

Surveillance de la salubrité alimentaire avec la série MiSeq^{MC} i100

Détection précise des agents pathogènes
d'origine alimentaire



Réduction de la durée
de manipulation grâce à
la préparation simplifiée
des bibliothèques



Durées d'analyse de
séquençage plus courtes
et débits plus élevés



Assemblage fiable de
génomés bactériens
de haute qualité

Introduction

L'utilisation de méthodes de laboratoire pour déterminer le sous-type des bactéries entériques, notamment celui de la *Salmonella*, a joué un rôle essentiel dans l'identification des épidémies potentielles et dans l'établissement de liens entre les maladies causées par des bactéries et les sources potentielles d'épidémie. À mesure que la résolution des méthodes de sous-typage en laboratoire s'est améliorée, la précision dans l'établissement de liens entre la maladie et une source s'est également accrue. Aux États-Unis, l'établissement de liens pour des cas d'épidémie potentiels est effectué via PulseNet, un réseau national de surveillance moléculaire pour les maladies entériques d'origine alimentaire, d'origine hydrique et associées au concept « One Health » (Une seule santé), que l'on trouve dans plus de 80 laboratoires étatiques, locaux et fédéraux. Le réseau PulseNet prévient plus de 250 000 maladies causées par des bactéries entériques chaque année et permet d'économiser un demi-milliard de dollars en coûts médicaux et en perte de productivité^{1,2}.

PulseNet a commencé à mettre en œuvre le séquençage du génome entier (WGS, Whole-genome Sequencing) à l'aide des instruments d'Illumina au début des années 2010 avec des partenaires fédéraux de la Food and Drug Administration (FDA) des États-Unis et du Food Safety and Inspection Service du département de l'agriculture des États-Unis (USDA-FSIS)³. Grâce au WGS, les scientifiques de la santé publique peuvent détecter de potentielles épidémies en identifiant les souches issues de cas hautement similaires sur le plan génétique et associées à des

maladies survenant au cours de la même période ou dans la même région géographique. Le WGS est également utilisé pour identifier les tendances en matière de souches bactériennes qui n'étaient pas auparavant caractérisées par les méthodes moléculaires de résolution inférieure^{4,5}. Ces souches sont réapparues au fil du temps et ont provoqué des épidémies, sont apparues et ont entraîné une augmentation des cas de maladies, ou ont persisté, provoquant l'apparition de maladies pendant des mois ou des années. Les Centers for Disease Control and Prevention (CDC, Centres pour le contrôle et la prévention des maladies) ont commencé à décrire ces souches comme des souches récurrentes, émergentes et **persistantes (REP, Reoccurring, Emerging, Persisting)**. Les souches REP peuvent être identifiées par leur lien génétique et d'autres caractéristiques déterminantes, notamment la multirésistance aux médicaments, la virulence ou d'autres marqueurs génétiques. Étant donné que seulement environ 10 % des bactéries responsables de maladies dans PulseNet sont associées à une épidémie, l'utilisation du WGS pour identifier les souches REP peut offrir la possibilité de prévenir d'autres maladies en reliant des souches supplémentaires qui peuvent partager une source commune.

Cette note d'application démontre la détection et la caractérisation précises d'une souche persistante de *Salmonella enterica* sérotype Infantis (**REP JFX01**) multirésistante à l'aide d'un flux de travail de séquençage de nouvelle génération (SNG) efficace qui intègre Illumina DNA Prep, la série MiSeq i100 et des outils d'analyse tiers (**figure 1**).



Méthodes

Échantillons

Les isolats bactériens recueillis dans le cadre de la surveillance de routine PulseNet ont été mis en culture sur base BBL Blood Agar Base (sans ajout de sang) (BD, référence n° 211037). L'ADN génomique (ADNg) a été extrait à l'aide de la trousse Wizard Genomic DNA Purification Kit (Promega, référence n° A1120) en suivant un protocole décrit précédemment pour l'isolement de l'ADNg des bactéries à Gram négatif⁶. L'ADNg purifié a été quantifié avec la trousse Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific, référence n° Q32853) avant la préparation des librairies.

Préparation des librairies

Les librairies prêtes pour le séquençage ont été préparées manuellement à partir de 100 ng d'ADNg extrait à l'aide d'Illumina DNA Prep (Illumina, référence n° 20060060) en apportant des modifications au protocole pour améliorer les performances sur la série MiSeq i100, notamment une phase supplémentaire de purification des billes avec un rapport bille/échantillon de 0,5× pour retirer les inserts courts⁷. La qualité et la concentration des librairies amplifiées par PCR ont été évaluées à l'aide de 4200 TapeStation System (Agilent Technologies, référence n° G2991BA) et la trousse Qubit 1X dsDNA High Sensitivity (HS) Assay Kit (Thermo Fisher Scientific, référence n° Q33231) avant le regroupement.

Séquençage

Les librairies préparées ont été regroupées par volume selon un protocole décrit précédemment⁶ avec une légère modification qui consistait en l'ajout d'une phase supplémentaire de purification des billes à 0,5× par immobilisation réversible en phase solide (SPRI, Solid-phase Reversible Immobilization) effectuée après quantification et regroupement équimolaire de chaque librairie. Les librairies regroupées ont été diluées à une concentration de chargement de 60 pM (32 librairies/analyse). Le séquençage a été effectué sur MiSeq i100 Plus System à l'aide d'une Flow Cell 25M avec une configuration d'analyse de 2 × 301 pb. Pour les études de plus grande envergure, les analyses de séquençage peuvent être étendues aux systèmes NextSeq 1000, NextSeq 2000, NovaSeq 6000 ou NovaSeq X.

Analyse des données

Une fois le séquençage terminé, les données ont été transmises dans BaseSpace^{MC} Sequence Hub et analysées à l'aide de l'application DRAGEN^{MC} FastQC + MultiQC pour fournir des statistiques et des indicateurs généraux, la cartographie des séquences et l'alignement. L'analyse a été effectuée dans l'application Web PulseNet 2.0.

Dans cette application, les indicateurs de contrôle de la qualité (CQ) basés sur la lecture et l'assemblage ont été calculés⁵, le typage de séquence multilocus du génome de base (cgMLST, Core Genome Multilocus Sequence Typing) a été effectué à l'aide du paramètre d'appel d'allèles PulseNet 2.0⁸, le sérotypage a été effectué à l'aide de SeqSero2⁸, le profilage de la résistance aux antimicrobiens (RAM) a été effectué à l'aide d'AMRFinderPlus⁹ et les plasmides ont été identifiés à partir des bases de données PlasmidFinder¹⁰. L'arbre du cgMLST et les génotypes associés ont été visualisés à l'aide de l'application Web iTOL v7¹¹.

Résultats

Indicateurs de séquençage

Les données de démultiplexage ont été générées après 20 cycles avec une durée d'analyse totale de 14 h et 20 min, confirmant ainsi la quantification et le regroupement corrects des librairies. MiSeq i100 Plus System a généré des données de séquençage avec une moyenne de 1,88 million de lectures par échantillon à une longueur de lecture de 2 × 301 pb avec ≥ 97,5 % de bases au-dessus de Q30 (figure 2A). Le pourcentage observé de la teneur en GC des lectures suivait une distribution à peu près normale, indiquant que la chimie de préparation des librairies et le séquençage n'étaient pas fortement biaisés en faveur des régions à teneur en GC supérieure ou inférieure (figure 2B).

Les 32 isolats bactériens qui ont été séquencés ont démontré une qualité de séquençage constamment élevée, avec un score de qualité médian de 37,1 (tableau 1). La profondeur de couverture estimée allait de 46,1× à 109,9×, avec une moyenne de 71,9×. Le pourcentage de locus de base du schéma de cgMLST de la *Salmonella* PulseNet 2.0 était supérieur à 99 % pour tous les échantillons, ce qui indique une exhaustivité suffisante du génome pour la génération d'amplifiats par cgMLST.

Prédiction du sérotype de la *Salmonella*

Les fichiers FASTQ ont été analysés à l'aide du logiciel SeqSero2 pour prédire le sérotype de la *Salmonella*. Trente-et-un échantillons ont été correctement prédits comme étant de sérotype *Salmonella* Infantis (I 7:r:1,5) et un échantillon a été correctement prédit comme étant de sérotype *Salmonella* Enteritidis (I 9:g,m:-) (figure 3).

Typage moléculaire des agents pathogènes

L'analyse des amplifiats de 3 002 locus de cgMLST dans le schéma *Salmonella* PulseNet 2.0 a déduit le degré de similarité génomique entre les isolats de l'ensemble d'échantillons. Les échantillons ont été regroupés comme prévu, les échantillons typés REPJFX01 ayant été regroupés ensemble.

Des échantillons supplémentaires de *Salmonella* Infantis et un échantillon de *Salmonella* Enteritidis qui n'ont pas été typés REPJFX01 se regroupent séparément comme prévu (figure 3). Sept types de séquence (ST, Sequence Type) multilocus de gènes ont été détectés à l'aide des *locus aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA* et *thrA*. Les types de séquence (ST) correspondaient au résultat attendu pour les 32 échantillons. La RAM prédite correspondait aux résultats attendus pour les 32 échantillons et, particulièrement, la mutation *gyrA_D87Y* caractéristique des souches REPJFX01 a été détectée dans tous les échantillons REP de l'ensemble de données (figure 3). Le contenu plasmidique correspondait aux résultats attendus pour les 32 échantillons, et le plasmide IncFIB (pN55391) a été détecté dans toutes les souches REPJFX01, ce qui représente également une caractéristique déterminante de cette souche REP (figure 3).

Tableau 1: Sélection des indicateurs de qualité

Échantillon	Qualité moyenne (Score de qualité)	Profondeur de couverture estimée	Pourcentage de locus de base appelés
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 %

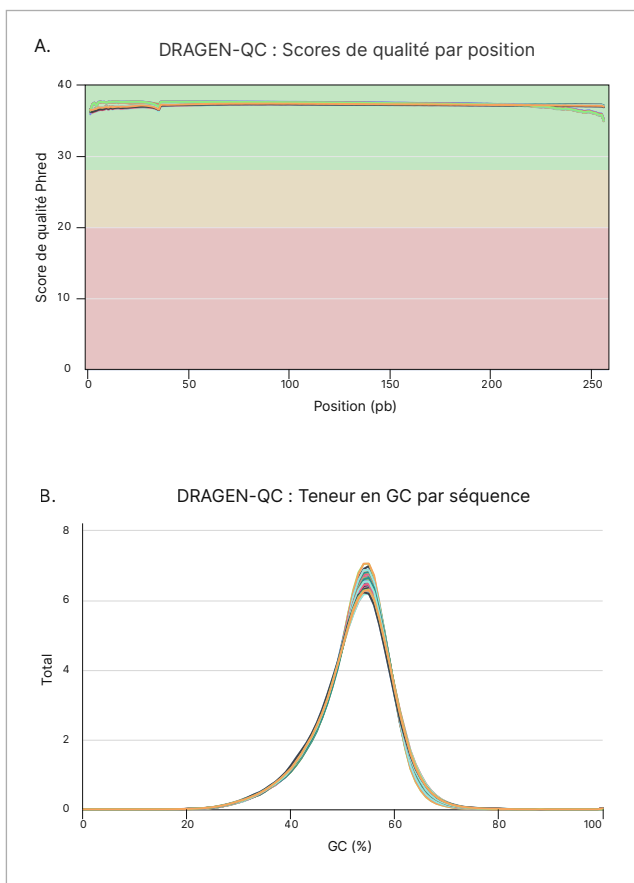


Figure 2 : Données de haute qualité pour la détection des agents pathogènes d'origine alimentaire

(A) MiSeq i100 Plus System génère des données de haute qualité avec $\geq 97,5$ % de bases au-dessus de Q30 (zone verte) pour tous les échantillons analysés (lignes colorées). (B) Le pourcentage de distribution de la teneur en GC pour les lectures séquencées sur MiSeq i100 Plus System a atteint les niveaux attendus pour les 32 échantillons analysés (lignes colorées). Les données ont été traitées à l'aide de l'application DRAGEN FastQC + MultiQC.

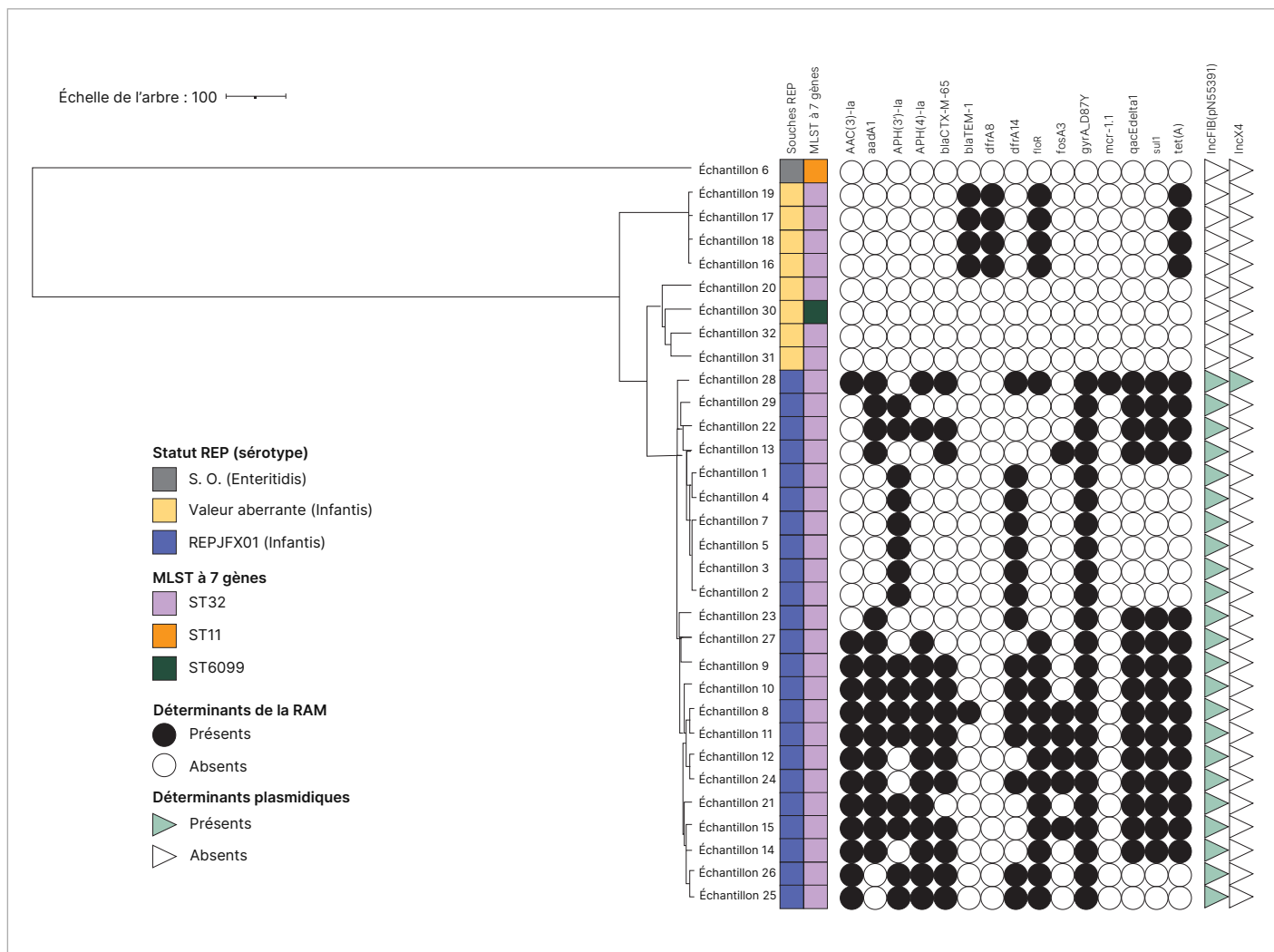


Figure 3 : Génération d'amplifiats des isolats de l'étude basée sur le cgMLST et les données de génotype associées

Arbre de la méthode des groupes de paires non pondérés avec moyenne arithmétique (UPGMA, Unweighted Pair Group Method with Arithmetic) basé sur le cgMLST de 32 échantillons de *Salmonella enterica* avec renseignements clés sur le génotype; les renseignements sur le statut REP et le sérotipe, le typage de séquence multi-locus (MLST, Multi-Locus Sequence Typing) à sept gènes, les déterminants de la RAM et les déterminants plasmidiques se trouvent à droite sur l'arbre.

Résumé

Les maladies d'origine alimentaire associées à la contamination microbienne par des bactéries représentent une menace importante pour la santé humaine. La série MiSeq i100 permet d'offrir un flux de travail de SNG rapide et complet qui permet une détection très précise des agents pathogènes bactériens d'origine alimentaire pour une surveillance efficace dans le cadre des efforts de santé publique.

En savoir plus

[Illumina DNA Prep](#)

[Série MiSeq i100](#)

Références

- 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. Publié le 21 septembre 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. Publié le 4 juin 2025. doi:10.3390/microorganisms13061310
- Illumina. Maximizing performance on the MiSeq i100 Series. [illumina.com/content/dam/illumina/gcs/assembly-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/assembly-assets/marketing-literature/miseq-i100-library-optimization-technote-m-gl-03322/miseq-i100-library-optimization-technote-m-gl-03322.pdf). Publié en 2024. Consulté le 5 février 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. Publié le 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. Publié le 16 juin 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



Numéro sans frais aux États-Unis : + (1) 800 809-4566 | Téléphone : + (1) 858 202-4566
techsupport@illumina.com | www.illumina.com

© 2025 Illumina, Inc. Tous droits réservés. Toutes les marques de commerce sont la propriété d'Illumina, Inc. ou de leurs détenteurs respectifs. Pour obtenir des renseignements sur les marques de commerce, consultez la page www.illumina.com/company/legal.html. M-GL-02912 FRA v2.0