

Sorveglianza della sicurezza alimentare con MiSeq™ i100 Series

Rilevamento accurato dei patogeni di origine alimentare



Riduzione degli interventi manuali grazie alla preparazione semplificata delle librerie



Durata inferiore della corsa di sequenziamento e output più elevati



Assemblaggio affidabile di genomi batterici di alta qualità

Introduzione

L'utilizzo di metodi di laboratorio per sottotipizzare i batteri enterici, compresa la *salmonella*, è stato fondamentale per identificare potenziali epidemie e collegare i batteri che causano malattie a potenziali fonti di epidemie. Con l'aumento della risoluzione dei metodi di sottotipizzazione in laboratorio, è aumentata anche la precisione nel collegare la malattia a una fonte. Negli Stati Uniti, il collegamento di potenziali casi di focolai epidemici viene effettuato tramite PulseNet, una rete nazionale di sorveglianza molecolare per malattie enteriche a trasmissione alimentare, acquosa e correlate a One Health riscontrate in oltre 80 laboratori statali, locali e federali. La rete PulseNet impedisce ogni anno oltre 250.000 malattie causate da batteri enterici e consente di risparmiare mezzo miliardo di dollari in costi medici e perdita di produttività.^{1,2}

PulseNet ha iniziato a implementare il sequenziamento dell'intero genoma (WGS, Whole-Genome Sequencing) utilizzando gli strumenti Illumina all'inizio degli anni 2010 con partner federali presso l'Ente statunitense preposto alla tutela di alimenti e medicinali (FDA, Food and Drug Administration) e il Servizio di ispezione e sicurezza alimentare del Dipartimento dell'agricoltura degli Stati Uniti (USDA-FSIS, Department of Agriculture Food Safety and Inspection Service).³ Utilizzando il WGS, gli scienziati della sanità pubblica possono identificare potenziali epidemie individuando ceppi da casi geneticamente molto simili e associati a malattie che si verificano nello stesso periodo o area geografica.

Il WGS viene utilizzato anche per identificare le tendenze nei ceppi batterici che non erano precedentemente caratterizzati da metodi molecolari a risoluzione inferiore.^{4,5} Questi ceppi si sono ripresentati nel corso di periodi di tempo e hanno causato epidemie, sono emersi e hanno causato un aumento della malattia o hanno continuato a causare malattie nell'arco di mesi o anni. I Centers for Disease Control and Prevention (CDC) hanno iniziato a descrivere questi ceppi come ceppi ricorrenti, emergenti e **persistenti (REP)**. I ceppi REP possono essere identificati in base alla correlazione genetica e ad altre caratteristiche, tra cui resistenza multifarmaco, virulenza o altri marker genetici. Poiché solo circa il 10% dei batteri che causano malattie in PulseNet è associato a un'epidemia, l'utilizzo del WGS per identificare i ceppi REP può fornire l'opportunità di prevenire ulteriori malattie collegando ulteriori ceppi che possono condividere una fonte comune.

Questa nota sull'applicazione dimostra il rilevamento e la caratterizzazione accurati di un ceppo persistente di *Salmonella enterica* sierotipo Infantis multifarmaco-resistente (**REP JFX01**) utilizzando un efficiente flusso di lavoro NGS che integra Illumina DNA Prep, MiSeq i100 Series e strumenti di analisi di terze parti (**Figura 1**).



Figura 1: flusso di lavoro NGS end-to-end per il rilevamento della *salmonella*

Combinare la preparazione delle librerie Illumina con il sequenziamento su MiSeq i100 Series e l'analisi secondaria DRAGEN per una sorveglianza accurata del patogeno di origine alimentare *salmonella*.

Metodi

Campioni

Gli isolati batterici raccolti nell'ambito della sorveglianza di routine PulseNet sono stati coltivati su BBL Blood Agar Base (senza aggiunta di sangue) (BD, n. di catalogo 211037). Il DNA genomico (gDNA, genomic DNA) è stato estratto utilizzando Wizard Genomic DNA Purification Kit (Promega, n. di catalogo A1120) seguendo un protocollo precedentemente descritto per l'isolamento del gDNA da batteri Gram-negativi.⁶ Il gDNA purificato è stato quantificato con Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific, n. di catalogo Q32853) prima della preparazione delle librerie.

Preparazione delle librerie

Le librerie pronte per il sequenziamento sono state preparate manualmente a partire da 100 ng di gDNA estratto utilizzando Illumina DNA Prep (Illumina, n. di catalogo 20060060) con modifiche al protocollo per migliorare le prestazioni su MiSeq i100 Series, incluso un ulteriore ciclo di purificazione delle microsfere con un rapporto microsfere/campioni di 0,5× per rimuovere gli inserti corti.⁷ La qualità e la concentrazione delle librerie amplificate mediante PCR sono state valutate utilizzando 4200 TapeStation System (Agilent Technologies, n. di catalogo G2991BA) e Qubit 1X dsDNA High Sensitivity (HS) Assay Kit (Thermo Fisher Scientific, n. di catalogo Q33231) prima del raggruppamento in pool.

Sequenziamento

Le librerie preparate sono state raggruppate in pool per volume seguendo un protocollo descritto in precedenza⁶ con una leggera modifica che includeva un ulteriore ciclo di purificazione delle microsfere a 0,5× SPRI eseguito dopo la quantificazione e il raggruppamento in pool equimolare di ciascuna libreria. Le librerie raggruppate in pool sono state diluite a una concentrazione di caricamento di 60 pM (32 librerie/corsa). Il sequenziamento è stato eseguito su MiSeq i100 Plus System utilizzando una cella a flusso 25M con una configurazione della corsa di 2 × 301 bp. Per studi più ampi, le corse di sequenziamento possono essere estese fino a NextSeq™ 1000, NextSeq 2000, NovaSeq™ 6000 o NovaSeq X System.

Analisi dei dati

Al termine del sequenziamento, i dati sono stati trasmessi a BaseSpace™ Sequence Hub e analizzati con l'app DRAGEN™ FastQC + MultiQC per fornire statistiche e metriche generali, mappatura delle sequenze e allineamento. L'analisi è stata eseguita nell'applicazione PulseNet 2.0 Web.

All'interno di questa applicazione, sono state calcolate le metriche di controllo qualità (QC, Quality Control) basate su lettura e assemblaggio,⁵ è stata eseguita la tipizzazione delle sequenze multilocus del genoma core (cgMLST, core genome Multilocus Sequence Typing) utilizzando l'identificatore allelico PulseNet 2.0,⁶ la sierotipizzazione è stata eseguita utilizzando SeqSero2,⁸ la profilazione AMR è stata eseguita utilizzando AMRFinderPlus⁹ e i plasmidi sono stati identificati dai database PlasmidFinder.¹⁰ L'albero cgMLST e i genotipi associati sono stati visualizzati utilizzando l'applicazione web iTOL v7.¹¹

Risultati

Metriche di sequenziamento

I dati del demultiplex sono stati generati dopo 20 cicli con una durata totale della corsa di 14 ore e 20 minuti, il che conferma che le librerie sono state quantificate e raggruppate correttamente. MiSeq i100 Plus System ha generato dati di sequenziamento con una media di 1,88 milioni di letture per campione a una lunghezza di lettura di 2 × 301 bp con una percentuale maggiore o uguale al 97,5% delle basi al di sopra di Q30 (Figura 2A). La percentuale osservata di contenuto di GC delle letture ha seguito una distribuzione approssimativamente normale, il che indica che la chimica di preparazione delle librerie e il sequenziamento non sono stati fortemente spinti verso regioni GC superiori o inferiori (Figura 2B).

I 32 isolati batterici sequenziati hanno dimostrato una qualità di sequenziamento costantemente elevata, con un punteggio qualitativo mediano di 37,1 (Tabella 1). La profondità di copertura stimata variava da 46,1× a 109,9×, con una media di 71,9×. La percentuale di loci core dello schema cgMLST di PulseNet 2.0 *Salmonella* era maggiore del 99% per tutti i campioni, il che indica una completezza del genoma sufficiente per il clustering mediante cgMLST.

Previsione dei sierotipi di salmonella

I file FASTQ sono stati analizzati utilizzando il software SeqSero2 per la previsione dei sierotipi di *salmonella*. Trentuno campioni sono stati stimati correttamente come *Salmonella* Infantis (I 7:r:1,5) e uno è stato stimato correttamente come *Salmonella* Enteritidis (I 9:g,m:-) (Figura 3).

Tipizzazione molecolare dei patogeni

L'analisi cluster di 3002 loci cgMLST nello schema PulseNet 2.0 *Salmonella* ha dedotto il grado di correlazione genomica tra gli isolati nel set di campioni.

Campioni raggruppati come previsto, con campioni tipizzati come REPJFX01 raggruppati insieme. Ulteriori campioni di *Salmonella Infantis* e un campione di *Salmonella Enteritidis* che non sono stati tipizzati come cluster REPJFX01 separatamente come previsto (Figura 3). Sono stati trovati sette tipi di sequenza (ST, Sequence Type) multilocus genica utilizzando i loci *aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA* e *thrA*. I tipi di sequenza (ST) corrispondevano al risultato previsto per tutti i 32 campioni. L'AMR previsto corrispondeva ai risultati previsti per tutti i 32 campioni e, in particolare, la mutazione *gyrA_D87Y* caratteristica dei ceppi REPJFX01 è stata rilevata in tutti i campioni REP del set di dati (Figura 3). Il contenuto di plasmidi corrispondeva ai risultati previsti per tutti i 32 campioni e il plasmide IncFIB (pN55391) è stato rilevato in tutti i ceppi REPJFX01, che è anche una caratteristica distintiva di questo ceppo REP (Figura 3).

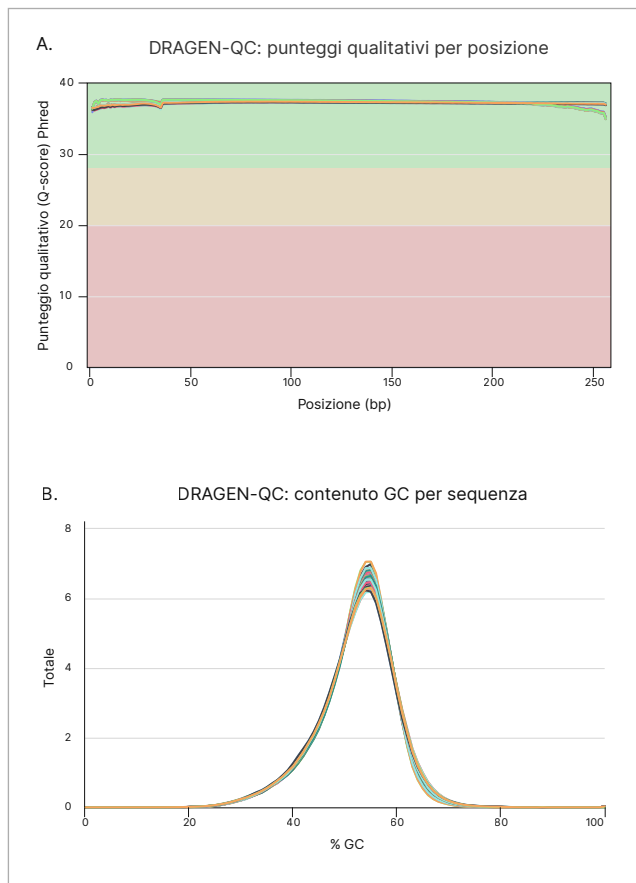


Figura 2: dati di alta qualità per il rilevamento di patogeni di origine alimentare

(A) MiSeq i100 Plus System genera dati di alta qualità con una percentuale maggiore o uguale al 97,5% delle basi al di sopra di Q30 (area verde) per tutti i campioni analizzati (linee colorate). (B) La distribuzione percentuale del contenuto di GC per le letture sequenziate su MiSeq i100 Plus System ha soddisfatto i livelli previsti per tutti i 32 campioni analizzati (linee colorate). I dati sono stati elaborati utilizzando l'app DRAGEN FastQC + MultiQC.

Tabella 1: metriche di qualità selezionate

Campione	Qualità media (punteggio qualitativo)	Profondità di copertura stimata	Percentuale di loci core identificati
1	37,1	76,7x	99,3%
2	37,1	93,0x	99,3%
3	37,1	68,3x	99,3%
4	37,1	67,2x	99,2%
5	37,1	65,5x	99,3%
6	37,0	96,9x	99,4%
7	37,1	77,3x	99,3%
8	37,0	62,1x	99,3%
9	37,1	59,9x	99,2%
10	37,0	57,7x	99,2%
11	37,0	63,4x	99,3%
12	37,0	52,8x	99,3%
13	37,1	86,2x	99,3%
14	37,0	55,7x	99,3%
15	37,0	66,8x	99,2%
16	37,0	55,6x	99,4%
17	37,1	109,9x	99,5%
18	37,1	70,5x	99,5%
19	37,0	54,8x	99,5%
20	37,0	46,1x	99,5%
21	37,1	77,0x	99,3%
22	37,1	78,2x	99,2%
23	37,1	76,6x	99,2%
24	37,1	100,6x	99,2%
25	37,1	80,1x	99,2%
26	37,2	78,0x	99,2%
27	37,1	63,0x	99,3%
28	37,0	73,4x	99,3%
29	37,1	88,8x	99,2%
30	37,0	61,6x	99,5%
31	37,1	71,3x	99,5%
32	37,1	66,8x	99,4%

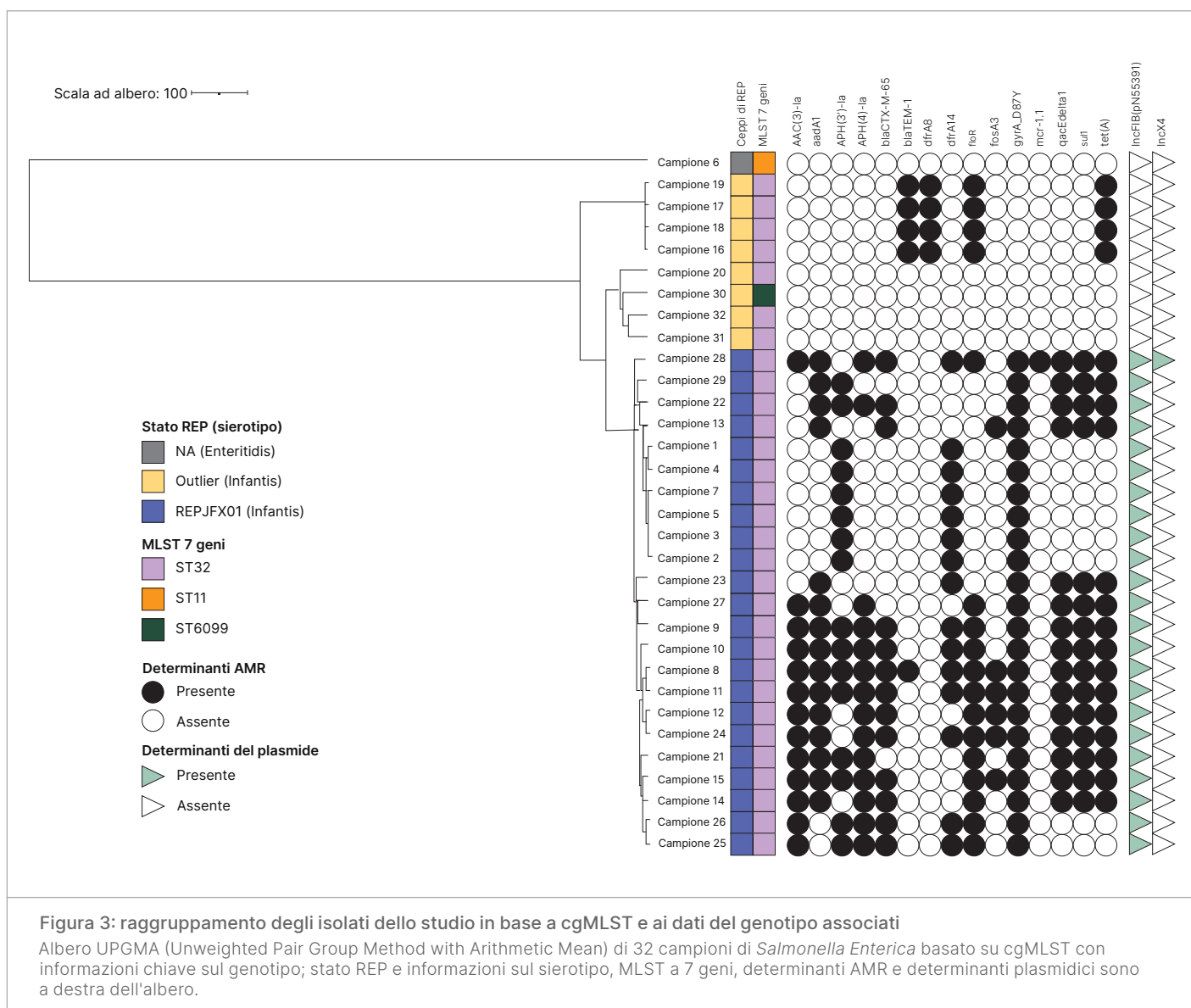


Figura 3: raggruppamento degli isolati dello studio in base a cgMLST e ai dati del genotipo associati

Albero UPGMA (Unweighted Pair Group Method with Arithmetic Mean) di 32 campioni di *Salmonella Enterica* basato su cgMLST con informazioni chiave sul genotipo; stato REP e informazioni sul sierotipo, MLST a 7 geni, determinanti AMR e determinanti plasmidici sono a destra dell'albero.

Riepilogo

Le malattie di origine alimentare correlate alla contaminazione microbica da batteri rappresentano una minaccia sostanziale per la salute umana. MiSeq i100 Series offre un flusso di lavoro NGS rapido e completo che consente un rilevamento altamente accurato dei patogeni batterici di origine alimentare per una sorveglianza efficace nell'ambito degli sforzi di salute pubblica.

Maggiori informazioni

[Illumina DNA Prep](#)

[MiSeq i100 Series](#)

Bibliografia

- Scharff RL, Besser J, Sharp DJ, Jones TF, Peter GS, Hedberg CW. [An Economic Evaluation of PulseNet: A Network for Foodborne Disease Surveillance](#). *Am J Prev Med*. 2016;50 (5 Suppl 1):S66-S73. doi:10.1016/j.amepre.2015.09.018
- Ribot EM, Freeman M, Hise KB, Gerner-Smidt P. [PulseNet: Entering the Age of Next-Generation Sequencing](#). *Foodborne Pathog Dis*. 2019;16(7):451-456. doi:10.1089/fpd.2019.2634
- Stevens EL, Carleton HA, Beal J, et al. [Use of Whole Genome Sequencing by the Federal Interagency Collaboration for Genomics for Food and Feed Safety in the United States](#). *J Food Prot*. 2022;85(5):755-772. doi:10.4315/JFP-21-437
- Tolar B, Joseph LA, Schroeder MN, et al. [An Overview of PulseNet USA Databases](#). *Foodborne Pathog Dis*. 2019;16(7):457-462. doi:10.1089/fpd.2019.2637
- Leeper MM, Tolar BM, Griswold T, et al. [Evaluation of whole and core genome multilocus sequence typing allele schemes for Salmonella enterica outbreak detection in a national surveillance network, PulseNet USA](#). *Front Microbiol*. 2023;14:1254777. Pubblicato il 21 settembre 2023. doi:10.3389/fmicb.2023.1254777
- Leeper MM, Schroeder MN, Griswold T, et al. [Validation of Core and Whole-Genome Multi-Locus Sequence Typing Schemes for Shiga-Toxin-Producing E. coli \(STEC\) Outbreak Detection in a National Surveillance Network, PulseNet 2.0, USA](#). *Microorganisms*. 2025;13(6):1310. Pubblicato il 4 giugno 2025. doi:10.3390/microorganisms13061310
- Illumina. [Maximizing performance on the MiSeq i100 Series](#). [illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/miseq-i100-library-optimization-technote-m-gl-03322/miseq-i100-library-optimization-technote-m-gl-03322.pdf](https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/miseq-i100-library-optimization-technote-m-gl-03322/miseq-i100-library-optimization-technote-m-gl-03322.pdf). Pubblicato nel 2024. Consultato il 5 febbraio 2025.
- Zhang S, den Bakker HC, Li S, et al. [SeqSero2: Rapid and Improved Salmonella Serotype Determination Using Whole-Genome Sequencing Data](#). *Appl Environ Microbiol*. 2019;85(23):e01746-19. Pubblicato il 14 novembre 2019. doi:10.1128/AEM.01746-19
- Feldgarden M, Brover V, Gonzalez-Escalona N, et al. [AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence](#). *Sci Rep*. 2021;11(1):12728. Pubblicato il 16 giugno 2021. doi:10.1038/s41598-021-91456-0
- Carattoli A, Zankari E, García-Fernández A, et al. [In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing](#). *Antimicrob Agents Chemother*. 2014;58(7):3895-3903. doi:10.1128/AAC.02412-14
- Letunic I, Bork P. [Interactive Tree of Life \(iTOL\) v6: recent updates to the phylogenetic tree display and annotation tool](#). *Nucleic Acids Res*. 2024;52(W1):W78-W82. doi:10.1093/nar/gkae268



Numero verde 1.800.809.4566 (U.S.A.) | Tel. +1.858.202.4566
techsupport@illumina.com | www.illumina.com

© 2025 Illumina, Inc. Tutti i diritti riservati. Tutti i marchi di fabbrica sono di proprietà di Illumina, Inc. o dei rispettivi proprietari. Per informazioni specifiche sui marchi di fabbrica, visitare la pagina web www.illumina.com/company/legal.html.
M-GL-02912 ITA v2.0