



HAL
open science

Epidémiologie des Salmonelles en filière animale par approche génomique

Valérie Michel, Madeleine De Sousa Violante, Carole Feurer, Michel-Yves Mistou, Ludovic Mallet

► **To cite this version:**

Valérie Michel, Madeleine De Sousa Violante, Carole Feurer, Michel-Yves Mistou, Ludovic Mallet. Epidémiologie des Salmonelles en filière animale par approche génomique. Innovations Agronomiques, 2023, 88, pp.178-191. 10.17180/ciag-2023-vol88-art15 . hal-04324486

HAL Id: hal-04324486

<https://hal.inrae.fr/hal-04324486>

Submitted on 5 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Epidémiologie des Salmonelles en filière animale par approche génomique

Michel Valérie¹, De Sousa Violante Madeleine¹, Feurer Carole², Mistou Michel-Yves³, Mallet Ludovic⁴

¹Département de Microbiologie Laitière, ACTALIA, 74800 La Roche sur Foron, France

²IFIP–Institut du Porc, La Motte au Vicomte B.P. 35104, 35651, Le Rheu Cedex, France

³INRAE, MaIAGE, Université Paris-Saclay, F-78352, Jouy-en-Josas, France

⁴Institut Universitaire du Cancer Toulouse - Oncopole, 31059 Toulouse, France

Correspondance : v.michel@actalia.eu

Résumé :

Le projet EMISSAGE a étudié la diversité génomique de souches françaises de 3 sérovars de *Salmonella enterica subsp. enterica* d'importance majeure en filière agro-alimentaire. Les génomes de 267 souches de *S. Typhimurium* et son variant monophasique issues d'abattoirs porcins de la région Grand Ouest et de 148 souches de *S. Mbandaka* isolées de fermes et d'ateliers laitiers du Nord-Ouest ont été séquencés. Leur diversité génomique a été analysée soit sur le coregénom (la part du génome partagée par l'ensemble des souches du sérovar étudié) avec un pipeline historique (iVARCall2), soit sur le pangénom (l'information génomique totalisée par l'ensemble des souches) avec un pipeline nouvellement développé (pgSNP). Ces sérovars montrent des diversités génomiques différentes. Les identifications de marqueurs génomiques spécifiquement liés à l'origine géographique ou de la spécificité d'hôte ne se sont pas avérées assez discriminantes pour une identification rapide de l'origine des souches sur le terrain.

Mot clés : *Salmonella*, abattoirs porcins, ateliers laitiers, diversité génomique

Abstract : Epidemiology of Salmonella in the animal chain using a genomic approach

During the EMISSAGE project, genomic diversity of French *Salmonella enterica subsp. enterica* strains isolated from two food sectors was investigated. The genomes of 267 strains of *S. Typhimurium* and its monophasic variant isolated from western pig slaughterhouses and of 148 strains of *S. Mbandaka* originating from North-West dairy farms and processing factories were sequenced. Their genomic diversity was characterized either based on coregenome data (genomic information common to the whole set of strains of a same serovar) with an historical pipeline (iVARCall2), and based on pangenome (the sum of the genomic information of the strains of a same serovar) with a newly developed pipeline (pgSNP). These serovars harbored different genomic diversity. The identifications of genomic markers of country location (for the pork serovar) and animal origin (for the dairy serovar) were not specific enough to be applied on a routine basis to rapidly identify the origin of the studied serovars.

Keywords: *Salmonella*, pork slaughterhouses, dairy farms, dairy units, genomic diversity

Introduction

Salmonella est l'un des 4 principaux agents pathogènes responsables de toxi-infections alimentaires dans le monde (WHO, 2015). En France, le nombre de personnes atteintes d'une salmonellose a été estimé à 200 000 par an en 2016 et bien que la plupart des cas soient bénins, on dénombre 4 000 hospitalisations et 72 décès en 2016 (Van Cauteren *et al*, 2017). En 2019, les charcuteries étaient suspectées d'être à l'origine de 14% des cas de toxi-infections collectives alimentaires suspectées ou confirmées en *Salmonella* tandis que les fromages et produits laitiers étaient suspectés pour 5 % des cas (Fournet *et al*, 2021). La surveillance de *Salmonella spp.* est accrue dans les principales filières alimentaires, en particulier dans la filière porcine et la filière lait et produits laitiers, touchées ces dernières années par plusieurs crises sanitaires liées à ces deux pathogènes (Bone *et al*, 2010 ; Gossner *et al*, 2012 ; Jourdan-DaSilva et Le Hello, 2012 ; Ung *et al*, 2019). Pour renforcer cette surveillance et dégager des moyens de gestion, les opérateurs des filières concernées, de la production primaire à la transformation, ont besoin de caractériser plus finement ces souches. Les liens génétiques entre les souches permettent de distinguer d'éventuels clusters (groupes de souches à forte parenté génomique) dominants, d'identifier des pistes de dissémination et de circulation par le lien établi avec les données épidémiologiques des souches concernées (Plateforme de Surveillance de la Chaîne Alimentaire, 2021) ou d'investiguer l'origine des souches. L'exploitation de l'information génomique globale des souches a été rendue possible par le développement des techniques de séquençage à haut débit. Ce séquençage global des souches, appelé selon l'acronyme anglais WGS pour Whole Genome Sequencing (WGS) est couramment utilisé en France et dans le monde pour le typage génétique des souches humaines (EFSA, 2018). Son utilisation dans le domaine de la production agricole et agro-alimentaire reste encore partielle et demande à être développée pour permettre aux acteurs de s'emparer de ces technologies innovantes, dans l'objectif d'avoir une meilleure caractérisation des souches mais également pour exploiter la connaissance de ces génomes à d'autres fins (identification de marqueurs génétiques de résistance, présence de gènes de pathogénicité, identification de séquences génétiques pouvant servir de marqueurs spécifiques...). L'un des besoins des opérateurs des filières agro-alimentaires concernées est d'être en mesure de caractériser rapidement l'origine (géographique, source) de toutes souches nouvellement isolées.

Plus précisément, pour la filière porcine, derrière l'origine s'entend l'origine géographique de la souche, de manière à pouvoir rapidement remonter à l'élevage de porcs porteur de la contamination en *S. Typhimurium* ou son variant monophasique (TMV). Pour la filière laitière, la notion d'origine s'adresse plutôt à la notion de source de contamination (faune sauvage, aliments du bétail, etc...).

C'est dans ce contexte que les trois partenaires du projet EMISSAGE (CASDAR-RT n°1701) ont uni leurs compétences scientifiques et techniques pour progresser sur la connaissance de 3 sérovars de *Salmonella* d'importance majeure en filières animales. Il s'agissait pour la filière porcine des sérovars *Salmonella* Typhimurium et de son variant mono-phasique (TMV) S. 1, 4, (5), 12 : i : - et pour la filière laitière du séovar Mbandaka. Le partenaire scientifique du projet EMISSAGE était le laboratoire de Sécurité des Aliments (LSAL) de l'Anses de Maisons-Alfort dont la mission GAMER (Genome Analysis Modelling and Risk) a participé au développement méthodologique nécessaire à l'interprétation des données génomiques générées, ainsi qu'à leur interprétation. Les deux partenaires techniques associés au projet EMISSAGE étaient l'Ifip (Institut Technique Agricole et Institut Technique Agro-Industriel de la filière porcine) et ACTALIA (Institut Technique Agro-Industriel des produits laitiers et de la Sécurité des Aliments), avec l'implication respective d'un chef de projet en microbiologie moléculaire et du responsable du laboratoire de microbiologie laitière.

Les objectifs principaux du projet EMISSAGE étaient de :

- Constituer une collection de référence de génomes des 3 sérovars de *Salmonella* issues du terrain par WGS ;

- De développer un ensemble de pipelines bio-informatiques automatisés de traitement et d'analyse des données génomiques permettant leur transfert au sein des instituts techniques, afin qu'ils les mettent au service des professionnels,
- D'identifier des marqueurs génétiques épidémiologiques et de risques permettant la traçabilité simple et rapide des isolats issus des auto-contrôles.

1- Démarche mise en œuvre

Le projet EMISSAGE a été organisé en plusieurs actions et tâches, reprises ci-après en autant de paragraphes :

1.1 Constituer une collection de souches de *Salmonella* isolées d'atelier de production, de découpe et de transformation agro-alimentaire.

Afin de récupérer des souches d'origine porcine, l'Ifip a réalisé 9 campagnes de prélèvements de souches dans des abattoirs de la région Grand-Ouest, 6 ayant été effectuées en 2018 et 3 en 2019. Des prélèvements sur des carcasses et des langues de porcs et leur environnement d'abattage (travées, fendeuses, goulottes, podium) ont été analysés pour la présence de *Salmonella*, avec la méthode alternative validée AFNOR BRD 07/11-12/05, au laboratoire de l'Ifip à Maisons-Alfort.

Pour chaque souche issue du porc, le département d'élevage a été recensé. Afin d'augmenter le panel de souches issues d'abattoir à séquencer, l'Ifip a pris contact avec la DGAI (Direction Générale de l'Alimentation) pour obtenir les souches isolées des plans de surveillance nationale de la DGAI entre 2017 et 2019 sur carcasses de porc, ainsi que leurs métadonnées. L'Ifip a centralisé les métadonnées reçues sous format informatique, de manière à établir un fichier de méta-données commun aux souches isolées de ses campagnes de prélèvements et mises à disposition par la DGAI.

Pour la filière laitière, ce sont des filières laitières du Nord-Ouest de la France et produisant des fromages au lait cru qui ont été sollicitées pour la mise à disposition de souches de *S. Mbandaka* issues tout au long de la chaîne de production et de transformation laitière. Les métadonnées (données écologiques) de ces souches demandées étaient des méta-données minimales : date isolement, code ferme, nature du prélèvement, département. ACTALIA a assuré la récupération des souches auprès des opérateurs laitiers, après avoir établi un accord de transfert mutuel (accord régissant les conditions de mise à disposition et d'utilisation des souches et des données qui en découlent) et a également centralisé les méta-données associées.

1.2 Obtenir le génome complet des souches

Pour ce faire, les ADN totaux des souches ont été extraits par Ifip et ACTALIA, chacun traitant leurs souches respectives. Les extractions et contrôle qualité des ADN extraits ont été réalisés selon le même protocole avant d'être séquencés par la technologie Illumina pour produire des lectures de séquençage par paires, comme décrit par Radomski *et al* (2019).

Après séquençage et contrôle qualité des séquences obtenues et vérification des sérovars, le jeu de données de génomes disponibles pour le projet était de 140 génomes pour *S. Mbandaka* et 267 génomes pour *S. Typhimurium* et son variant monophasique. Les génomes ont été assemblés selon le workflow développé par l'Anses et dénommé ARTwork (Vila Nova *et al*, 2019).

1.3. Développer un pipeline bio-informatique pour analyser le pangéome

L'analyse des séquences d'ADN dans un but de typage et de définition de proximité génétique entre souches d'un même sérovar est le plus souvent réalisée sur le coregénom, qui est la partie du génome conservée par l'ensemble des souches du panel étudié. Cette analyse, déjà très discriminante, porte sur une partie du génome disponible (entre 70 et 80% du génome pour *S. Typhimurium* (Fu *et al*, 2017)). En incluant le génome accessoire (c'est-à-dire la part du génome compris dans l'ensemble des souches), le pangéome est ainsi analysé. Pour ce faire, un pipeline d'analyse du pan-génom, dénommé pgSNP a été développé spécifiquement au cours du projet EMISSAGE et validé sur des jeux de données issues de toxi-infections alimentaires collectives (TIAC).

1.4. Procéder à l'analyse génomique des souches

Dans un premier temps, une analyse phylogénomique de chaque sérovar a été menée de manière à :

- Etudier la diversité génétique des souches ;
- La comparer à celle de souches du même sérovar mais issues d'un contexte différent (autre département pour les sérovats d'intérêt pour la filière porc, autre origine pour le sérovar d'intérêt laitier).

Tous les génomes ont été caractérisés *in silico* par MLST (MultiLocus Sequence Type) basé sur les séquences des 7 gènes de ménage (*aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA*, et *thrA*) décrits dans la base de données PubMLST (Jolley et Maiden, 2010). Le Séquence Type (ST) de chaque génome a été obtenu avec l'outil MLST Tseeman (<https://github.com/tseemann/mlst>).

La comparaison a été menée d'abord par une analyse reposant sur le coregénom. Les inférences phylogénomiques ont été réalisées sur un modèle d'évolution calculé par IQ-TREE (Nguyen *et al*, 2015) (inférence par maximum de vraisemblance) sur les SNP. Le terme SNP correspond à Single Nucleotide Polymorphism pour « polymorphisme basé sur un unique nucléotide », ce qui consiste à compter le nombre de nucléotides différents entre deux séquences alignées. Le workflow iVARCall2 (Lee *et al*, 2015) a été utilisé dans cet objectif. Les arbres phylogénétiques ont été ensuite visualisés et annotés en utilisant l'outil interactif Tree of Life (iTOL, <https://itol.embl.de/>), selon Letunic et Bork (2016).

1.5. Rechercher des marqueurs spécifiques de ces sérovats pour une caractérisation plus rapide de leur origine

L'existence de marqueurs génétiques spécifiques à l'origine géographique ou à l'origine liée à l'hôte bovin ou volaille a été recherchée, respectivement pour les 2 sérovats d'intérêt pour la filière porcine et pour le sérovar d'intérêt pour la filière laitière. Cette recherche a été menée *in silico* par une approche d'association de génomes en comparant pour les sérovats *S. Typhimurium* et TMV, les 267 génomes récupérés par l'Ifip à 83 génomes de souches françaises issues d'abattoirs porcins des mêmes sérovats isolées d'autres départements. Ces souches ont été fournies et leur génome séquencé par le partenaire ANSES. Pour le sérovar *S. Mbandaka*, 164 génomes de souches issues de volailles, isolées dans les départements identiques ou proches sur le plan géographique, ont été mis à disposition pour le projet EMISSAGE par l'Anses (Unités HQPAP de Ploufragan et LNR antibiorésistance de Ploufragan). La récupération de souches supplémentaires des 3 sérovats étudiés a été possible grâce à la mobilisation des laboratoires partenaires du réseau *Salmonella* de l'Anses et de leur volonté de coopérer au recensement et à la fourniture de souches et des méta-données minimales associées.

La recherche de marqueurs (gènes ou variants) uniques ou en combinaison ont été faites.

Pour les gènes uniques, les fichiers GFF produits par le logiciel Prokka (<https://github.com/tseemann/prokka>) ont été utilisés par l'outil Panaroo (Tonkin-Hill *et al*, 2020), outil d'extraction du pangéome permettant de prendre en compte les erreurs introduites par l'annotation des génomes. Le seuil d'identité au sein des regroupements de gènes était de 90%. Pour les variants uniques, les fichiers VCF issues de iVarCall2 ont été fusionnés pour identifier la présence ou l'absence de variants dans chaque génome. La position des variants a été définie vis-à-vis de la souche *Salmonella* Mbandaka SA20026234 et de la souche *S. Typhimurium* LT2 prises comme référence.

Pour explorer les possibles combinaisons de marqueurs, un script sous Python appelé MarkerFindr a été développé. Il calcule la combinaison de 3 gènes ou variants au maximum pour éliciter celui donnant le meilleur score de discrimination basé sur le critère recherché (origine géographique, origine animale). Les scores d'aptitude se basent sur des valeurs de spécificité et sensibilité. L'outil développé MarkerFindr est disponible en ligne : <https://github.com/madeleinevlt/MarkerFindr>. Les gènes ou variants identifiés comme discriminants ont été annotés en utilisant les outils Blast et Uniprot (Uniprot 2021).

2 Résultats obtenus

2.1 Une collection de souches de *Salmonella* des 3 sérovars étudiés issues de bassins de production

Pour la filière porcine, l'Ifip a réalisé un total de 871 prélèvements en abattoirs au cours de 9 campagnes de prélèvements, réalisées dans 6 abattoirs différents notés de A à F dans le Tableau 1. 116 souches de *Salmonella* ont ainsi été isolées soit de l'environnement de l'abattoir, soit directement des animaux.

Tableau 1 : Nombre de souches de *Salmonella* isolées des campagnes menées en abattoirs par l'Ifip en 2018 et 2019.

Abattoir	Nb positifs	Dont environnement	Dont Porc (carcasse ou langue)	Nombre de souches récoltées
A	2/138	0	2	4
B	3/140	2	1	5
C	14/138	8	6	14
D	12/150	12	0	35
E	28/149	14	14	47
F	6/156	4	2	11
Total	65/871	40/105	25/766	116
Taux de positifs	4,1% [3,4-4,9]	38% [29,3-47,7]	3,2% [2,2-4,7]	

Pour chaque souche issue du porc, le département d'élevage a été recensé.

Afin d'augmenter le panel de souches issues d'abattoir à séquencer, l'Ifip a pris contact avec la DGAI pour récupérer les souches isolées des plans de surveillance nationale de la DGAI entre 2017 et 2019 sur carcasses de porc. 111 souches des sérovars d'intérêt (24 de *S. Typhimurium* et 87 de son TMV) ont été transmises par la mission antibiorésistance du Laboratoire de Sécurité des Aliments (LSAL) de l'Anses et de certains LDA à l'Ifip. **267 souches appartenant aux deux sérovars d'intérêt pour la filière porcine furent disponibles pour le projet EMISSAGE.** L'Ifip a centralisé les méta-données reçues sous format informatique, dans un fichier de méta-données commun aux souches isolées de ses campagnes de prélèvements et mises à disposition par la DGAI.

Pour le sérovar *S. Mbandaka*, des souches des fermes laitières ou de leur environnement, ainsi que des souches isolées dans les ateliers de production ou dans les produits finis (fromages au lait cru) ont été recensées. **Les 148 souches de *S. Mbandaka* constituant le noyau de souches terrain étaient issues de 23 fermes laitières, livrant 3 ateliers de production différents**, isolées entre les années 2016 à 2019. Les méta-données (données écologiques) de ces souches étaient des méta-données minimales : date isolement, code ferme, nature du prélèvement, département. A cette centralisation, s'est ajoutée la structuration des méta-données des souches sous format informatique.

2.2 Développement et validation du pipeline bio-informatique pgSNP, pour une discrimination des souches basée sur le pangéome

Le pipeline bio-informatique développé avait pour objectif de prendre en compte les variants du coregénom (c'est-à-dire les régions homologues) et du génome accessoire (c'est-à-dire les régions non-homologues) qui peuvent expliquer certains processus d'adaptation à des pressions sélectives. Son développement a consisté à reconstruire des inférences phylogénomiques sur la base à la fois des variants du core-génom et du génome accessoire, ce qui est habituellement réalisé indépendamment par différentes méthodes analytiques. Pour chaque sérovar, un pangénom de référence a été construit à partir d'un assemblage de génomes considérés. Le pangénom de référence représente donc au final toutes les séquences génomiques des échantillons. **Durant ces développements, il a été montré que 50% du génome de toutes les souches de *S. Typhimurium* et son TMV ne sont pas pris en compte lors des analyses focalisant uniquement sur le core-génom.** De plus, dans ce génome accessoire, de nombreux éléments adaptatifs ont été détectés, dont certains éléments partagés par 99% des souches françaises, comme des fragments génomiques contenant des gènes de résistance aux métaux lourds.

Ensuite, le mapper BWA et le variant caller FreeBayes (intégrés dans le workflow Snippy) ont permis respectivement d'aligner les données de séquençage génomique global au pangénom de référence et identifier les variants des génomes à l'échelle du pangénom. Un pseudogénom contenant les variants pour chaque échantillon est par la suite construit et cet alignement est découpé en morceaux pour inférer des phylogénies pour chacun de ces derniers. Les arbres phylogénomiques sont produits avec IQ-Tree pour chaque morceau FastRFS. Des vérifications ont été réalisées sur chaque étape du pipeline, comme la présence ou non de matériel génétique redondant sur la référence reconstruite *de novo* ou la comparaison des arbres phylogénétiques obtenus selon la méthode utilisée. ERaBLE, qui est un estimateur de longueur de branche phylogénétique sur des supertrees (l'arbre final obtenu avec le pipeline pgSNP), a été testé et implémenté au pipeline. **Le pipeline bio-informatique obtenu, dénommé pgSNP (Figure 1) a ensuite été validé sur 3 jeux de données différents issues de TIAC.**

Le premier jeu de données est un jeu de données de *S. Typhimurium* et son variant monophasique qui contient 192 souches avec 4 clusters épidémiologiques, publié par Radomski *et al* (2019). Il a été tout d'abord montré sur ce jeu de données que pgSNP ajoute environ 1,6 Mb d'information génétique par rapport à l'analyse basée sur le core-génom. Par rapport aux clusters épidémiologiques, pgSNP est capable d'identifier et de grouper les souches épidémiques entre elles, à l'exception de trois souches. Deux d'entre-elles ont déjà été décrites dans le papier, tandis que la troisième provient d'une souche de TMV qui se retrouve reliée à un cluster de *S. Typhimurium* dans l'arbre pangénomique. Ce résultat est appuyé par le peu de distance entre les souches. Il a également été montré que pgSNP induit des différences topologiques sur les variants monophasiques, qui ont peu de différences au niveau du core-génom, par rapport aux *S. Typhimurium* dont la topologie reste préservée entre les deux méthodes. Ce résultat montre que l'ajout du génome accessoire impacte principalement des souches avec peu de variabilité sur le core-génom, même si les méthodes de core-génom évaluent la topologie sur environ 80% des données. Globalement, pgSNP obtient des résultats concordants avec les données épidémiologiques sur ce jeu de données.

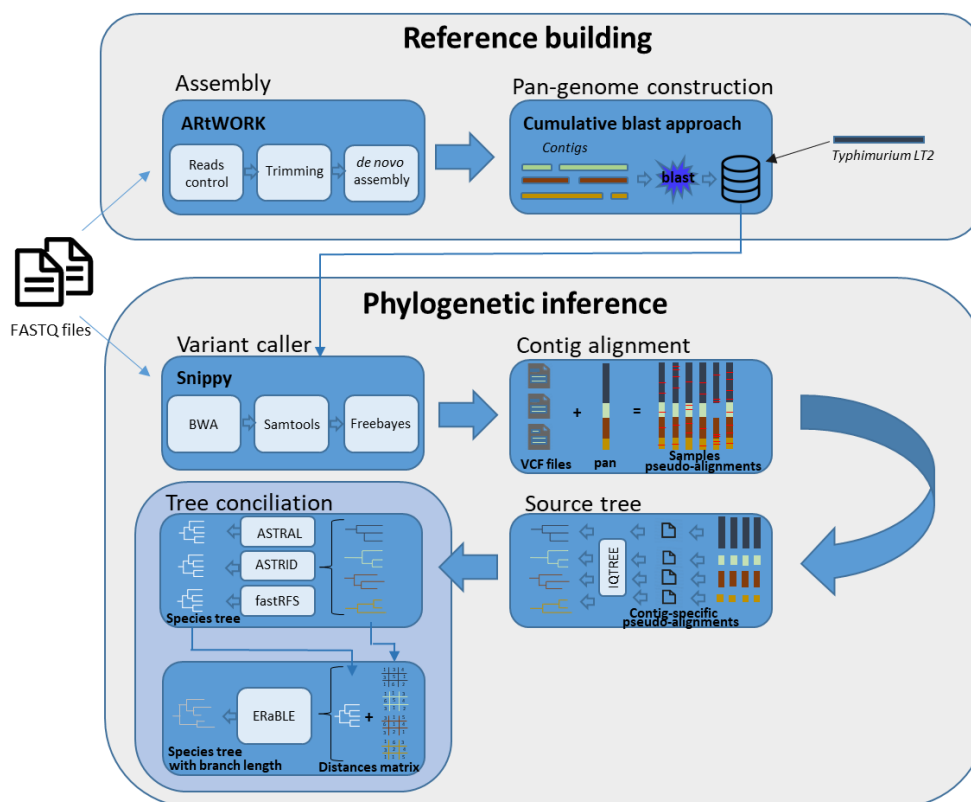


Figure 1. Représentation schématique du pipeline d'analyse pangénomique, pgSNP

Le deuxième jeu de données correspond à un jeu de données de *Escherichia coli* O157:H7 publié par Rumore *et al* (2018). Ce jeu de données contient 210 souches d'origine humaine, avec 8 clusters épidémiologiques qui ont très peu de différences sur le core-génome (< 5 SNPs). Avec pgSNP, la plupart des clusters épidémiologiques identifiés sont retrouvés en adéquation avec les résultats publiés par les auteurs. En revanche, l'utilisation du pangénome de référence révèle un changement topologique d'une souche épidémique avec des souches sporadiques. Ce changement topologique est dû à la présence de 206 kb d'ADN core-génome (plasmides, ADN chromosomique) présent dans le pangénome de référence mais absent dans la référence de la publication. PgSNP démontre ici l'avantage de l'utilisation d'un pangénome de référence lors d'une mauvaise sélection de génome de référence.

Enfin, pgSNP a été testé sur un jeu de données de clusters épidémiologiques de *Neisseria meningitidis*. Comparé aux deux autres jeux de données, *Neisseria meningitidis* est très recombinaut, avec un génome de petite taille. Ce jeu de données contenait également des informations précises sur les souches sporadiques. Avec pgSNP, il a été montré de nouvelles réconciliations pour quelques souches. Dans l'ensemble, les pgSNP ajoutent une distance génétique entre les souches épidémiologiques, notamment en raison des éléments génétiques mobiles et de la nature recombinaut de *Neisseria meningitidis*. Ces changements permettent de relier des souches sporadiques et des souches épidémiques qui ont des méta-données très similaires, et qui donc pourraient potentiellement être reliées. pgSNP a également montré que les distances génétiques supplémentaires peuvent rendre les enquêtes sur les foyers épidémiques difficiles. Cependant, les échantillons de foyers, dont le core-génome et le génome accessoire sont proches, sont toujours regroupés dans l'arbre pangénomique, ce qui démontre l'importance du génome accessoire dans les enquêtes épidémiologiques.

2.3 Description de la diversité génomique des souches de *Salmonella* Typhimurium et de son variant monophasique (TMV)

La diversité génomique des souches porcines de ces sérovars a été décrite (Figure 2). La différence génomique entre les génomes des variants monophasiques est très faible, l'hypothèse pourrait être émise qu'un seul clone est disséminé en France. En regardant plus précisément, la différence moyenne de SNP des 152 échantillons de TMV est de 64 SNPs. Topologiquement, un nœud interne a divisé les 152 TMV en deux groupes de 104 et 48 échantillons de TMV, avec une moyenne intra-groupe de 49 et 51 SNPs, respectivement. Cette faible diversité des TMV peut être expliquée par l'apparition récente de ce variant, comparé aux souches de *S. Typhimurium*. En examinant le contenu génomique, le nombre élevé de gènes de résistance aux antibiotiques, aux métaux lourds et aux biocides semblent associés à la prévalence de ces deux sérovars dans les élevages porcins et pourrait expliquer leur adaptation.

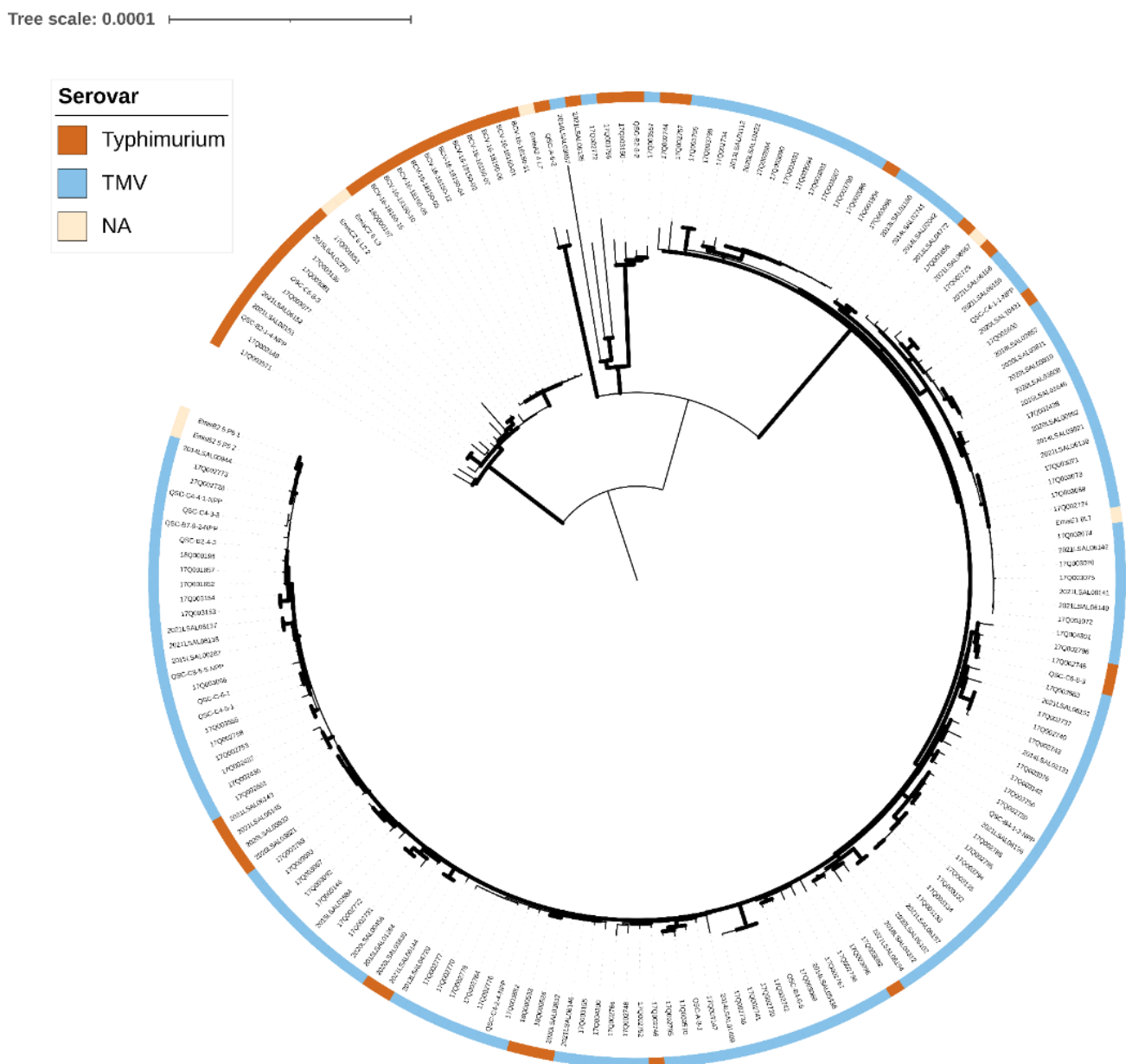


Figure 2. Arbre phylogénomique des souches de *S. Typhimurium* (orange) et de son TMV (en bleu)

Le questionnement principal de la filière était de savoir s'il existe un lien entre les diversités géographiques et les diversités génomiques chez *S. Typhimurium* et TMV. Pour cela, 188 souches provenant de 3 régions productrices de porc ont été sélectionnées. Certaines souches proviennent des élevages porcins, mais la plupart sont des souches issues de prélèvements faits à l'abattoir avec l'identification du département de provenance de l'animal. **Ce jeu de données a permis de mettre en lumière une répartition dans toute la France de ces souches, sans adaptation particulière à la géographie** (Figure 3).

Tree scale: 0.0001

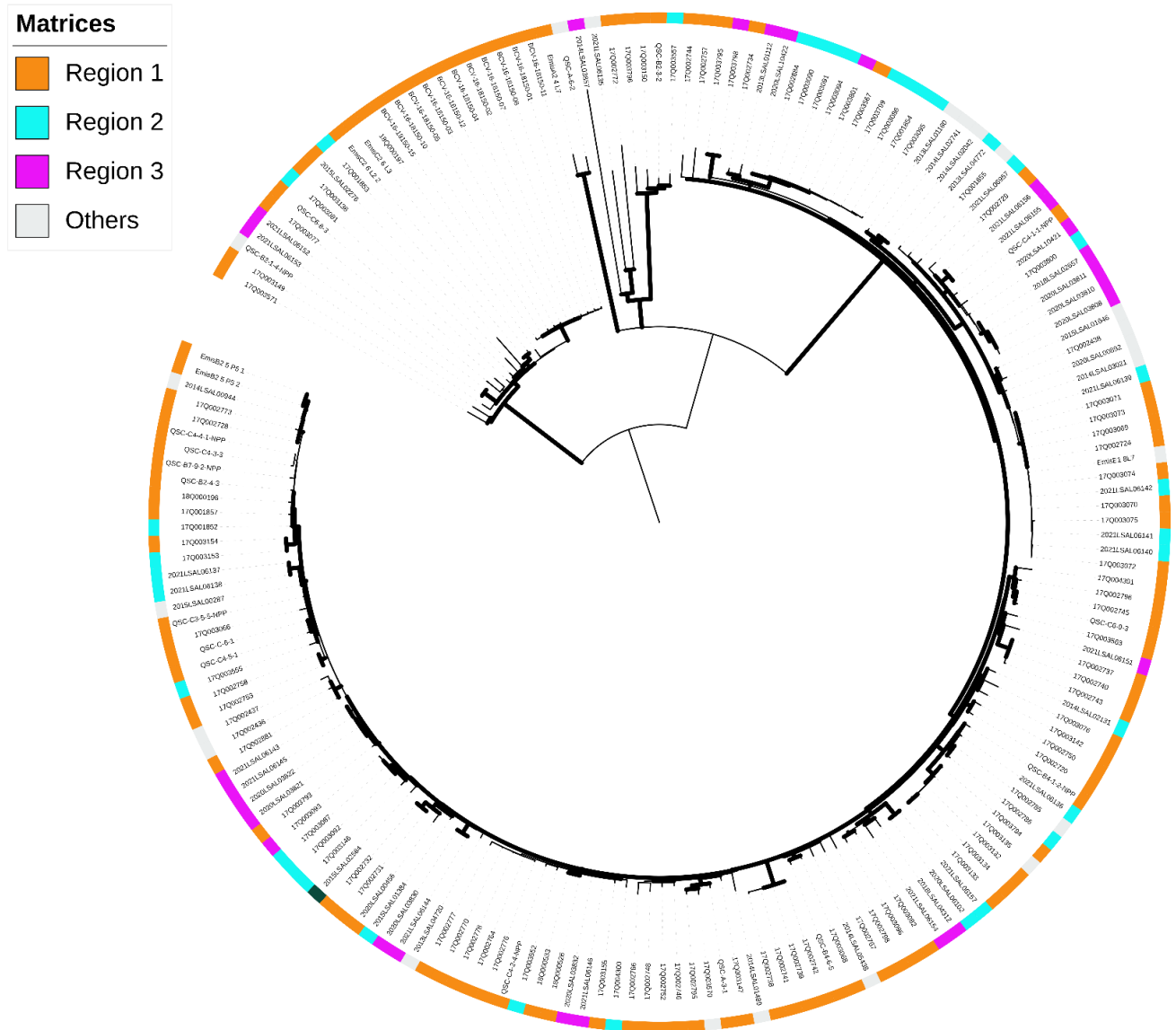


Figure 3. Arbre phylogénomique des souches de *S. Typhimurium* et de son TMV. Le cercle extérieur figure la localisation des souches selon leur région d'isolement.

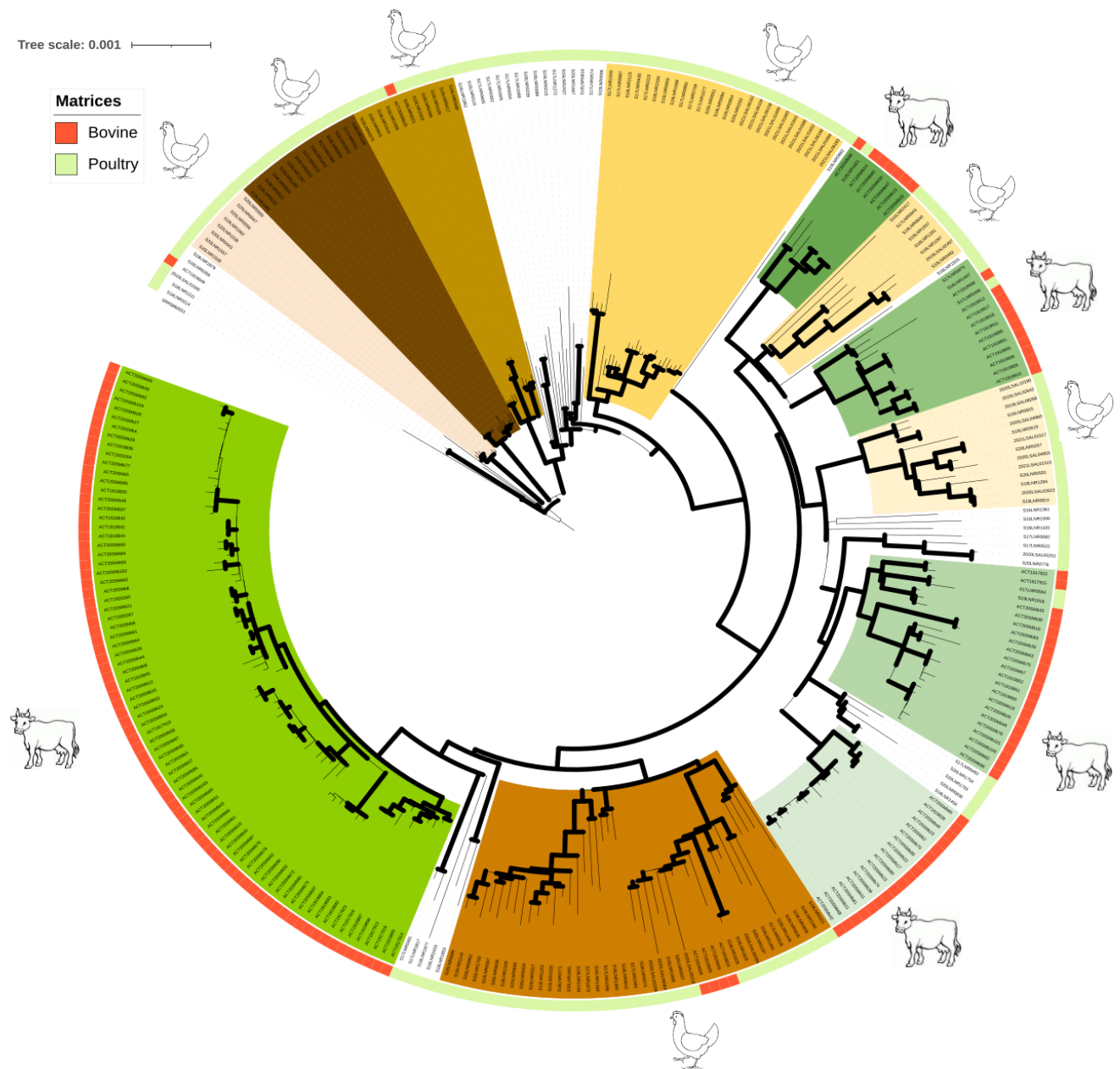
L'analyse des variants ou des gènes n'a pas mis en évidence de discrimination entre les échantillons provenant de différentes régions. pgSNP a également été appliqué à ce jeu de données, afin de vérifier si l'adaptation géographique n'était pas localisée dans le génome accessoire de ces sérovars. Grâce au

pipeline, le contenu du génome accessoire a été rapidement identifié, avec la présence de phages et de plasmides sur certaines souches. Les analyses ont été réalisées sur 30% de bases supplémentaires pour la reconstruction de l'arbre phylogénomique. Ces nouvelles réconciliations, obtenues grâce au génome accessoire, n'ont pas pu être reliées à la géographie.

2.4 Description de la diversité génomique des souches de Salmonella Mbandaka

Toutes les souches de S. Mbandaka isolées en France appartiennent au même profil MLST : ST413. L'arbre phylogénomique inféré sur les souches de volailles et laitières a montré qu'il n'y a pas de spécificité à l'hôte, mais plutôt des clades aviaires intercalés par des clades bovins (Figure 4). **Sur les différents clades d'hôtes, il n'y a pas de spécificité de matrices, que ce soit entre les échantillons de volailles isolés des poules pondeuses et des poulets à chair ou les matrices de la filière laitière bovine.** Des échantillons de dinde sont disséminés tout autour de l'arbre, mais avec une longueur de branche unique, et ont donc été identifiés comme des singletons. Les méta-données géographiques révèlent que la plupart du temps, les échantillons de volailles sont regroupés avec des isolats provenant de la même région ou des mêmes régions limitrophes. Cependant, dans la partie supérieure de l'arbre phylogénomique, les souches de différentes régions sont regroupées ensemble, sans lien de proximité géographique entre les régions. Cette faible diversité entre les souches de différentes sources géographiques concerne presque exclusivement les isolats de volailles, ce qui nous permet d'émettre l'hypothèse de l'existence d'une source de reproduction commune. En effet, les parents se trouvent dans les élevages et s'occupent des naissances des poules, qui sont ensuite triées en poulets de chair et poules pondeuses. Ces poules sont ensuite redistribuées dans les fermes, et peuvent propager la maladie à partir de la même source, ce qui expliquerait ce clade.

Figure 4. Arbre phylogénomique des souches de *S. Mbandaka*. Le cercle extérieur figure l'origine bovin lait (orange) ou volaille (vert) des souches. Les clades d'hôte sont surlignés en couleurs.



L'outil bio-informatique pgSNP a également été appliqué à ce jeu de données, et malgré le nombre de bases supplémentaires analysées sur l'arbre phylogénomique (+ 27%), très peu de différences sur les clades d'hôtes déjà identifiés grâce au coregénom ont été trouvées. Enfin, la piste d'une contamination commune entre les deux filières a été investiguée, sur la base des observations des opérateurs laitiers de la région Nord-Ouest. La piste de la transmission par voie aérienne a été proposée, à partir de données publiques de souches d'oiseaux sauvages américains. **La proximité de ces souches avec des souches de volailles proches des côtes, mais également d'une souche de fumier du Nord-Ouest suggère la possible contamination par la faune sauvage.** Des données supplémentaires européennes sont nécessaires afin de valider ces hypothèses. Une étude de cette filière, en lien avec des pays frontaliers

pourrait valider cette piste de contamination par la faune sauvage, qui a déjà été observée chez *S. Typhimurium* avec le portage de rongeur.

2.5 Recherche de marqueurs génétiques associés à une localisation géographique des souches de *S. Typhimurium* et son TMV

Une analyse comparative génomique a été menée *in silico* pour tenter d'identifier des marqueurs génomiques ou couples de marqueurs génomiques qui seraient spécifiques soit de l'origine des souches pour *S. Mbandaka* soit de leur origine géographique pour *S. Typhimurium*. De tels marqueurs ont pu être identifiés mais leur opérationnalité reste faible. Dans l'objectif d'être en mesure de déployer un test PCR qui reste relativement simple et de coût financier raisonnable, il a été décidé que le nombre de marqueurs discriminants ne devait pas excéder 3. Ainsi par exemple, pour la filière laitière, le jeu de données analysées différenciait 12 clusters de souches différents isolés de volaille ou de bovin (Figure 4). Il existe des marqueurs spécifiques des souches bovines ou volailles mais pas de marqueurs qui soient discriminants à 100%. La combinaison de 2 ou 3 marqueurs ne permet jamais d'atteindre une spécificité et une sensibilité dépassant 95% pour les deux critères, conduisant à une justesse de prédiction comprise entre 80% et 95%, ce qui n'est pas suffisant. L'intégration de marqueurs supplémentaires, appartenant au génome accessoire, n'a pas permis d'obtenir des marqueurs discriminants avec des critères de performance suffisant. La même approche a été utilisée sur les sérovars porcins pour tenter d'associer des marqueurs génétiques spécifiques à une région ou un département. Cette diversité française a pu être caractérisée en utilisant une combinaison de gènes et de variants, qui est capable de discriminer les souches françaises avec une précision de 86%. Les critères de performance n'ont pas été jugés suffisamment discriminants pour être applicables en routine. Une analyse plus poussée des variants accessoires pourrait également révéler une meilleure précision.

En conclusion, on peut dire qu'il **est possible de définir des marqueurs plus ou moins associés aux critères de discrimination souhaités par les filières. Cependant, les conditions de leur mise en œuvre pour les rendre abordables sur le plan économique et pertinents d'un point de vue technique (justesse de prédiction) ne sont pas remplies.** Des tests PCR d'application n'ont donc pas été développés.

Conclusion – Perspectives du projet

Le projet CASDAR-RT EMISSAGE mené entre un partenaire académique, l'Anses de Maisons-Alfort et deux instituts techniques (Ifip et ACTALIA), a permis à l'ensemble de ses partenaires d'avoir une meilleure connaissance de 3 sérovars de *Salmonella enterica subsp. enterica*, d'importance majeure dans les filières porc et lait. **La diversité génomique des souches circulant en France au niveau des abattoirs porcins pour les sérovars Typhimurium et son variant monophasique, ainsi que celle des souches bovines dans la région Nord-Ouest de production de fromages au lait cru et dans la filière volaille a été caractérisée pour la première fois. Ce travail a été mené sur un nombre conséquent de souches, issu d'un effort important d'isolements en conditions de terrain de production.** Un outil de traitement des données génomiques, visant à maximiser l'information génétique en se basant sur l'ensemble de l'information génomique (le pangenome) a été développé pour comparer des souches entre elles : c'est le pipeline de bio-informatique pgSNP. La pertinence du pipeline pgSNP a été validé sur des jeux de données issues de TIAC et, en maximisant, l'information génétique disponible, il se révèle très discriminant. Cette approche utilisant le pangenome pour différencier des souches est en plein développement actuellement et pgSNP constitue une avancée importante dans ce domaine.

Cependant, **la connaissance de la diversité génomique et de la proximité des souches au sein d'un même sérovar (pour le variant monophasique de Typhimurium) ou entre souches bovines et**

aviaires sont des éléments importants pour mieux investiguer les contaminations du terrain. La mise en évidence par le pgSNP d'un clone unique du TMV circulant dans la filière porcine et spécifiquement français est un apport de connaissance majeure lors d'investigations épidémiologiques. L'hypothèse d'une contamination inter-espèce entre volaille et bovin est envisagée pour certains clusters de souches de *S. Mbandaka*. La mise à disposition des données écologiques de ces souches (contexte exact de prélèvements et d'isolement), malheureusement non disponibles, aurait permis de valider ou non, cette hypothèse.

En termes de poursuites, la connaissance génomique de ces sérovars a progressé. L'existence d'un clone unique du TMV français interroge sur son origine (élevage ou abattoir). La persistance constatée de certains clusters de souches des 3 sérovars dans leur filière respective pose la question des propriétés d'adaptation et/ou de résistance de ces sérovars. Les premières analyses génomiques faites sur *S. Mbandaka* et sur *S. Typhimurium* ou son TMV indiquent la présence de facteurs génétiques de résistance : pour ces trois sérovars, de nombreux gènes de résistances à des antibiotiques, aux métaux lourds et à certains biocides ont été identifiés. Il serait intéressant de vérifier si ces gènes sont caractéristiques de ce sérovar ou de ces filières en comparaison d'autres sérovars ou d'autres filières. Si tel est le cas, des stratégies de criblage pour déterminer les souches les plus résistantes pourraient être développées pour permettre une meilleure surveillance et prévention des contaminations au niveau des élevages et améliorer la gestion du risque.

Références bibliographiques

Bone A., Noel H., Le Hello S., Pihier N., Danan C., Raguenaud ME., Salah S., Bellali H., Vaillant V., Weill F.X., Jourdan-da Silva N., 2010. Nationwide outbreak of *Salmonella enterica* serotype 4,12:i:- infections in France, linked to dried pork sausage, March-May 2010. *Euro Surveill* 15(24):pii=19592.

EFSA, 2018. Outcome of EC/EFSA questionnaire (2016) on use of Whole Genome Sequencing (WGS) for food- and waterborne pathogens isolated from animals, food, feed and related environmental samples in EU/EFTA countries. *EFSA Journal*. 49 pages ; doi:10.2903/sp.efsa.2018.EN-1432.

Fournet N., Laurent E., Jones G., Tourdjman M., Chereau F., Nisavanh A., Jourdan Da Silva N., de Valk H., 2021. Surveillance des toxi-infections alimentaires collectives. Données de la déclaration obligatoire, 2019. Mars 2021. *Santé publique France*. 2021. 12 pages.

Fu S., Hiley L., Octavia S., Tanaka M., Sintchenko V., Lan R., 2017. Comparative genomics of Australian and international isolates of *Salmonella Typhimurium*: correlation of core genome evolution with CRISPR and prophage profiles, *Scientific Reports* 7, 9733.

Gossner C.M, Van Cauteren D., Le Hello S., Weill F-X., Terrien E., Tessier S., Janin C., Brisabois A., Dusch V., Vaillant V., Jourdan-da Silva N., 2012. Nationwide outbreak of *Salmonella enterica* serotype 4,[5],12 :i :- infection associated with consumption of dried pork sausage, France, November to December 2011. *Euro Surveill* : 17(5):pii=20071.

Jolley, K. A., Maiden, M. C., 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11, 595.

Jourdan-da Siva N., Le Hello S., 2012. Salmonelloses en France, 2002-2010: tendances en épidémiologie humaine, émergence de la souche monophasique, principaux aliments impliqués dans les dernières épidémies. *Bulletin épidémiologique, santé animale et alimentation* no 50/Spécial Risques alimentaires microbiologiques ; 31-35 Eds : Anses- InVS – Paris.

Lee R., Radomski N., Proulx J.F., Manry J., McIntosh F., Desjardins F., Soualhine H., Domenech P., Reed M.B., Menzies D., 2015. Reemergence and Amplification of Tuberculosis in the Canadian Arctic. *The Journal of Infectious Diseases* 211(12), 1905–1914.

Letunic I., Bork P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* 44(W1), 242-5.

Nguyen L.T., Schmidt H.A., von Haeseler A., Minh B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular and Biology Evolution* 32(1), 268-74.

Plateforme de Surveillance de la Chaine Alimentaire, 2021. Etude d'épidémiologie-surveillance génomique rétrospective de Salmonella Dublin en Franche-Comté, publié le 08/09/2021, 27 pages, disponible at Rapport Technique de l'étude génomique rétrospective S. Dublin en Franche-Comté | Plateforme SCA (plateforme-sca.fr).

Radomski N., Cadel-Six S., Cherchame E., Felten A., Barbet P., Palma F., Mallet L., Le Hello S., Weill F.X., Guillier L., Mistou M.Y., 2019. A Simple and Robust Statistical Method to Define Genetic Relatedness of Samples Related to Outbreaks at the Genomic Scale - Application to Retrospective Salmonella Foodborne Outbreak Investigations. *Frontiers Microbiology* 10, 2413.

Rumore J., Tschetter L., Kearney A., Kandar R., McCormick R., Walker M., Peterson C., Reimer A., Nadon C., 2018. Evaluation of whole-genome sequencing for outbreak detection of Verotoxigenic Escherichia coli O157:H7 from the Canadian perspective. *BMC Genomics* 19.1, p. 870.

Tonkin-Hill G., MacAlasdair N., Ruis C., Weimann A., Horesh G., Lees J. A., Gladstone R. A., Lo S., Beaudoin C., Floto R. A., Frost S. D. W., Corander J., Bentley S. D., Parkhill J., 2020. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biology* 21(1), 180.

Ung A., Baidjoe A., Van Cauteren D., Fawal N., Fabre L., Guerrisi C., Danis K., Morand A., Donguy M.P., Lucas E., Rossignol L., Lefèvre S., Vignaud M.L., Cadel-Six S., Lailier R., Jourdan-da Silva N., Le Hello S., 2019. Disentangling a complex nationwide Salmonella Dublin outbreak associated with raw-milk cheese consumption, France, 2015 to 2016. *Euro Surveillance* 24(3):pii=1700703.

UniProt C., 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1), D480-D489.

Van Cauteren D., Le Strat Y., Sommen C., Bruyand M., Tourdjman M., Da Silva N., Desenclos J., 2017. Estimated Annual Numbers of Foodborne Pathogen-Associated Illnesses, Hospitalizations, and Deaths, France, 2008–2013. *Emerging Infectious Diseases* 23(9), 1486-1492.

Vila Nova M., Durimel, K., La K., Felten A., Bessieres P., Mistou M. Y., Mariadassou M., Radomski N., 2019. Genetic and metabolic signatures of Salmonella enterica subsp. enterica associated with animal sources at the pangenomic scale. *BMC Genomics* 20(1), 814.

World Health Organization, 2015. WHO estimates of the global burden of foodborne diseases: foodborne disease burden epidemiology reference group 2007-2015. *World Health*, 265 pages.



Cet article est publié sous la licence Creative Commons (CC BY-NC-ND 4.0)

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Pour la citation et la reproduction de cet article, mentionner obligatoirement le titre de l'article, le nom de tous les auteurs, la mention de sa publication dans la revue *Innovations Agronomiques* et son DOI, la date de publication.