

IRIDA-ARIES, UNA RISORSA CHIAVE IN ITALIA PER LA SORVEGLIANZA GENOMICA ONE HEALTH DELLE MALATTIE INFETTIVE E IL SUO IMPIEGO NELLA SORVEGLIANZA DELLE MALATTIE A TRASMISSIONE ALIMENTARE



Arnold Knijn, Valeria Michelacci, Federica Gigliucci, Rosangela Tozzoli, Paola Chiani, Fabio Minelli, Gaia Scavia, Eleonora Ventola, Alfonsina Fiore, Antonietta Gattuso, Marco Francesco Ortoffi, Gianni Ciccaglioni, Umberto Agrimi e Stefano Morabito in collaborazione con il Gruppo utenti di IRIDA-ARIES STEC*, IRIDA-ARIES Listeriosi* e il Registro Italiano Sindrome Emolitico Uremica*
Dipartimento di Sicurezza Alimentare, Nutrizione, Sanità Pubblica Veterinaria, ISS

RIASSUNTO - La genomica dei microrganismi patogeni sta trasformando la sorveglianza delle malattie infettive, approfondendo la nostra comprensione sull'evoluzione e sulla diffusione degli agenti eziologici, sulle interazioni ospite-patogeno e la resistenza antimicrobica. L'Istituto Superiore di Sanità (ISS) ha sviluppato un sistema informativo per la raccolta, l'analisi e la condivisione di dati genomici a supporto della sorveglianza delle malattie infettive e delle epidemie di origine alimentare. La piattaforma, IRIDA ARIES, è dotata di algoritmi di identificazione di cluster e di sistemi di allerta automatica e fornisce una console per monitorare la circolazione degli agenti infettivi a livello regionale e al contempo rappresentando un osservatorio in tempo reale delle dinamiche delle malattie infettive a livello nazionale. La piattaforma IRIDA-ARIES ospita il sistema nazionale di sorveglianza delle infezioni da *Listeria monocytogenes* ed *Escherichia coli* produttore della tossina Shiga.

Parole chiave: sorveglianza genomica; malattie a trasmissione alimentare; gestione epidemica

SUMMARY (IRIDA-ARIES, a key resource in Italy for One Health genomic surveillance of infectious diseases and its use for surveillance of foodborne diseases) - Genomics of pathogenic microorganisms is transforming surveillance of infectious diseases, deepening our understanding of evolution and diffusion of etiological agents, host-pathogen interactions, and antimicrobial resistance. The National Institute of Health in Italy (ISS) has developed an Information System for the collection analysis and sharing of genomic data to support surveillance of infectious epidemics and foodborne outbreaks. The platform, IRIDA ARIES, is equipped with cluster detection algorithms and automatic alerts systems and provides a console to monitor the circulation of infectious agents at regional level and at the same time represents a real-time observatory of the dynamics of infectious diseases at national level. The IRIDA-ARIES platform hosts the Italian national surveillance system for *Listeria monocytogenes* and Shiga toxin-producing *Escherichia coli* infections.

Key words: genomic surveillance; foodborne diseases; cluster investigation

stefano.morabito@iss.it

La diffusione del sequenziamento genomico (Whole Genome Sequencing - WGS) dei microrganismi patogeni ha migliorato la sorveglianza delle malattie infettive aumentandone la risoluzione in termini di identificazione precoce di focolai, nonché la comprensione dell'evoluzione e della diffusione degli agenti eziologici, delle interazioni ospite-patogeno e della resistenza antimicrobica.

Il WGS, inoltre, offre un'eccellente opportunità per l'applicazione dell'approccio *One Health* (1, 2) attraverso l'integrazione dei dati provenienti dai diversi domini (umano, veterinario, alimentare e ambientale).

Nel 2014 l'Autorità per la Sicurezza Alimentare (European Food Safety Authority, EFSA) e il Centro Europeo per il Controllo delle Malattie (European Centre for Disease Prevention and Control, ECDC) hanno avviato un progetto pilota congiunto sulla caratterizzazione molecolare degli agenti di malattie a trasmissione alimentare (MTA) isolati da alimenti e da casi di malattia nell'uomo. Il progetto è culminato nel 2023, con il rilascio del sistema "One Health WGS", focalizzato su tre agenti di MTA prioritari a livello europeo: 1) *Escherichia coli* produttori di Shiga tossine (STEC); 2) *Listeria monocytogenes*; 3) Salmonella.

(*) L'elenco del Gruppo utenti di IRIDA-ARIES STEC, IRIDA-ARIES Listeriosi e del Registro Italiano Sindrome Emolitico Uremica è riportato a p. 12.

L'Istituto Superiore di Sanità (ISS) ha sviluppato, già dal 2018, una piattaforma per la sorveglianza genomica delle malattie infettive denominata IRIDA-ARIES. Questa infrastruttura è stata concepita con lo scopo di fornire alle strutture del Servizio Sanitario Nazionale (SSN) una piattaforma per la raccolta, l'analisi e la condivisione delle informazioni di sequenziamento genomico dei microrganismi patogeni alimentari in supporto alla sorveglianza. Poiché nel nostro Paese l'SSN è organizzato e gestito a livello regionale, la piattaforma è stata disegnata con lo scopo di facilitare la diffusione dell'informazione sia in senso orizzontale (all'interno del territorio regionale) tra laboratori, ospedali e autorità sanitarie locali, sia verticale tra Regioni diverse che con l'autorità sanitaria centrale. In questo modo, la piattaforma costituisce uno strumento di aggregazione, sintesi e analisi a livello nazionale, capace di restituire e condividere con le Regioni tutte le informazioni utili e quindi in grado di integrare i dati anche a livello territoriale.

La piattaforma IRIDA-ARIES

Il sistema informativo che supporta la sorveglianza genomica nella piattaforma IRIDA-ARIES è composto da due soluzioni software open-source: Galaxy-ARIES (Advanced Research Infrastructure for Experimentation in genomics) (3, 4) e IRIDA (Integrated Rapid Infectious Disease Analysis) (5) (Figura 1).

La piattaforma Galaxy-ARIES è stata pubblicata nel 2015 sui server dell'ISS con lo scopo di fornire strumenti bioinformatici per l'analisi di dati di sequenziamento genomico appositamente sviluppati per la microbiologia in sanità pubblica (<https://aries.iss.it>). Lo scopo principale era fornire uno spazio per queste analisi a operatori di laboratorio che non avessero le conoscenze informatiche necessarie per utilizzare i software di analisi bioinformatica. Oggi Galaxy-ARIES conta oltre 800 utenze in tutto il mondo distribuite tra istituti di sanità pubblica, enti di ricerca e università (Figura 2).

Per utilizzare la piattaforma Galaxy per la sorveglianza delle malattie infettive è stato necessario dotare ARIES di un'interfaccia per il caricamento dei dati e il lancio di analisi automatizzate. Inoltre, è stato sviluppato un database per raccogliere le informazioni derivanti dalle analisi in modo che potessero essere utilizzate per la sorveglianza e condivise tra utenti sul territorio nazionale. A questo scopo si è ricorso all'utilizzo di una versione ampiamente modificata del sof-

tware open-source IRIDA (<http://irida.ca>). Il software originario, inizialmente sviluppato per la ricerca, ha richiesto numerose modifiche per adattarne le funzionalità alle esigenze della sorveglianza e dell'SSN regionalizzato, ad esempio, la gestione di privilegi di accesso differenti a seconda della provenienza territoriale degli utenti e l'utilizzo di dati di sequenziamento prodotti da piattaforme di sequenziamento differenti (*paired ends*, *single ends*) e a diversi stadi di elaborazione (analisi di sequenza *raw reads and fasta*). Inoltre, IRIDA è stata dotata di algoritmi automatici per l'identificazione di cluster genomici e di un sistema di allerta automatico.

In Italia, ciascuna Regione/Provincia Autonoma (PA) gestisce i propri sistemi di sorveglianza in modo autonomo attraverso le aziende sanitarie locali (6). Per riflettere questa realtà, l'organizzazione di IRIDA-ARIES è stata sviluppata in progetti regionali che condividono parzialmente le informazioni con un progetto nazionale che risulta accessibile dalle Regioni, dall'ISS e dal Ministero della Salute allo scopo di analizzare la situazione a livello nazionale. I metadati sensibili, eventualmente presenti nei progetti regionali, non vengono condivisi nel progetto nazionale nel rispetto del Regolamento Generale sulla Protezione dei Dati (GDPR, UE n. 2016/679). In caso di cluster multiregionali gli utenti delle Regioni coinvolte possono visualizzare i contatti delle altre Regioni coinvolte nel cluster per stabilire interlocuzioni dirette finalizzate alle indagini epidemiologiche.

Il flusso informativo IRIDA-ARIES

Per popolare la piattaforma, sono necessarie due operazioni: i) creazione di un campione; ii) caricamento della/e sequenza/e.

Al momento della creazione di un campione, i metadati epidemiologici possono essere aggiunti manualmente o recuperati attraverso l'integrazione dei dati da fonti esterne, se disponibili, utilizzando protocolli di *record linkage* (Figura 1). In seguito al caricamento delle sequenze, viene avviato automaticamente un flusso di lavoro specifico per ogni patogeno che esegue elaborazioni di assemblaggio, controlli di qualità, tipizzazione inclusa l'identificazione dei determinanti di antibiotico-resistenza e il clustering basato sull'analisi del core-genome MLST (cgMLST) (7, 8).

Concluso il flusso di lavoro automatizzato, viene generata automaticamente e inviata agli utenti, che hanno caricato i dati, una e-mail che informa della conclusione dell'analisi e contiene un sommario ►

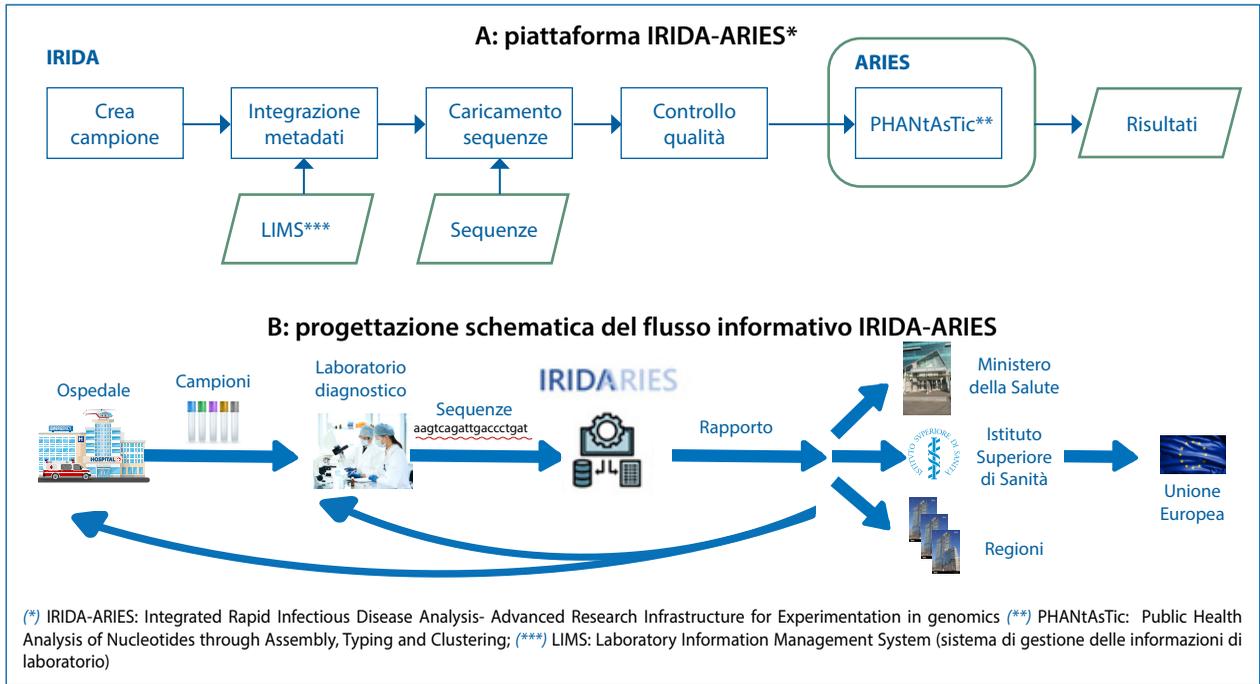


Figura 1 - A: quando si crea un campione, i metadati si recuperano sia dai sistemi informativi di dati molecolari che epidemiologici locali. Quando le sequenze genomiche vengono caricate, si esegue il controllo di qualità e si lanciano le analisi automatiche con la pipeline patogeno-specifica PHANtAsTic. I risultati vengono restituiti a IRIDA e visualizzati in una pagina web. **B:** il flusso viene attivato a livello locale da un centro che invia un campione di un caso a un laboratorio che esegue il sequenziamento dell'intero genoma. Le sequenze vengono caricate su IRIDA-ARIES e i risultati vengono restituiti a livello locale e inoltrati a livello regionale e nazionale. Se necessario, i dati possono essere portati a livello europeo



Figura 2 - . Distribuzione geografica degli utenti dell'istanza di Galaxy ARIES (<https://aries.iss.it>)

dei risultati dell'analisi di clustering (se il campione è parte di un cluster, cioè il suo profilo genetico è simile a quello di altri campioni entro una certa soglia di differenze) e della tipizzazione. In caso di identificazione di un cluster che coinvolge più Regioni, la comunicazione viene inviata dal sistema agli utenti delle altre Regioni coinvolte, al Ministero della Salute e all'ISS per garantire il coordinamento e la gestione di eventuali focolai, rappresentando una risorsa chiave del sistema di allerta rapido.

I flussi di lavoro di tipizzazione molecolare

Tutti i flussi di lavoro analitici sono stati progettati combinando software open-source già pubblicati e di nuovo sviluppo. Il flusso di lavoro PHANtAsTiC (Public Health Analysis of Nucleotides through Assembly, Typing and Clustering) è stato sviluppato per eseguire una serie di tipizzazione specifiche per ogni microrganismo. Questi includono la sierotipizzazione, la tipizzazione del tipo di MLST, la caratterizzazione del viruloma (insieme dei geni di virulenza) e dei determinanti di resistenza agli antibiotici e del profilo cgMLST (9). Un avviso viene attivato nel caso in cui venga identificato un cluster (la soglia è di 4 differenze alleliche in cgMLST per *L. monocytogenes* e 10 per STEC).

Il flusso di lavoro è stato adattato per riflettere la corrispondenza con il metodo di tipizzazione cgMLST eseguito a livello europeo dal "One Health WGS System" (10) di EFSA/ECDC.

Risultati

La piattaforma IRIDA-ARIES ospita attualmente il sistema nazionale di sorveglianza genomica delle infezioni da *L. monocytogenes* e il sistema nazionale di sorveglianza genomica *One Health* delle infezioni da *E. coli* produttore di STEC, nell'ambito delle MTA. Inoltre, nel 2021, a causa della necessità di tracciare le varianti del virus SARS-CoV2, la piattaforma è stata dotata di moduli specifici per la raccolta, analisi e condivisione dei dati di sequenziamento del genoma virale.

Nel contesto della sorveglianza delle MTA, il sistema conta oggi 71 utenti distribuiti tra il livello nazionale e regionale. Per quanto concerne la listeriosi, all'11 aprile 2024, sulla piattaforma sono presenti 1.911 sequenze genomiche di ceppi isolati da 1.713 casi di malattia insorti nel periodo 2002-2024, e da 70 campioni storici di animali/alimenti/mangimi. Infine,

sono stati caricati in piattaforma 128 campioni provenienti da eventi epidemici avvenuti in diversi Paesi europei condivisi attraverso la piattaforma di ECDC, EpiPulse (11) al fine di verificare la potenziale presenza di genomi italiani riferibili a focolai multinazionali.

Nell'ambito della sorveglianza genomica delle infezioni causate da ceppi STEC, all'11 aprile 2024, sono presenti sulla piattaforma IRIDA-ARIES un totale di 2.086 sequenze genomiche ottenute da campioni prelevati nel periodo 1989-2024, comprendenti 1.023 sequenze di ceppi isolati da casi di malattia nell'uomo, 968 da ceppi provenienti da animali/alimenti/mangimi e 95 provenienti da eventi epidemici condivisi attraverso la piattaforma EpiPulse (11).

La piattaforma IRIDA-ARIES è stata al centro delle indagini sui recenti focolai di listeriosi verificatisi in Italia. Nel 2022, gli algoritmi che operano sulla piattaforma hanno rilevato la presenza di due cluster di genomi di sequence type (ST) ST8 e ST155. Conseguentemente, è stato costituito dal Ministero della Salute un Gruppo di Lavoro composto dall'ISS, dagli Istituti Zooprofilattici Sperimentali, dal Laboratorio Nazionale di Riferimento per *L. monocytogenes* e dalle Regioni/PA coinvolte nel cluster. Il lavoro di questo Gruppo ha supportato l'indagine epidemiologica stabilendo una correlazione tra i casi clinici e il consumo di alcuni prodotti alimentari. Durante l'indagine, l'analisi dei profili cgMLST ha consentito la rapida identificazione di campioni appartenenti al cluster, restringendo il processo analitico. Le pipeline (un servizio o un insieme di azioni che elaborano i dati in sequenza) filogenetiche integrate nella piattaforma sono state utilizzate dal Gruppo di Lavoro per la redazione dei report periodici, nonché in autonomia dagli stessi utenti regionali. Come affermato dal sottosegretario al Ministero della Salute in un'interrogazione parlamentare: "L'attuale situazione legata alla Listeriosi è emersa grazie al lavoro del Ministero della Salute, attraverso la sorveglianza ordinaria e attraverso la banca dati IRIDA-ARIES dell'Istituto Superiore di Sanità, che ha permesso di verificare l'aumento dei casi umani su tutto il territorio nazionale" (12). ■

Dichiarazione sui conflitti di interesse

Gli autori dichiarano che non esiste alcun potenziale conflitto di interesse o alcuna relazione di natura finanziaria o personale con persone o con organizzazioni, che possano influenzare in modo inappropriato lo svolgimento e i risultati di questo lavoro.

Riferimenti bibliografici

1. One Health High-Level Expert Panel (OHHLEP), Adisasmito WB, Almuhairi S, et al. One Health: A new definition for a sustainable and healthy future. Dvorin JD, curatore. *PLoS Pathog* 2022;18:e1010537 (doi:10.1371/journal.ppat.1010537).
2. Dente MG, Riccardo F, Milano A, et al. A One Health-based Conceptual Framework for comprehensive and coordinated prevention and preparedness to health threats. *Int J Infect Diseases* 2022;116:S108-9 (doi:10.1016/j.ijid.2021.12.256).
3. Afgan E, Baker D, Batut B, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* 2018;46(W1):W537-44 (doi:10.1093/nar/gky379).
4. Knijn A, Michelacci V, Orsini M, et al. Advanced Research Infrastructure for Experimentation in genomicS (ARIES): a lustrum of Galaxy experience. *bioRxiv* 2020 (doi:10.1101/2020.05.14.095901).
5. Matthews TC, Bristow FR, Griffiths EJ, et al. The Integrated Rapid Infectious Disease Analysis (IRIDA) Platform. *bioRxiv* (doi:10.1101/381830).
6. Humboldt-Dachroeden S. A governance and coordination perspective - Sweden's and Italy's approaches to implementing One Health. *SSM - Qualitative Research in Health* 2022;2:100198 (doi: 10.1016/j.ssmqr.2022.100198).
7. Silva M, Machado MP, Silva DN, et al. chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. *Microb Genom* 2018;4(3):e000166 (doi:10.1099/mgen.0.000166).
8. Moura A, Criscuolo A, Pouseele H, et al. Whole genome-based population biology and epidemiological surveillance of *Listeria monocytogenes*. *Nat Microbiol* 2016;2:16185 (doi:10.1038/nmicrobiol.2016.185).
9. Zhou Z, Alikhan NF, Sergeant MJ, et al. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res* 2018(9);28:1395-404 (doi:10.1101/gr.232397.117).
10. European Food Safety Authority (EFSA), Costa G, Di Piazza G, et al. Guidelines for reporting Whole Genome Sequencing-based typing data through the EFSA One Health WGS System. EFSA Supporting Publications 2022:EN-7413 2022 (doi:10.2903/sp.efsa.2022.EN-7413).
11. European Centre for Disease Prevention and Control (ECDC). Launch of EpiPulse, a new portal to strengthen the prevention and control of infectious diseases; 2021 (<https://www.ecdc.europa.eu/en/news-events/launch-epipulse-new-portal-strengthen-prevention-and-control-infectious-diseases>).
12. Acts of the Italian Chamber of deputies 2023 (<https://aic.camera.it/aic/scheda.html?numero=5-00004&ramo=C&leg=19>).

Il codice del software sviluppato internamente utilizzato per questo studio può essere trovato nei seguenti repository GitHub:

IRIDA-ARIES: <https://github.com/aknijn/irida>

PHANtAsTiC: <https://github.com/aknijn/phantastic-galaxy>

TAKE HOME MESSAGES

- L'epidemiologia genomica rappresenta lo strumento più avanzato per il controllo delle malattie infettive.
- La piattaforma IRIDA-ARIES è un asset strategico per la sorveglianza delle malattie infettive e la gestione dei focolai in chiave *One Health*.
- La condivisione controllata dei dati a livello regionale e nazionale operata dalla piattaforma IRIDA-ARIES e la disponibilità di algoritmi automatici di identificazione e comunicazione dei cluster di casi di infezione, rappresentano strumenti innovativi per l'identificazione dei segnali e la gestione dei focolai.

(*) Gruppo utenti IRIDA-ARIES STEC, IRIDA-ARIES Listeriosi e Registro Italiano Sindrome Emolitico Uremica

Gruppo utenti IRIDA-ARIES STEC: Sara Arnaboldi, Giorgia Borriello, Teresa Bossù, Paola Chiani, Maria Chironna, Germana Colarusso, Maria Grazia Coppola, Maria Laura De Marchis, Paola De Santis, Virginia Filipello, Giuliano Garofolo, Alessandra Gazzola, Federica Gigliucci, Arnold Knijn, Daniela Loconsole, Nadia Losio, Sarah Lovari, Giulia Magagna, Valeria Michelacci, Stefano Morabito, Rubina Paradiso, Antonio Parisi, Anna Sallustio, Gaia Scavia, Michela Tilola, Rosangela Tozzoli, Eleonora Ventola.

Gruppo utenti IRIDA-ARIES Listeriosi: Richard Aschbacher, Giuliana Blasi, Giorgia Borriello, Teresa Bossù, Cesare Camma, Daniela Cecconi, Lisa Chenal, Maria Chironna, Veronica Cibir, Gianni Ciccaglioni, Germana Colarusso, Maria Grazia Coppola, Valeria Cosma, Mauro Cravero, Maria Laura De Marchis, Paola De Santis, Lucia Decastelli, Adriano Di Pasquale, Teresa Fasciana, Federica Ferraro, Alfonsina Fiore, Alessia Franco, Giuliano Garofolo, Laura Gasperetti, Antonietta Gattuso, Elisabetta Giacobazzi, Claudio Giacomazzi, Anna Giammanco, Federica Gigliucci, Maria Gori, Arnold Knijn, Daniela Loconsole, Daniela Lombardi, Sarah Lovari, Paola Marconi, Elisa Masi, Riccardo Mazzocca, Elena Mazzolini, Valeria Michelacci, Onofrio Mongelli, Stefano Morabito, Marina Morganti, Marco Francesco Ortoffi, Elisabetta Pagani, Barbara Palombo, Rubina Paradiso, Antonio Parisi, Monica Pitti, Francesco Pomilio, Stefano Pongolini, Erika Scaltriti, Gaia Scavia, Cristina Schellenberger, Elisabetta Tanzi, Rosangela Tozzoli, Francesco Vairo, Eleonora Ventola, Teresa Zaccaria, Cristina Zappetti.

Registro Italiano Sindrome Emolitico Uremica: Francesca Becherucci, Milena Brugnara, Roberta Camilla, Valentina Capone, Roberto Chimenz, Ciro Corrado, Alessandra Gianviti, Mario Giordano, Edoardo La Porta, Gabriele Malgieri, Andrea Pasini, Ilse Maria Ratsch, Fabrizio Pugliese, Enrico Vidal.