

Cette présentation a été effectuée le 3 décembre 2024, au cours de la journée « La génomique : un outil pour la détection et l'investigation d'éclosions » dans le cadre des 27es Journées annuelles de santé publique.



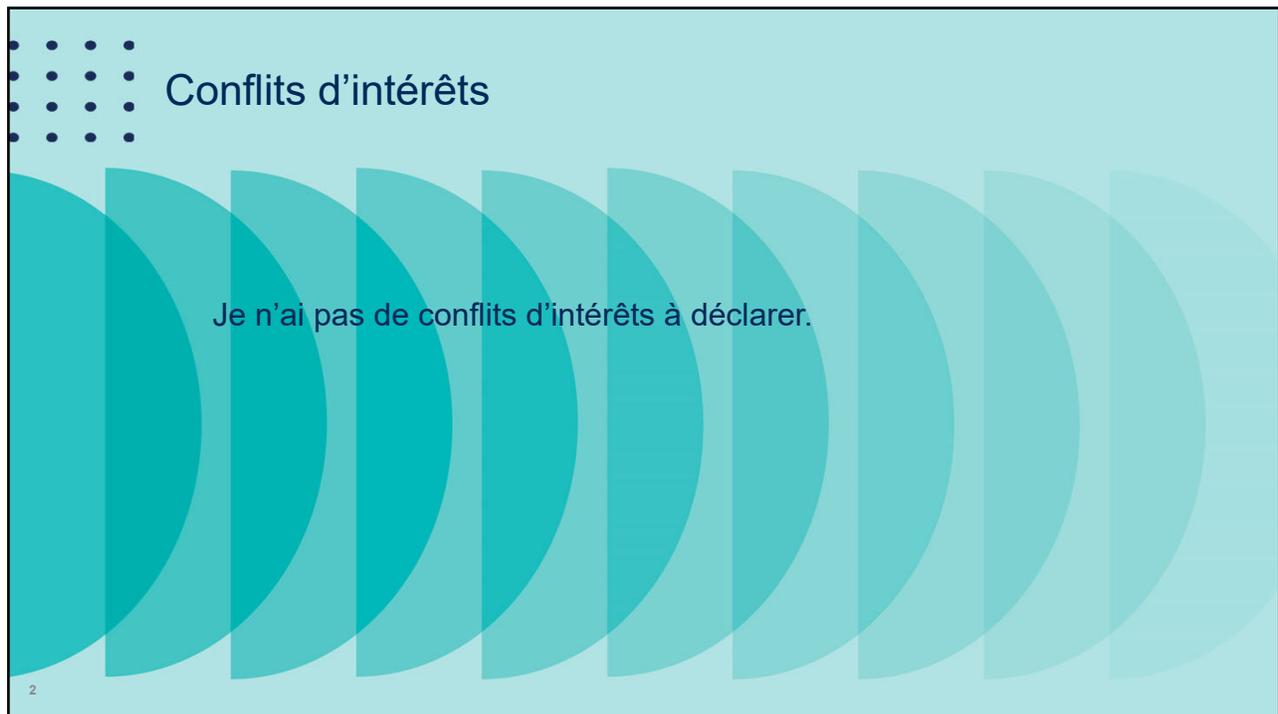
27^{ES} JOURNÉES ANNUELLES DE SANTÉ PUBLIQUE

D'ACQUIS et D'AUDACE

Concepts de base en génomique pertinents pour les professionnel(le)s de santé publique

Sadjia Bekal- Laboratoire de santé publique du Québec
2023-12-03

1



Conflits d'intérêts

Je n'ai pas de conflits d'intérêts à déclarer.

2

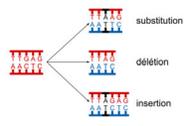
2

•••••

Définitions

- * **Génome**: l'ensemble du contenu génétique représenté par l'ADN, dans une cellule.
- * **Séquencer**: déterminer la suite des bases (A, C, G, T) de l'ADN
- * **Pourquoi séquencer ?** Parce que l'ADN est en évolution sous pression et chaque mutation représente un évènement génétique qu'on peut exploiter comme empreinte génétique, pour l'identification, la phylogénie, le typage..etc...





3

3

Le séquençage, une histoire de générations

4

Séquençage première génération (Sanger)

- * Dès 1977
- * Séquençer un fragment par réaction
- * Lire jusqu'à 1000 pb
- * Marche sur le chromosome pour séquençer plus
- * Lent et fastidieux

Principales retombées:

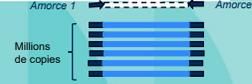
Séquençage du gène *rrs* (16S) et développement de la phylogénie

Séquençage du génome humain

(A) ADN extrait d'une cellule vivante



(B) Amplifier la cible



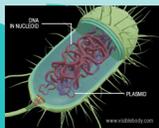
(C) Séquençer la cible



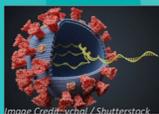
5

5

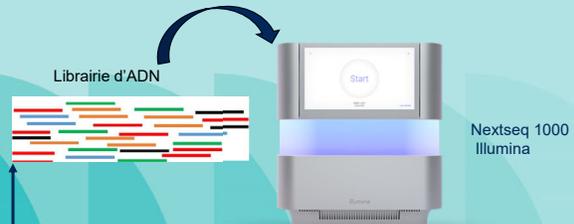
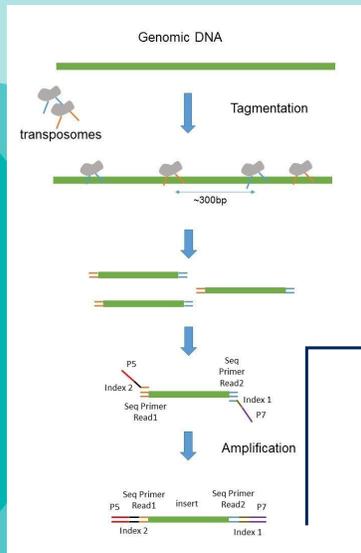
Séquençage de deuxième génération (2005)



Bactéries: Moy. du génome 5 Mb (10^9 pb)



Virus: Moy. du génome 5 kb (10^3 pb)



Read length	NextSeq 1000/2000 P1 Reagents	NextSeq 1000/2000 P2 Reagents
Single-end reads	100M	400M (300M for 2 × 300 bp)
1 × 50 bp	–	–
2 × 50 bp	10 Gb	40 Gb
2 × 100 bp	–	80 Gb
2 × 150 bp	30 Gb	120 Gb
2 × 300 bp	60 Gb	180 Gb

Read length	NextSeq 1000/2000 P1 Reagents	NextSeq 1000/2000 P2 Reagents
1 × 50 bp	–	–
2 × 50 bp	10 hr	13 hr
2 × 100 bp	–	21 hr
2 × 150 bp	19 hr	29 hr
2 × 300 bp	34 hr	44 hr*

6

6

Assemblage des reads pour reconstituer le génome: Un grand casse-tête

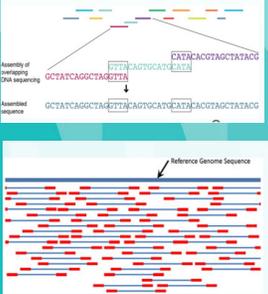
Données génomiques sorties du séquenceur sous formes de courtes lectures ou **reads** (exp. 150 pb)



Chaque base ADN est lue des dizaines de fois = Séquence fiable 😊

Assemblage de novo

Sur un génome de référence



Le génome est assemblé sous forme de **contigs** (entre 30 et 500 contigs)



On n'obtient pas une seule molécule d'ADN 😞

7

7



Méthodes d'analyses génomiques validées pour la phylogénie et leur interprétation

8

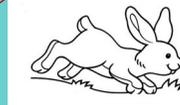
Un bon typage moléculaire doit savoir exploiter l'horloge moléculaire



- Sous pression de la sélection, les gènes évoluent mais pas à la même vitesse
- Les gènes les plus importants évoluent généralement plus lentement



Gènes de ménage (ex. 16S, cpn60)
Évolution lente
Utiles pour l'identification
Genre/espèce



Zones à évolution rapide
Utiles pour les éclosions
(SNP, allèles)
Souche/clone

Vitesse des mutations

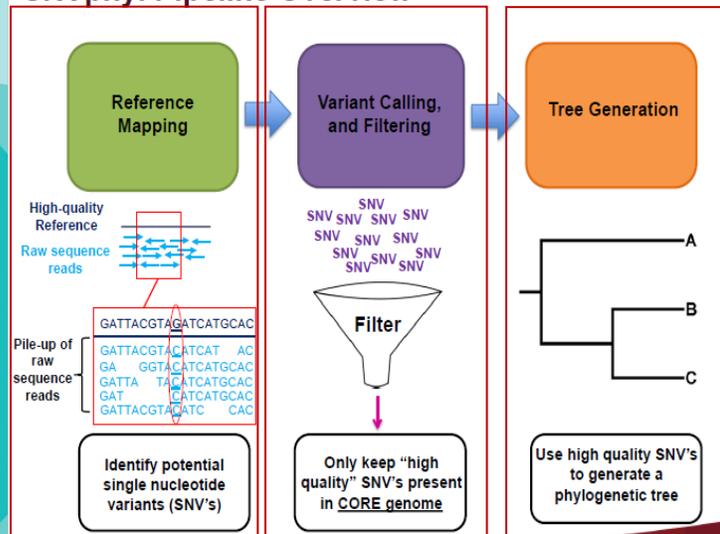
9

9

Approche 1

- SNP (single nucleotide polymorphism)
- Méthode: SNVPhyl
- Nécessite un génome de référence
- Ne considère que les gènes hérités verticalement
- Validée par PulseNet
- Non harmonisable pour surveillance inter-laboratoire
- Non retenue pour la surveillance multi-provinciale

SNVphyl Pipeline Overview



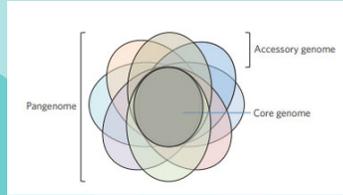
PUBLIC HEALTH AGENCY OF CANADA >

10

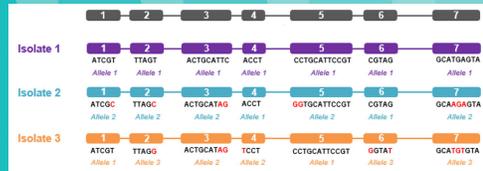
Approche 2: core ou whole genome MLST

- Comparaison Gène par gène
- Extension du MLST à 7 gènes à tous les gènes (loci)
- Ne nécessite pas un génome de référence
- Prend en considération tout gène hérité verticalement ou horizontalement
- Validée par PulseNet
- Harmonisable pour une surveillance inter-laboratoire
- Retenue pour la surveillance PulseNet

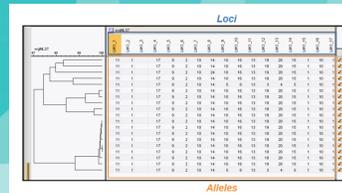
- wgMLST: comparaison avec le pangénome
- cgMLST: locci présent dans plus de 95% des souches



Base de données de tous les locci avec tous leurs allèles numérotés (ex. *Salmonella*: 15 874 locci)



Pour chaque isolat, si le locus est présent, l'allèle est déterminé

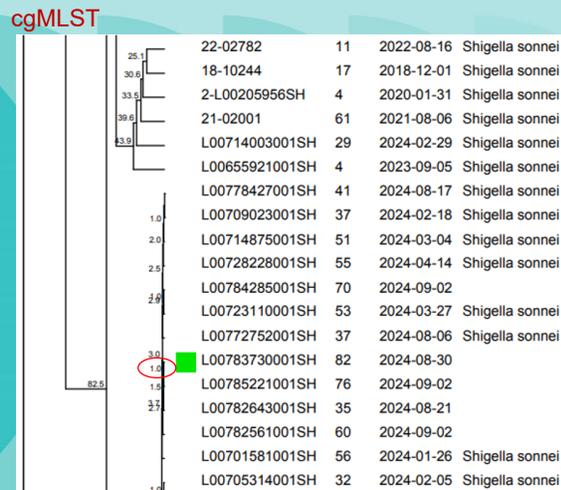
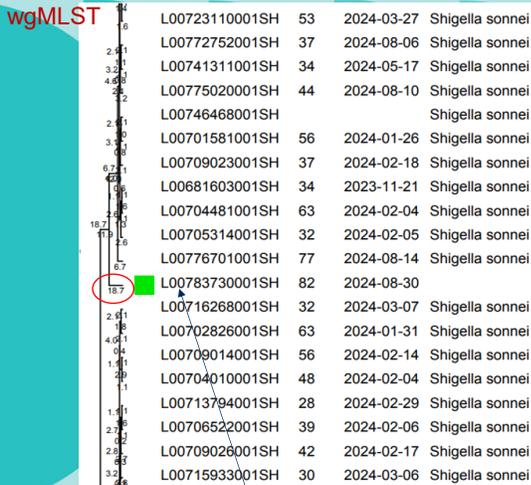


Les allèles de chacun des locci sont comparés et un arbre phylogénétique est construit.

11

11

(A) Utiliser la mauvaise approche d'analyse génomique affecte le lien génétique et donc le lien épidémiologique



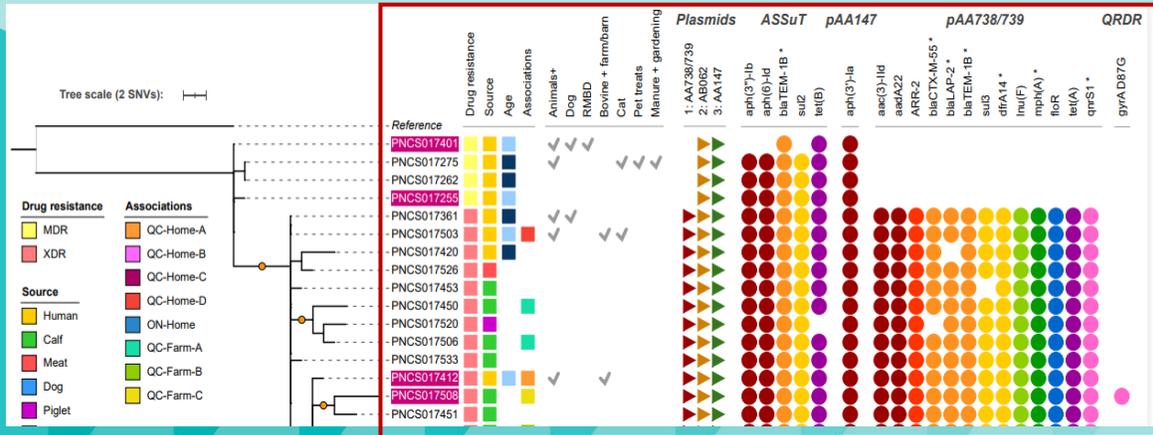
Cette souche avait un plasmide de plus que celles de l'éclouaison

12

12



(B) Se baser sur le profil de résistance pour l'inclusion épidémiologique n'est pas toujours bon



Des souches de *Salmonella* appartenant à l'agrégat 2009STWGS-2QC-MP (XDR) ont des profils d'antibiorésistance différents (plasmides différents)

13

13

Retombées du séquençage génomique Exemples des bactéries entériques



14

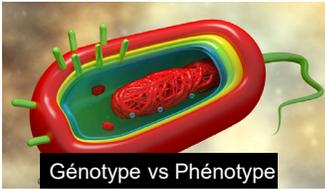
Informations livrées par la génomique

À partir de la séquence d'une souche (Reads)

<p>AMR Detection Pipeline</p> <p>A pipeline to search microbial genomes for antimicrobial resistance genes (AMR). Uses the tools RGI and StarAMR.</p> <p>Select</p> <p style="text-align: center;">Prédiction de profil de sensibilité</p>	<p>Assembly and Annotation Pipeline</p> <p>Shovill assembly, Prokka annotation and QUAST assembly assessment</p> <p>Select</p>	<p>AssemblyAnnotationCollection Pipeline</p> <p>Shovill assembly, Prokka annotation with ZIP bundling of outputs and QUAST assembly assessment</p> <p>Select</p>
<p>ECTYPER Pipeline (E.coli serotyping)</p> <p>This pipeline serotypes E.coli genomes by using ECTYPER on assembled genomes. ECTYPER is available at https://github.com/phac-nml/ecoli_serotyping.</p> <p>Select</p> <p style="text-align: center;">Détermination du sérotype de E. coli/Shigella</p>	<p>MentalIST MLST Pipeline</p> <p>Genotype bacterial samples directly from reads, using an efficient k-mer based algorithm.</p> <p>Select</p> <p style="text-align: center;">Détermination du type MLST</p>	<p>RefSeqMasher Pipeline</p> <p>Find what NCBI RefSeq genomes most closely match or are contained in your sample sequences</p> <p>Select</p>
<p>SISTR Pipeline</p> <p>Generates in silico typing results using the Salmonella In Silico Typing Resource (SISTR). This assembles a genome and runs the resulting contigs through https://github.com/peterk87/sistr_cmd to generate the final result.</p> <p>Select</p>	<p>SISTR Pipeline v1.1.1 (PLUGIN)</p> <p>Generates Salmonella in silico typing results from assemblies using the Salmonella In Silico Typing Resource (SISTR). The pipeline assembles a genome from paired reads and runs the resulting contigs through the https://github.com/phac-nml/sistr_cmd tool.</p> <p>Select</p> <p style="text-align: center;">Détermination du sérotype de Salmonella</p>	<p>SNVPhyl Phylogenomics Pipeline</p> <p>Generate a Whole Genome Phylogeny from a set of samples and a reference genome based on Single Nucleotide Polymorphisms (SNVs) using the SNVPhyl pipeline. This will provide a dendrogram as well as a table of all SNVs used and a SNV distance matrix between each sample. An example of how to run SNVPhyl and a description of the parameters can be found at https://irida.corefacility.ca/documentation/user/tutorial/snphy/.</p> <p>Select</p> <p style="text-align: center;">Phylogénie/type génomique</p>

IRIDA: INTEGRATED RAPID INFECTIOUS DISEASE ANALYSIS

15



Génotype vs Phénotype

Génomique et identification/sérotypage/sérogroupage

16

Prédiction du sérotype des *E. coli* producteurs de shigatoxines

Antigènes capsulaires (ex. K1, K12)

Antigènes somatiques (paroi) (O157, O104)

Nomenclature antigénique courante = sérotype
O157:H7
O104:H4

Antigènes flagellaires H (ex. H7, H4)

Agglutination (phénotype)	Antigènes détectés par WGS	Ectyper (génomique)
O157:H7	O157, H7	O157:H7
O157:non mobile	O157, H7	O157:H7
O non sérotypable:H21	O88,H21	Escherichia coli 88:H21

Meilleure précision du sérotype

Pas de données sur les souches non mobiles

17

Prédiction des sérotypes de *Salmonella*

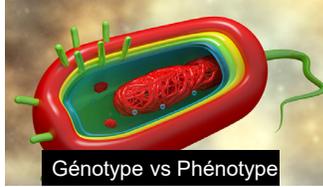
Agglutination (phénotype)	Antigènes détectés par WGS	SISTR (génomique)
I 4,[5],12:i:-	I 4,[5],12:i:1,5	Typhimurium
I 4,[5],12:i:-	I 4,[5],12:i:-	I 4,[5],12:i:-
I 4,12:-:-	I 4,12:fg:[1,2]	Derby
I 6,8:e,h:-	I 6,8:e,h:1,2	Newport

Meilleure précision du sérotype

Données historiques discontinues

*Schéma: Patricia Fields National Salmonella Reference Lab CDC

18



Génomique et antibiorésistance

19

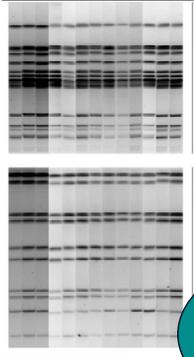
Phénotype versus prédiction génomique

- Le phénotype (CMI) sert à caractériser la réponse de la bactérie vivante en présence d'une concentration déterminée d'un antibiotique pour informer sur le traitement clinique
- La prédiction génomique consiste en la comparaison du génome de la bactérie avec une base de données de gènes de résistances et de mutations (SNP) élaborés à partir de génomes bactériens bien caractérisés (connus).
- La prédiction génomique pourrait avoir des discordances avec le phénotype (CMI):
 - Faux positif : gène présent mais non exprimé
 - Faux négatif : résistance codée par un nouveau gène ou nouveau mécanisme inconnu

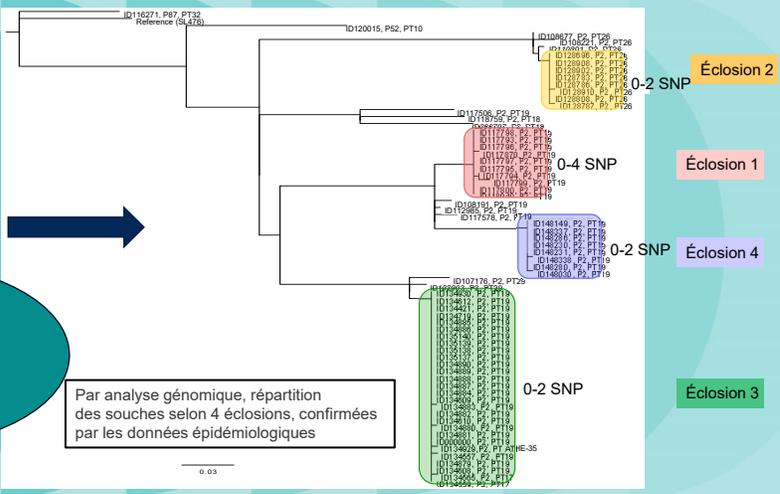
20

Comparaison de 4 éclosions à *Salmonella* Heidelberg par EGCP (PFGE) et par analyse génomique

Par Électrophorèse sur gel en champs pulsés, les souches des différentes éclosions sont identiques



Meilleur pouvoir de discrimination génétique 😊



Par analyse génomique, répartition des souches selon 4 éclosions, confirmées par les données épidémiologiques

Bekal S et al. Usefulness of High-Quality Core Genome Single-Nucleotide Variant Analysis for Subtyping the Highly Clonal and the Most Prevalent *Salmonella enterica* Serovar Heidelberg Clone in the Context of Outbreak Investigations. J Clin Microbiol. 2016

Critères d'interprétation des agrégats

Critères d'interprétation par SNVPhyl et wgMLST

Organisme	Nombre de différences génétiques		
	Isolats reliés	Non concluants	Possiblement non reliés
<i>Listeria monocytogenes</i>	0-10	11-15	>15
<i>Escherichia coli</i> O157:H7	0-10	11-15	>15
<i>Salmonella</i>	0-10	11-20	>20

Zones grises ?

- L'interprétation doit toujours être soutenue par l'enquête épidémiologique
- Ces critères peuvent évoluer selon les données acquises

Nomenclature des agrégats pour PulseNet

Année/mois/organisme/WGS- numéro séquentiel/province ou multi-province

Exemples: 1809StahWGS-1QC
1901SENWGS-4MP
1905SINWGS-1MP
1903SSWGS-2MP
1903SFWGS-1QC

Ces codes d'agrégats pourraient être modifiés dans le temps, à cause de l'évolution des critères d'interprétation

25

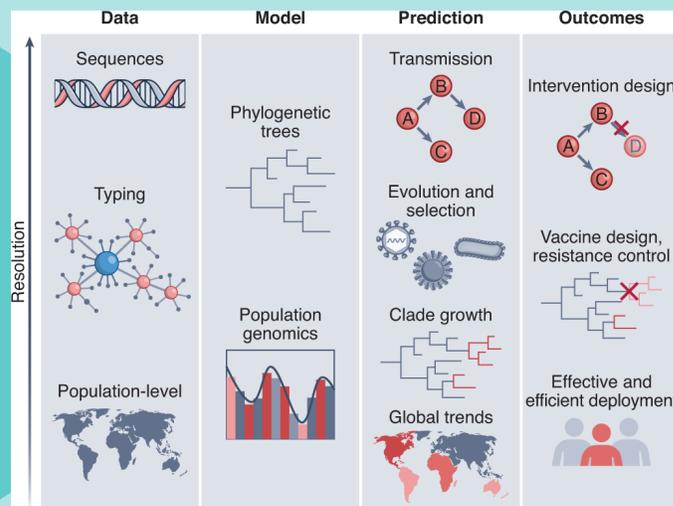
Interface génomique/maladies infectieuses/santé publique

The potential of genomics for infectious disease forecasting

Jessica E. Stockdale, Pengyu Liu & Caroline Collin

Nature Microbiology, 7, 1736–1743 (2022) | [Cite this article](#)

17k Accesses | 21 Citations | 89 Altmetric | [Metrics](#)



<https://www.nature.com/articles/s41564-022-01233-6>

26

26



27