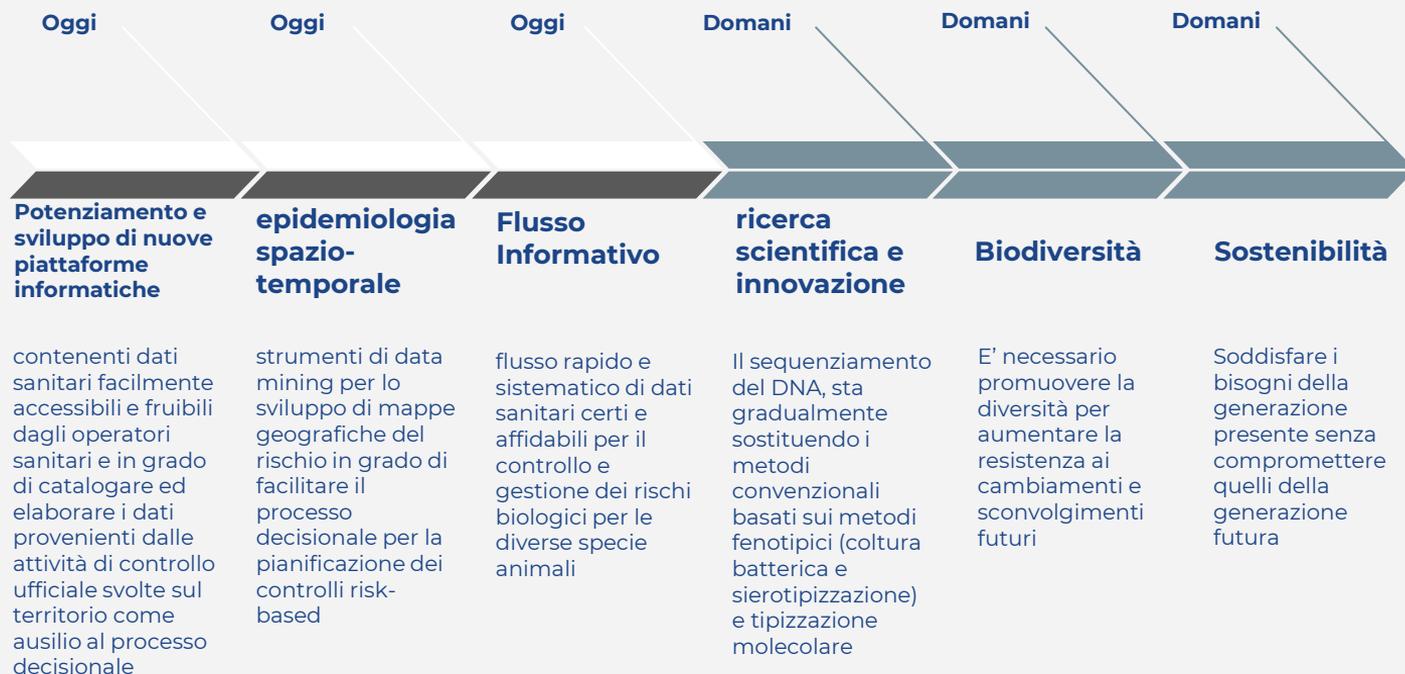
I big data offrono nuove intuizioni su reti, dinamiche spaziali e temporali per comprendere i sistemi umani, animali e ambientali a livello sistemico e per rilevare interazioni e non linearità tra variabili



Un'organizzazione Data-Driven è quella che comprende l'importanza di utilizzare i dati per prendere ogni tipo di decisione, è focalizzata nel renderli disponibili a chi ne ha bisogno quando ne ha bisogno e favorisce la collaborazione e l'interfunzionalità.



Gli orientamenti strategici futuri a cui la veterinaria pubblica dovrà guardare per assicurare la sostenibilità, la credibilità e l'efficacia delle azioni future possono comprendere attività strategiche riassumibili nei seguenti domini:





Innovazioni e Sostenibilità

L'integrazione dei Big Data nella ricerca biomedica promette di rivoluzionare la scoperta scientifica, migliorando l'efficienza dei processi di ricerca e promuovendo approcci sostenibili attraverso l'ottimizzazione delle risorse e la riduzione dei costi associati alla sperimentazione.

Futuro della Ricerca con i Big Data



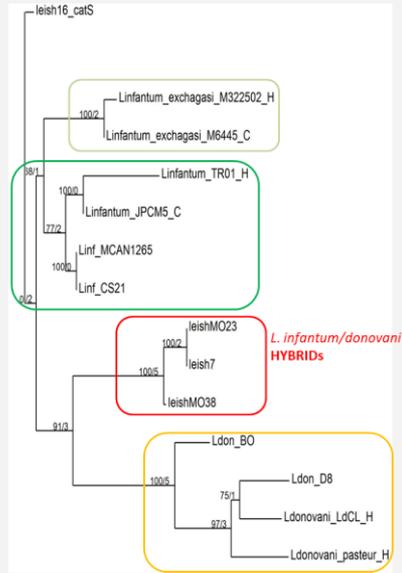
Genomic and epidemiological evidence for the emergence of a *L. infantum*/*L. donovani* hybrid with unusual epidemiology in northern Italy

F. Bruno,¹ G. Castelli,¹ B. Li,² S. Reale,¹ E. Carra,³ F. Vitale,¹ S. Scibetta,¹ M. Calzolari,⁴ S. Varani,⁴ M. Ortali,^{4,5} E. Franceschini,⁴ W. Gennari,⁶ G. Rugna,⁷ G. F. Späh⁸

AUTHOR AFFILIATIONS

- ¹WOAH Leishmania Reference Laboratory, Istituto Zooprofilattico Sperimentale della Sicilia, Centro di Referenza Nazionale per le Leishmaniosi (C.Re.Na.L.), Palermo, Italy
- ²Bioinformatics and Biostatistics Hub, Institut Pasteur, Université Paris Cité, Paris, France
- ³Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna "B. Ubertini", Brescia, Italy
- ⁴Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy
- ⁵IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy
- ⁶Infectious Disease Unit, Azienda Ospedaliera Universitaria di Modena, Modena, Italy
- ⁷Virology and Molecular Microbiology Unit, University Hospital of Modena, Modena, Italy
- ⁸Unité de Parasitologie moléculaire et Signalisation, INSERM U1201, Institut Pasteur, Université Paris Cité, Paris, France

Lo studio combina approcci di genomica comparativa e epidemiologia molecolare per analizzare la diversità genetica e l'eterozigosi all'interno della specie *Leishmania infantum* in Italia, importante sia per la salute pubblica che per la veterinaria.



Le analisi genomiche rivelano una diversità significativa all'interno del gruppo MON-1 di *L. infantum*, indicando un certo isolamento geografico, ma anche una complessità nelle interazioni ecologiche e nelle dinamiche di trasmissione.

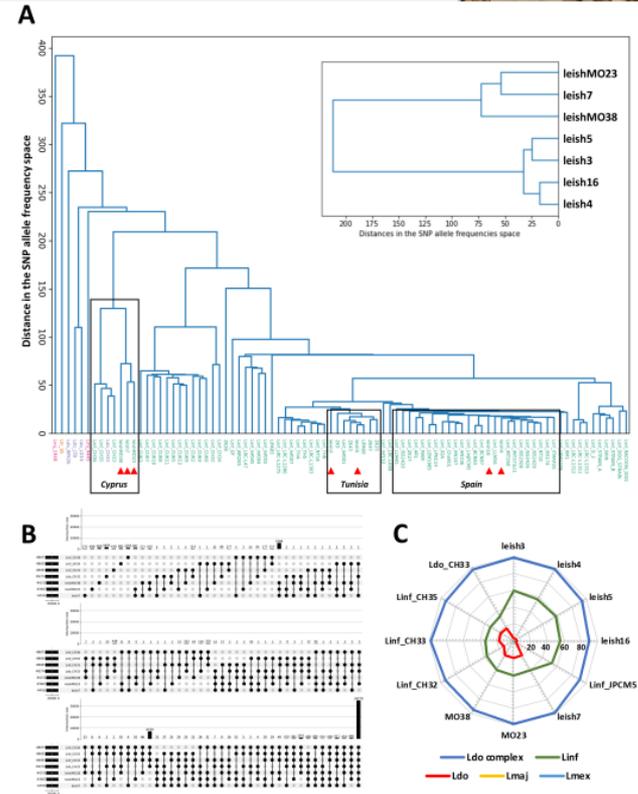


FIG 4 Putative origin of the leish7 hybrid based on SNP analysis. (A) Cluster analysis based on pairwise Euclidean distances between all-allele frequencies across SNPs. The tree is constructed from these distances using the UPGMA method. On the main tree, sample names are colored according to their recorded species. The samples that were not sequenced in the present study have their names prefixed by a tag indicating the species: Lmj (*major*), Ltr (*tropicus*), Ldo (*donovani*), or Linf (*infantum*). The samples that were sequenced in the present study are indicated by a red triangle. The top right inset shows the cluster analysis of two additional hybrid strains (leishMO23 and leishMO38) in comparison to the other strains sequenced in this study. (B) Bar plot of SNPs identified in the different strains indicated. The bars visualize the number of SNPs common to a given combination of samples. The numbers in the bar plot indicate shared SNPs for the given combination. The numbers on the left side of the strain ID indicate the total number of SNPs in a given isolate. Note that the upset plot continues in the lower panels. (C) Radar plot summarizing the results of the Kraken analysis. Along each radial axis, the distance from the center indicates the percentage of reads unambiguously assigned to a given taxon (color encoded) for the corresponding sample. Notable amounts of reads assigned to *L. donovani* (red) are only observed in the (putative) hybrids.



WOAH Reference Laboratory for Leishmaniasis



Risultati Principali:

- Eterogeneità Genetica:** Sono state identificate diverse isolate di *L. infantum* in Italia, correlate a genotipi di parassiti di diverse origini geografiche, incluso un genotipo ibrido descritto in precedenza a Cipro.
- Origini Regionali:** L'analisi suggerisce che l'Italia si trova all'incrocio delle infezioni da *L. infantum* nel Mediterraneo, evidenziando un legame genetico tra isolati della Sardegna e Sicilia con quelli della Spagna e Tunisia.
- Popolazioni Distinte:** Le isolate denominate leish3 e leish5 appartengono alla popolazione MON-1 del Nord Africa, differente dalla popolazione europea MON-1, suggerendo una complessità nella distribuzione geografica dei vari ceppi.
- Epidemiologia Complessa:** La diffusione del parassita è influenzata da dinamiche di trasmissione, distribuzione dei vettori, esposizione delle popolazioni umane, parametri ambientali e presenza di ospiti serbatoi. In Italia, i cani sono i principali serbatoi del parassita, anche se altri mammiferi possono giocare un ruolo.

Conclusioni:

L'Italia rappresenta un nodo centrale per lo studio delle infezioni da *L. infantum* nel contesto mediterraneo, riflettendo storie di migrazione e interazioni ecologiche nel tempo. Queste conclusioni sottolineano l'importanza di monitorare e studiare la diversità genetica dei ceppi di *L. infantum* e i potenziali ibridi per comprendere meglio la diffusione della leishmaniosi in Italia, specialmente alla luce delle recenti segnalazioni nel nord del paese.

Messaggi Chiave:

- Strain Ibridi Potenziali:** Sono state identificate ceppi ibridi (leish7, leishMO23, leishMO38) nella regione RER, potenzialmente collegati alla riemersione della leishmaniosi viscerale (VL) in quest'area.
- Espansione della Leishmaniosi:** Sebbene la leishmaniosi umana e canina sia stata prevalentemente segnalata nel centro e sud Italia, si osserva una recente diffusione anche nel nord, inclusa la regione RER.
- Analisi Genomica:** Le analisi di Multilocus Microsatellite Typing (MLMT) e di sequenziamento dell'intero genoma hanno rivelato la trasmissione di popolazioni ibride in RER, correlate a ceppi divergenti di *L. infantum* originariamente descritti a Cipro, con 34,778 SNP condivisi (74% del totale).
- Strain Cipro:** I ceppi ciprioti, come il ceppo CH33, sono stati precedentemente classificati come un nuovo gruppo di *L. donovani* sensu lato, suggerendo una linea evolutiva distintiva o un ibrido antico tra *L. infantum* e *L. donovani*.



WOAH Reference Laboratory
for Leishmaniasis

Reference Centre  World Organisation
for Animal Health
Founded on OIE



Grazie!

Fabrizio Vitale

*Direttore Area Biologia Molecolare
Responsabile Centro di Referenza Nazionale Leishmaniosi
(CRENAL)
WOAH Leishmania Reference Lab. Expert
Responsabile Laboratorio Entomologia e Controllo Vettori
Ambientale (EVA)
Istituto Zooprofilattico Sperimentale della Sicilia
via Gino Marinuzzi 3 - 90129 – Palermo
fabrizio.vitale@izssicilia.it
+390916565368
+393357895724*



Istituto Zooprofilattico Sperimentale Sicilia "A. Mirri"

WOAH Reference Laboratory
for Leishmaniasis

Reference Centre  World Organisation
for Animal Health
Founded in 1967

