

# Vigilância da segurança alimentar com o MiSeq™ i100 Series

Detecção precisa de patógenos transmitidos por alimentos



Menor tempo de trabalho prático com preparo simplificado da biblioteca



Tempos de corrida de sequenciamento mais curtos e maior rendimento



Montagem completa de genomas bacterianos de alta qualidade

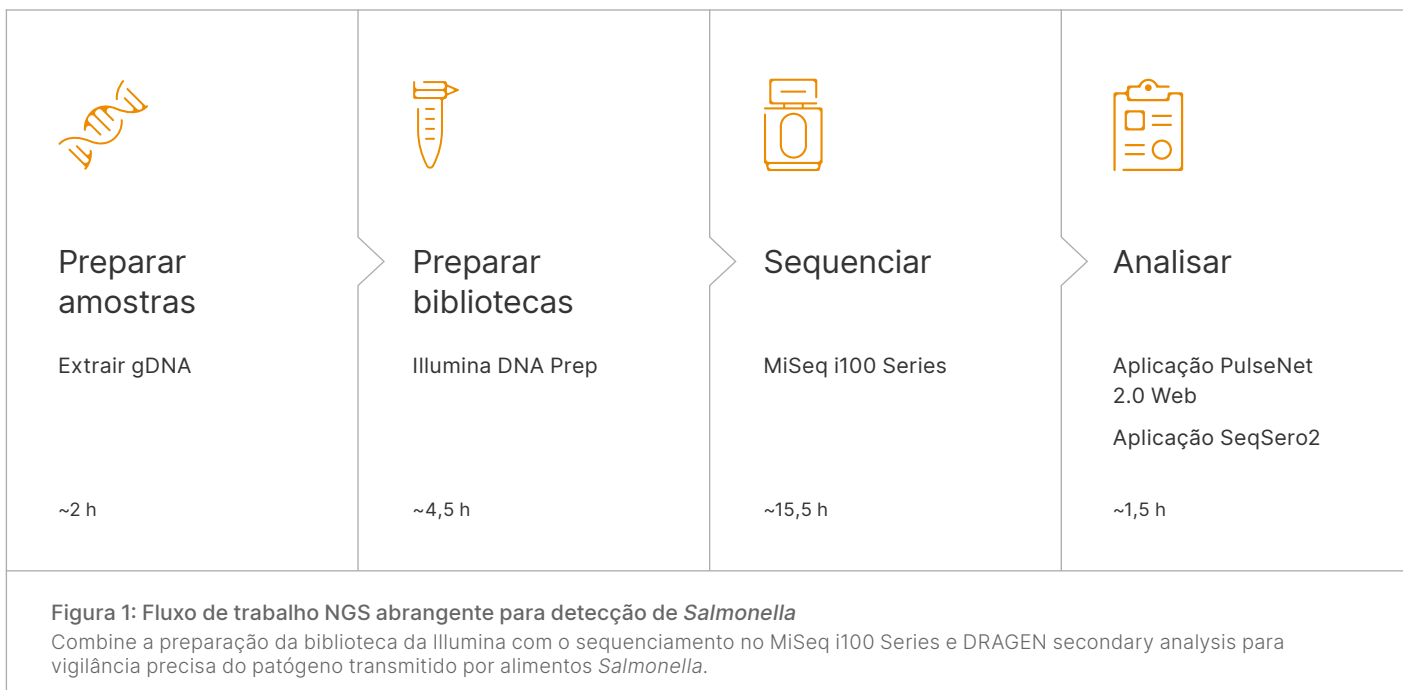
## Introdução

O uso de métodos laboratoriais para subtipar bactérias entéricas, incluindo *Salmonella*, tem sido fundamental para identificar possíveis surtos e relacionar doenças causadas por bactérias a fontes potenciais de surtos. Conforme a resolução dos métodos de subtipagem em laboratório avançou, também aumentou a precisão em relacionar doenças a uma fonte. Nos Estados Unidos, a ligação de possíveis casos de surtos é feita por meio do PulseNet, uma rede nacional de vigilância molecular para doenças entéricas transmitidas por alimentos, pela água e relacionadas à One Health encontradas em mais de 80 laboratórios estaduais, locais e federais. A rede PulseNet previne mais de 250 mil casos de doenças causadas por bactérias entéricas a cada ano e economiza meio bilhão de dólares em custos médicos e produtividade perdida.<sup>1,2</sup>

O PulseNet começou a implementar o sequenciamento de genoma completo (WGS) usando instrumentos da Illumina no início dos anos 2010 com parceiros federais da Food and Drug Administration (FDA) dos EUA e do US Department of Agriculture Food Safety and Inspection Service (USDA-FSIS).<sup>3</sup> Usando o WGS, os cientistas de saúde pública podem identificar possíveis surtos identificando cepas de casos que são altamente geneticamente semelhantes e associados a doenças que ocorrem no mesmo período ou região

geográfica. O WGS também é usado para identificar tendências em cepas bacterianas que não foram caracterizadas anteriormente por métodos moleculares de resolução mais baixa.<sup>4,5</sup> Essas cepas ocorreram novamente ao longo de períodos de tempo e causaram surtos, surgiram e causaram doenças crescentes ou persistiram causando doenças ao longo de meses ou anos. O Centers for Disease Control and Prevention (CDC) começou a descrever essas cepas como cepas recorrentes, emergentes e **persistentes (REP)**. As cepas REP podem ser identificadas por relação genética e outras características definidoras, incluindo resistência a múltiplos medicamentos, virulência ou outros marcadores genéticos. Como apenas cerca de 10% das bactérias causadoras de doenças no PulseNet estão associadas a um surto, o uso do WGS para identificar cepas REP pode fornecer uma oportunidade de prevenir outras doenças, ligando cepas adicionais que podem compartilhar uma fonte comum.

Esta nota de aplicação demonstra a detecção e caracterização precisas de uma cepa persistente de *Salmonella enterica* sorotipo Infantis resistente a múltiplos medicamentos (**REP JFX01**) usando um fluxo de trabalho NGS eficiente que integra o Illumina DNA Prep, o MiSeq i100 Series e ferramentas de análise de terceiros (**Figura 1**).



## Métodos

### Amostras

Os isolados bacterianos coletados como parte da vigilância de rotina da PulseNet foram cultivados em base de ágar sanguíneo BBL (sem a adição de sangue) (BD, N.º do catálogo 211037). O DNA genômico (gDNA) foi extraído usando o Wizard Genomic DNA Purification Kit (Promega, N.º do catálogo A1120) seguindo um protocolo descrito anteriormente para o isolamento de gDNA de bactérias Gram-negativas.<sup>6</sup> O gDNA purificado foi quantificado com o Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific, N.º do catálogo Q32853) antes da preparação da biblioteca.

### Preparação da biblioteca

As bibliotecas prontas para sequenciamento foram preparadas manualmente a partir de 100 ng de gDNA extraído usando o Illumina DNA Prep (Illumina, N.º do catálogo 20060060) com modificações ao protocolo para melhorar o desempenho no MiSeq i100 Series, incluindo uma rodada adicional de purificação de beads com uma proporção de beads para amostra de 0,5× para remover insertos curtos.<sup>7</sup> A qualidade e a concentração das bibliotecas amplificadas por PCR foram avaliadas usando o 4200 TapeStation System (Agilent Technologies, N.º do catálogo G2991BA) e o Qubit 1X dsDNA High Sensitivity (HS) Assay Kit (Thermo Fisher Scientific, N.º do catálogo Q33231) antes do pooling.

### Sequenciamento

As bibliotecas preparadas foram agrupadas por volume seguindo um protocolo descrito anteriormente<sup>8</sup> com uma pequena modificação que incluiu uma rodada adicional de purificação de beads a 0,5× SPRI realizada após quantificação e pooling equimolar de cada biblioteca. As bibliotecas agrupadas foram diluídas a uma concentração de carga de 60 pM (32 bibliotecas/corrída). O sequenciamento foi realizado no MiSeq i100 Plus System usando uma lâmina de fluxo de 25M com configuração de corrida de 2 × 301 bp. Para estudos maiores, as corridas de sequenciamento podem ser dimensionadas para os sistemas NextSeq 1000, NextSeq 2000, NovaSeq 6000 ou NovaSeq X.

### Análise de dados

Após a conclusão do sequenciamento, os dados foram transmitidos para o BaseSpace™ Sequence Hub e analisados com a aplicação DRAGEN™ FastQC + MultiQC para fornecer estatísticas e métricas gerais, mapeamento de sequências e alinhamento.

A análise foi realizada na aplicação Web PulseNet 2.0. Dentro dessa aplicação, métricas de controle de qualidade (CQ) baseadas em leitura e montagem foram calculadas,<sup>5</sup> a tipagem de sequência multilocus do genoma principal (cgMLST) foi realizada usando o identificador de alelos PulseNet 2.0,<sup>6</sup> a sorotipagem foi realizada com o SeqSero2,<sup>8</sup> o perfil de AMR foi realizado usando o AMRFinderPlus,<sup>9</sup> e plasmídeos foram identificados a partir dos bancos de dados PlasmidFinder.<sup>10</sup> A árvore cgMLST e os genótipos associados foram visualizados usando a aplicação web iTOL v7.<sup>11</sup>

## Resultados

### Métricas de sequenciamento

Os dados de demultiplexação foram gerados após 20 ciclos com um tempo de corrida total de 14 horas e 20 minutos, confirmando que as bibliotecas foram quantificadas e agrupadas corretamente. O MiSeq i100 Plus System gerou dados de sequenciamento com uma média de 1,88 milhão de leituras por amostra em um comprimento da leitura de 2 × 301 bp com ≥ 97,5% das bases acima de Q30 (**Figura 2A**). O conteúdo percentual de GC observado das leituras seguiu uma distribuição aproximadamente normal, indicando que a química e o sequenciamento da preparação da biblioteca não foram fortemente inclinados para regiões de GC mais altas ou mais baixas (**Figura 2B**).

Os 32 isolados bacterianos sequenciados demonstraram qualidade de sequenciamento consistentemente alta, com uma Q-score mediana de 37,1 (**Tabela 1**). A profundidade da cobertura estimada variou de 46,1× a 109,9×, com uma média de 71,9×. A porcentagem de loci core do esquema cgMLST de *Salmonella* PulseNet 2.0 foi > 99 % para todas as amostras, indicando completude genômica suficiente para agrupamento por cgMLST.

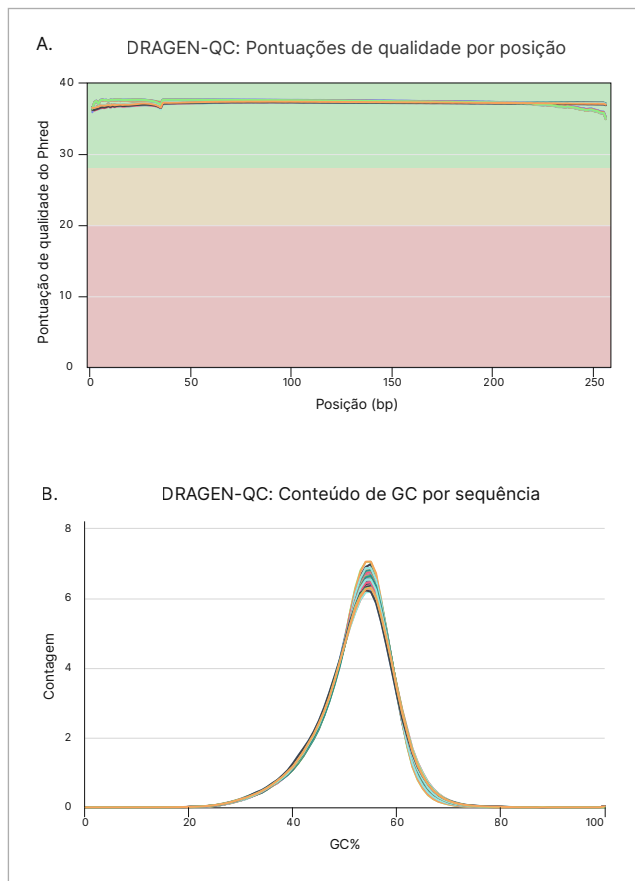
### Previsão do sorotipo *Salmonella*

Os arquivos FASTQ foram analisados usando o software SeqSero2 para previsão do sorotipo de *Salmonella*. Trinta e uma amostras foram corretamente previstas como *Salmonella* Infantis (I 7:r:1,5) e uma foi corretamente prevista como *Salmonella* Enteritidis (I 9:g,m:-) (**Figura 3**).

### Tipagem molecular de patógenos

A análise de cluster de 3002 loci cgMLST no esquema de *Salmonella* PulseNet 2.0 inferiu o grau de relação genômica entre os isolados no conjunto de amostras.

As amostras se agruparam conforme esperado, com as amostras tipadas como REPJFX01 formando um cluster conjunto. Amostras adicionais de *Salmonella* Infantil e uma amostra de *Salmonella* Enteritidis que não foram tipadas como cluster REPJFX01 separadamente, conforme esperado (Figura 3). Sete tipos de sequência multilocus (ST) de gene foram encontrados usando *loci* *aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA* e *thrA*. Os tipos de sequência (ST) corresponderam ao resultado esperado para todas as 32 amostras. O AMR previsto correspondeu aos resultados esperados para todas as 32 amostras e, notavelmente, a mutação *gyrA\_D87Y* que é característica das cepas REPJFX01 foi detectada em todas as amostras REP do conjunto de dados (Figura 3). O conteúdo de plasmídeo correspondeu aos resultados esperados para todas as 32 amostras, e o plasmídeo IncFIB (pN55391) foi detectado em todas as cepas REPJFX01, que também é uma característica definidora dessa cepa REP (Figura 3).



**Figura 2: Dados de alta qualidade para detecção de patógenos transmitidos por alimentos**

(A) O MiSeq i100 Plus System gera dados de alta qualidade com  $\geq 97,5\%$  das bases acima de Q30 (área verde) para todas as amostras analisadas (linhas coloridas). (B) A distribuição percentual do conteúdo de GC para leituras sequenciadas no MiSeq i100 Plus System atendeu aos níveis esperados para todas as 32 amostras analisadas (linhas coloridas). Os dados foram processados usando a aplicação DRAGEN FastQC + MultiQC.

**Tabela 1: Selecionar métricas de qualidade**

Amostra	Qualidade média (Q-score)	Profundidade estimada da cobertura	Porcentagem de loci core identificados
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%

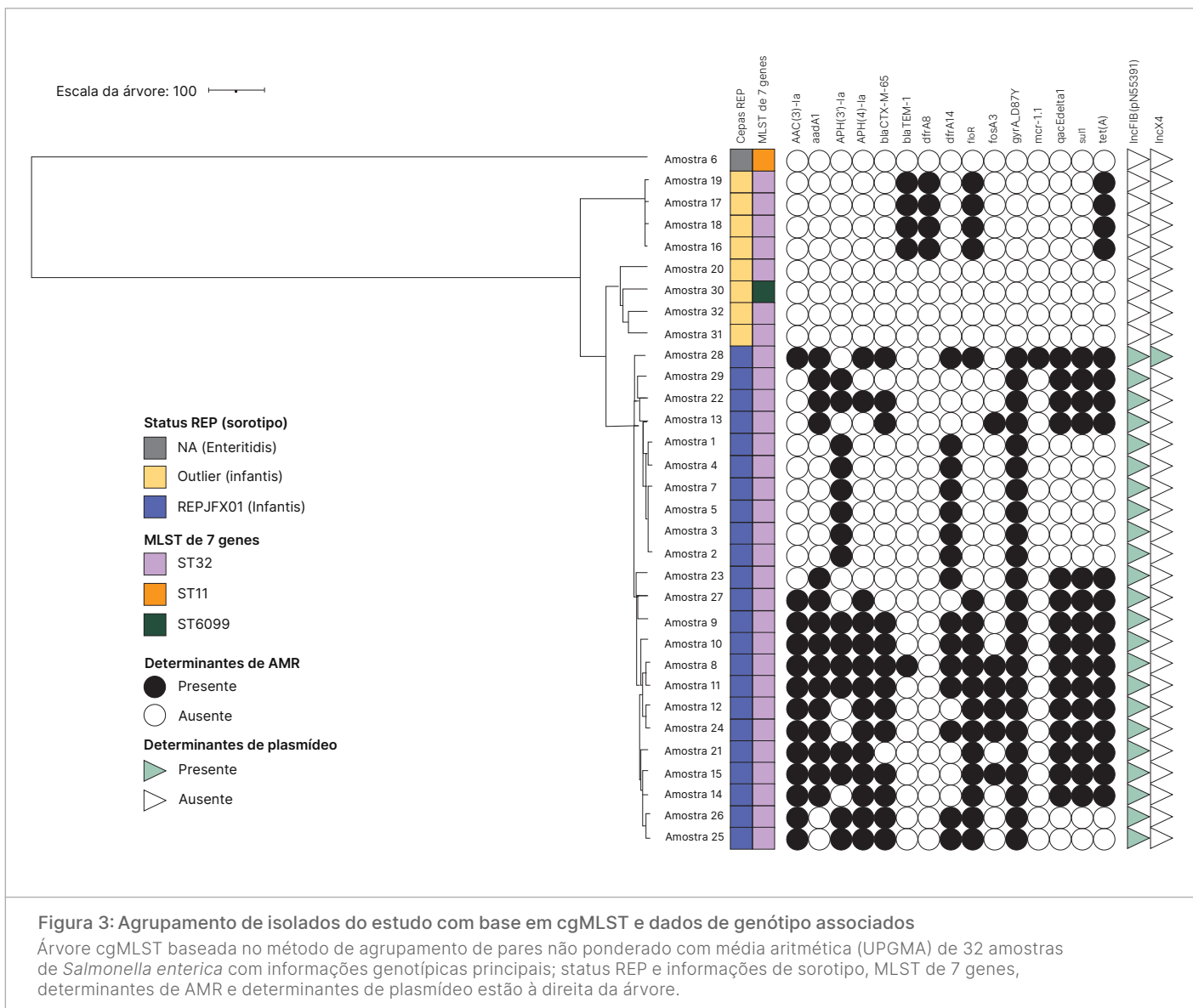


Figura 3: Agrupamento de isolados do estudo com base em cgMLST e dados de genótipo associados

Árvore cgMLST baseada no método de agrupamento de pares não ponderado com média aritmética (UPGMA) de 32 amostras de *Salmonella enterica* com informações genotípicas principais; status REP e informações de sorotipo, MLST de 7 genes, determinantes de AMR e determinantes de plasmídeo estão à direita da árvore.

## Resumo

Doenças alimentares de origem bacteriana constituem uma ameaça significativa à saúde humana. O MiSeq i100 Series ajuda a fornecer um fluxo de trabalho NGS rápido e abrangente que permite a detecção altamente precisa de patógenos bacterianos transmitidos por alimentos, apoiando a vigilância eficaz em iniciativas de saúde pública.

## Saiba mais

[Illumina DNA Prep](#)

[MiSeq i100 Series](#)

## Referências

- 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. Publicado em 21 de setembro de 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. Publicado em 4 de junho de 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). Publicado em 2024. Acessado em 5 de fevereiro de 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. Publicado em 14 de novembro de 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. Publicado em 16 de junho de 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



+1 (800) 809-4566, ligação gratuita (EUA) | tel. +1 (858) 202-4566  
techsupport@illumina.com | www.illumina.com

© 2025 Illumina, Inc. Todos os direitos reservados. Todas as marcas comerciais pertencem à Illumina, Inc. ou aos respectivos proprietários. Para obter informações específicas sobre marcas comerciais, consulte [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html).  
M-GL-02912 PTB v2.0