Epidemiology of Salmonella in the livestock sector using a genomic approach

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. Epidemiology of Salmonella in the livestock sector using a genomic approach. Innovations Agronomiques, 2024, 88, pp.137-149. 10.17180/ciag-2024-vol88-art12-GB . hal-04715279

HAL Id: hal-04715279
https://hal.inrae.fr/hal-04715279v1

Submitted on 30 Sep 2024

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

Epidemiology of *Salmonella* in the livestock sector using a genomic approach

Valérie MICHEL[1], Madeleine DE SOUSA VIOLANTE[1], Carole FEURER[2], Michel-Yves MISTOU [3], Ludovic MALLET [4]

[1]Dairy Microbiology Department, ACTALIA, 74800 La Roche sur Foron, France

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

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

[4]Toulouse University Cancer Institute - Oncopole, 31059 Toulouse, France

Correspondence : v.michel@actalia.eu

Abstract :

During the EMISSAGE project, the 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 the 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 is one of the 4 main pathogens responsible for food poisoning worldwide (WHO, 2015). In France, the number of people suffering from salmonellosis was estimated at 200,000 per year in 2016 and although most cases are mild, there were 4,000 hospitalisations and 72 deaths in 2016 (Van Cauteren *et al* , 2017). In 2019, cold meats were suspected of being the cause of 14% of cases of suspected or confirmed *Salmonella* collective foodborne outbreaks, while cheese and dairy products were suspected of being the cause of 5% of cases (Fournet *et al* , 2021). Surveillance for *Salmonella spp*. has been stepped up in the main food sectors, particularly in the pork and milk and dairy sectors, which have been affected in recent years by several health crises linked to these two pathogens (Bone *et al* , 2010 ; Gossner *et al*, 2012 ; Jourdan-DaSilva et Le Hello, 2012 ; Ung *et al* , 2019). In order to strengthen this surveillance and release management resources spanning from the primary production to the processing, operators in the aforementioned sectors, needed to characterise these strains in greater detail. Genetic links between strains and their characterisation is the basis upon which such details can be looked up for. They allow to distinguish any dominant clusters (groups of strains with a high degree of genomic relatedness), to identify dissemination and circulation routes when crossing with the epidemiological data (*Plateforme de Surveillance de la Chaine Alimentaire*, 2021) or to investigate the origin of the strains. The Exploitation of the genomic information on strains is enabled by the global development of high-throughput sequencing techniques. This comprehensive sequencing of strains, called Whole Genome Sequencing (WGS), is commonly used in France and throughout the world for the genetic typing of human strains (EFSA, 2018). Its use in the field of agricultural and agri-food production is still partial and needs to be

developed to enable stakeholders to take advantage of these innovative technologies, with the aim of better characterising strains but also exploiting knowledge of these genomes for targeted purposes (identification of genetic markers of resistance, presence of pathogenicity genes, identification of genetic sequences that can be used as specific markers, etc.). One of the needs of operators in the agri-food sectors is to be able to rapidly characterise the origin (geographical, source) of any newly isolated strains.

More specifically, for the pig industry, the term "origin" refers to the geographical origin of the strain, so that it can be quickly traced back to the pig farm carrying the *S.* Typhimurium contamination or its monophasic variant (TMV). For the dairy sector, the notion of origin refers more to the source of contamination (wildlife, feed, etc.).

It was facing this backdrop that the three partners in the EMISSAGE project (CASDAR-RT no. 1701) pooled their scientific and technical skills to gain a better understanding of 3 *Salmonella* serovars of major importance in the animal production chain. For the pig industry, these were *Salmonella* serovars Typhimurium and its mono-phasic variant (TMV) S. 1, 4, (5), 12 : i : - and the Mbandaka serovar for the dairy industry. The scientific partner in the EMISSAGE project was the Food Safety Laboratory (LSAL) at the Anses in Maisons-Alfort, whose GAMER (Genome Analysis Modelling and Risk) mission was involved in the methodological development required to interpret the genomic data. The two technical partners associated with the EMISSAGE project were Ifip (Institut Technique Agricole and Institut Technique Agro-Industriel de la filière porcine) and ACTALIA (Institut Technique Agro-Industriel des produits laitiers et de la Sécurité des Aliments), with the respective involvement of a molecular microbiology project leader and the head of the dairy microbiology laboratory.

The main objectives of the EMISSAGE project were as follow:

- Build a reference collection of genomes of the 3 *Salmonella* serovars from the field using WGS;
- Develop a set of automated bioinformatics pipelines for processing and analysing genomic data, so that it can be transferred to technical institutes for use by professionals;
- Identify epidemiological and risk-associated genetic markers to enable quick and easy screenings from self-monitoring isolates.

1-Implemented approach

The EMISSAGE project has been organised into a number of actions and tasks, listed in the following paragraphs:

1.1 Build up a collection of Salmonella strains isolated from food production, cutting and processing plants.

In order to recover strains of porcine origin, Ifip carried out 9 strain sampling campaigns in slaughter-houses in the Grand-Ouest region, of which 6 were carried out in 2018 and 3 in 2019. Samples of pig carcasses and tongues and their slaughter environment (bays, splitters, chutes, podiums) were analysed for the presence of *Salmonella*, using the AFNOR BRD 07/11-12/05 validated alternative method, at the Ifip laboratory in Maisons-Alfort.

For each strain from pigs, the breeding department was identified. In order to increase the panel of slaughterhouse strains to be sequenced, Ifip contacted DGAl (Direction Générale de l'Alimentation) to obtain strains isolated from DGAl national surveillance plans between 2017 and 2019 on pig carcasses, as well as their metadata. Ifip centralised the metadata received in digital format, so as to establish a file of metadata common to both the strains isolated from its sampling campaigns and those made available by the DGAl.

In the dairy sector, farms in north-western France producing raw milk were asked to provide strains of *S.* Mbandaka from throughout the dairy production and processing chain. The metadata (ecological data) for these requested strains were -minimal: isolation date, farm code, type of sample, geographical origin (French *département*, or *"département"* thereafter). ACTALIA recovered the strains from the dairy operators, after contracting a mutual transfer agreement (an agreement governing the conditions for making the strains and the resulting data available and using them), and also centralised the associated metadata.

1.2 Obtaining the complete genome of strains

To this end, total DNA from the strains was extracted by Ifip and ACTALIA, each processing their respective strains. Extractions and quality control of the extracted DNA were carried out using the same protocol before being sequenced using Illumina technology to produce paired-end sequencing reads, as described by Radomski *et al* (2019).

After sequencing and quality control of the sequences obtained and verification of the serovars, the genome dataset available for the project was 140 genomes for *S.* Mbandaka and 267 genomes for *S.* Typhimurium and its monophasic variant. The genomes were assembled according to the bioinformatics workflow developed by Anses and called ARTwork (Vila Nova *et al*, 2019).

1.3. Developing a bioinformatics pipeline to analyse the pangenome

DNA sequence analysis for the purposes of typing and defining genetic proximity between strains of the same serovar is most often carried out on the coregenome, which is the part of the genome conserved by all the strains in the panel studied. This analysis, which is already highly discriminating, focuses on part of the available genome (between 70 and 80% of the genome for *S.* Typhimurium (Fu *et al,* 2017). By including the accessory genome (i.e. the part of the genome that some strains may miss), the pangenome is analysed. To do this, a pan-genome analysis pipeline called pgSNP was developed specifically during the EMISSAGE project and validated on datasets from collective food poisoning outbreaks (CFTIs).

1.4. Genomic analysis of strains

Firstly, a phylogenomic analysis of each serovar was carried out in order to :

- Studying the genetic diversity of strains;
- Compare it with strains of the same serovar but from a different context (another *département* for serovars of interest to the pig industry, another origin for the serovar of interest to the dairy industry).

All genomes were characterized *in silico* by MLST (MultiLocus Sequence Type) based on the sequences of the 7 housekeeping genes (*aroC, dnaN, hemD, hisD, purE, sucA,* and *thrA*) described in the PubMLST database (Jolley and Maiden, 2010). The Sequence Type (ST) of each genome was obtained using the Tseeman MLST tool (https://github.com/tseemann/mlst).

The comparison was first carried out using an analysis based on the coregenome. Phylogenomic inferences were made using an evolutionary model calculated by IQ-TREE (Nguyen *et al*, 2015) (maximum likelihood inference) on SNPs. The term SNP stands for Single Nucleotide Polymorphism, which consists of counting the number of different nucleotides between two aligned sequences. The iVARCall2 workflow (Lee *et al*, 2015) was used for this purpose. The phylogenetic trees were then

visualised and annotated using the interactive Tree of Life tool (iTOL, https://itol.embl.de/), according to Letunic and Bork (2016).

1.5. Search for markers specific to these serovars to more rapidly characterise their origin

The existence of genetic markers specific to geographical origin or origin linked to the bovine or poultry host was sought, respectively for the 2 serovars of interest for the pig industry and for the serovar of interest for the dairy industry. This search was carried out *in silico* using a genome association approach by comparing the serovars *S.* Typhimurium and TMV serovars, the 267 genomes recovered by Ifip were compared with 83 genomes from French pig slaughterhouse strains of the same serovars isolated from other *départements*. These strains were supplied and their genomes sequenced by the partner ANSES. For the serovar *S.* Mbandaka, 164 genomes of strains from poultry, isolated in identical or geographically close departments, were made available to the EMISSAGE project by Anses (HQPAP units in Ploufragan and LNR antibiorésistance units in Ploufragan. The recovery of additional strains of the 3 serovars studied was made possible thanks to the mobilisation of partner laboratories in the Anses *Salmonella* network and their willingness to cooperate in the identification and supply of strains and associated minimum metadata.

The search for single or combined markers (genes or variants) was carried out.

For single genes, the GFF files produced by the Prokka software (https://github.com/tseemann/prokka) were used with the Panaroo (Tonkin-Hill *et al*, 2020)a pangenome extraction tool that takes into account errors introduced by genome annotation. The identity threshold within gene groupings was 90%. For single variants, the VCF files from iVARCall2 were merged to identify the presence or absence of variants in each genome. The position of the variants was defined in relation to the *Salmonella* Mbandaka SA20026234 strain and the *S.* Typhiumirum LT2 strain.

To explore possible marker combinations, a Python script called MarkerFindr was developed. It computes the combination of a maximum of 3 genes or variants to elicit the one giving the best discrimination score based on the criterion sought (geographical origin, animal origin). Ability scores are based on specificity and sensitivity values. The tool MarkerFindr is available online: https://github.com/madeleinevlt/MarkerFindr. The genes or variants identified as discriminating were annotated using the Blast and Uniprot tools (Uniprot 2021).

2 Results

2.1 A collection of Salmonella strains of the 3 studied serovars from food production regions

For the pork sector, Ifip sampled total of 871 specimen in slaughterhouses over 9 campaigns and carried out in 6 different slaughterhouses noted from A to F in Table 1. 116 *Salmonella* strains were isolated either from the slaughterhouse environment or directly from the animals.

Table 1: Number of Salmonella strains isolated from campaigns carried out in abattoirs by Ifip in 2018 and 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]	

For each strain derived from pigs, the breeding department was identified.

In order to increase the panel of slaughterhouse strains to be sequenced, Ifip contacted the DGAl to recover the strains isolated from the DGAl's national surveillance plans between 2017 and 2019 on pig carcasses. 111 strains of the serovars of interest (24 of *S.* Typhimurium and 87 of its TMV) were sent to Ifip by the antimicrobial resistance mission of the Food Safety Laboratory (LSAL) of the Anses and some local biological laboratories (LDAs). **267 strains belonging to the two serovars of interest to the pig industry were available for the EMISSAGE project**. Ifip centralised the meta-data received in digital format, in a common metadata file for the strains isolated from its sampling campaigns and made available by the DGAl.

For the serovar *S.* Mbandaka, strains from dairy farms or their environment, as well as strains isolated in production workshops or in finished products (raw milk cheeses) were identified. **The 148 *S.* Mbandaka strains came from 23 dairy farms, supplying 3 different production workshops, and** were isolated between 2016 and 2019. The metadata (ecological data) for these strains was minimal: date of isolation, farm code, type of sample, *département*. In addition to this collection, the metadata for the strains was structured in digital format.

2.2 Development and validation of the pgSNP bioinformatics pipeline for pangenome-based strain discrimination

The aim of the bioinformatics pipeline developed during the project was to take into account both core-genome variants (i.e. homologous regions) and accessory-genome variants (i.e. non-homologous regions) that may explain certain adaptation processes to selective pressures. Its development consisted in reconstructing phylogenomic inferences on the basis of both core-genome and accessory-genome variants, which is usually done independently using different analytical methods. For each serovar, a reference pangenome was constructed from an assembly of the genomes considered. The reference pangenome therefore ultimately represents all the genomic sequences in the samples. **During these developments, it was shown that 50% of the genome of all the strains of *S.* Typhimurium and its TMV are not taken into account during analyses focusing solely on the core genome**. In addition, numerous adaptive elements have been detected in this accessory genome, including certain elements shared by 99% of French strains, such as genomic fragments containing heavy metal resistance genes.

Next, the BWA mapper and the FreeBayes variant caller (integrated into the Snippy workflow) were used respectively to align the global genomic sequencing data with the reference pangenome and to identify genome variants at the pangenome scale. A pseudogenome containing the variants for each sample is then constructed and this alignment is cut into pieces, or chunks thereafter to infer phylogenies for each chunk and ultimately by coalescing chunk phylogenies into a pangenome phylogeny of all samples.

Phylogenomic chunk trees are produced using IQTree -for each FastRFS -chunk. Checks were carried out at each stage of the pipeline, such as the presence or absence of redundant genetic material on the *de novo* reconstructed reference or the comparison of phylogenetic trees obtained according to the method used. ERaBLE, which is a phylogenetic branch length estimator on supertrees (the final tree obtained with the pgSNP pipeline), was tested and implemented in the pipeline. **The resulting bioinformatics pipeline, called pgSNP** (Figure 1), was then **validated on 3 different CFTI datasets.**

The first dataset is a collection of *S.* Typhimurium and its monophasic variant containing 192 strains with 4 epidemiological clusters, published by Radomski *et al* (2019). It was first shown on this dataset that pgSNP adds about 1.6 Mb (about 30 percent) of genetic information compared to core-genome based analysis. Compared with epidemiological clusters, pgSNP was able to identify and group epidemic strains together, with the exception of three strains. Two of these have already been described in the paper, while the third comes from a TMV strain that is linked to an *S.* Typhimurium cluster in the genome-wide tree. This result is supported by the small distance between the strains. It was also shown that pgSNP induces topological differences in monophasic variants, which have few differences at the core-genome level, compared with *S.* Typhimurium whose topology remains preserved between the two methods. This result shows that the addition of the accessory genome mainly impacts strains with little core-genome variability, even though the core-genome methods assess topology on around 80% of the data. Overall, pgSNP obtains results that are consistent with the epidemiological data for this dataset.

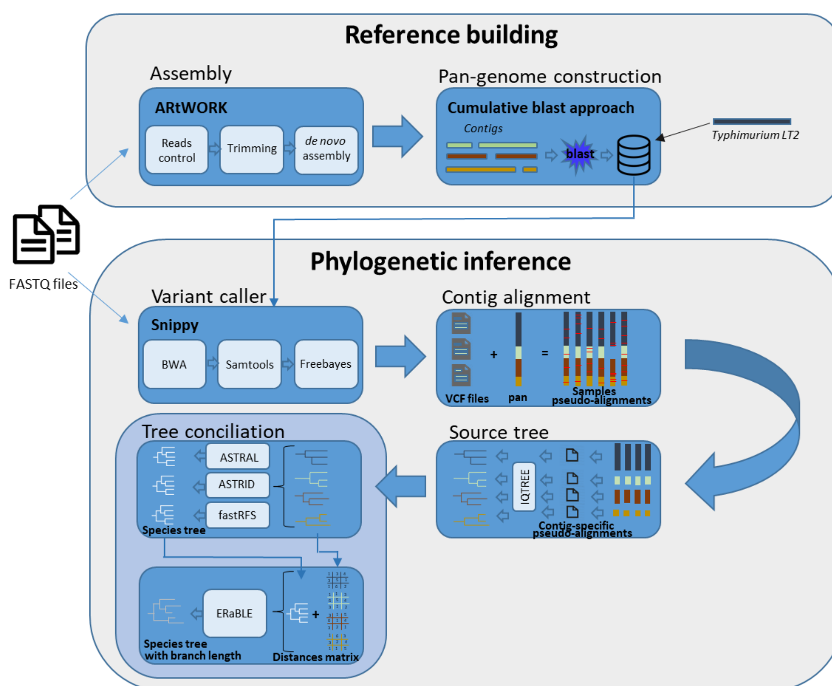


Figure 1. Schematic representation of the genome-wide analysis pipeline, pgSNP

The second dataset corresponds to an *Escherichia coli* O157:H7 dataset published by Rumore *et al* (2018). This dataset contains 210 strains of medical origin from human cases, with 8 epidemiological clusters that have very few differences on the core-genome (< 5 SNPs). Using pgSNP, most of the identified epidemiological clusters were found to be consistent with the results published by the authors. However, using the reference pangenome revealed a topological change from an epidemic strain to sporadic strains. This topological change is due to the presence of 206 kb of core-genome DNA (plasmids, chromosomal DNA) present in the reference pangenome but absent in the publication reference. PgSNP

demonstrates here the advantage of using a reference pangenome in the event of poor reference genome selection.

Finally, pgSNP was tested on a dataset of epidemiological clusters of *Neisseria meningitidis*. Compared with the other two datasets, *Neisseria meningitidis* is highly recombinant, with a small genome size. This dataset also contained precise information on sporadic strains. With pgSNP, new reconciliations were shown for some strains. Overall, pgSNPs add genetic distance between epidemiological strains, notably due to the mobile genetic elements and recombinant nature of *Neisseria meningitidis*. These changes make it possible to link sporadic and epidemic strains that have very similar meta-data, and which could therefore potentially be linked. pgSNP has also shown that the additional genetic distances can make outbreak investigations difficult. However, samples from outbreaks with close core and accessory genomes are always grouped together in the genome-wide tree, demonstrating the importance of the accessory genome in epidemiological investigations.

2.3 Description of the genomic diversity of Salmonella Typhimurium strains and its monophasic variant (TMV)

The genomic diversity of pig strains of these serovars has been described (Figure 2**). The genomic difference between the genomes of the monophasic variants is very small, and it could be hypothesised that a single clone is disseminated in France**. Looking more closely, the average pairwise difference between the 152 TMV samples is 64 SNPs. Topologically, an internal node divided the 152 TMVs into two groups of 104 and 48 TMV samples, with an intra-group mean of 49 and 51 SNPs, respectively. This low TMV diversity may be explained by the recent appearance of this variant, compared with *S.* Typhimurium strains. By examining the genomic content, the high number of genes for resistance to antibiotics, heavy metals and biocides seem to be associated with the prevalence of these two serovars in pig farms and could explain their adaptation.

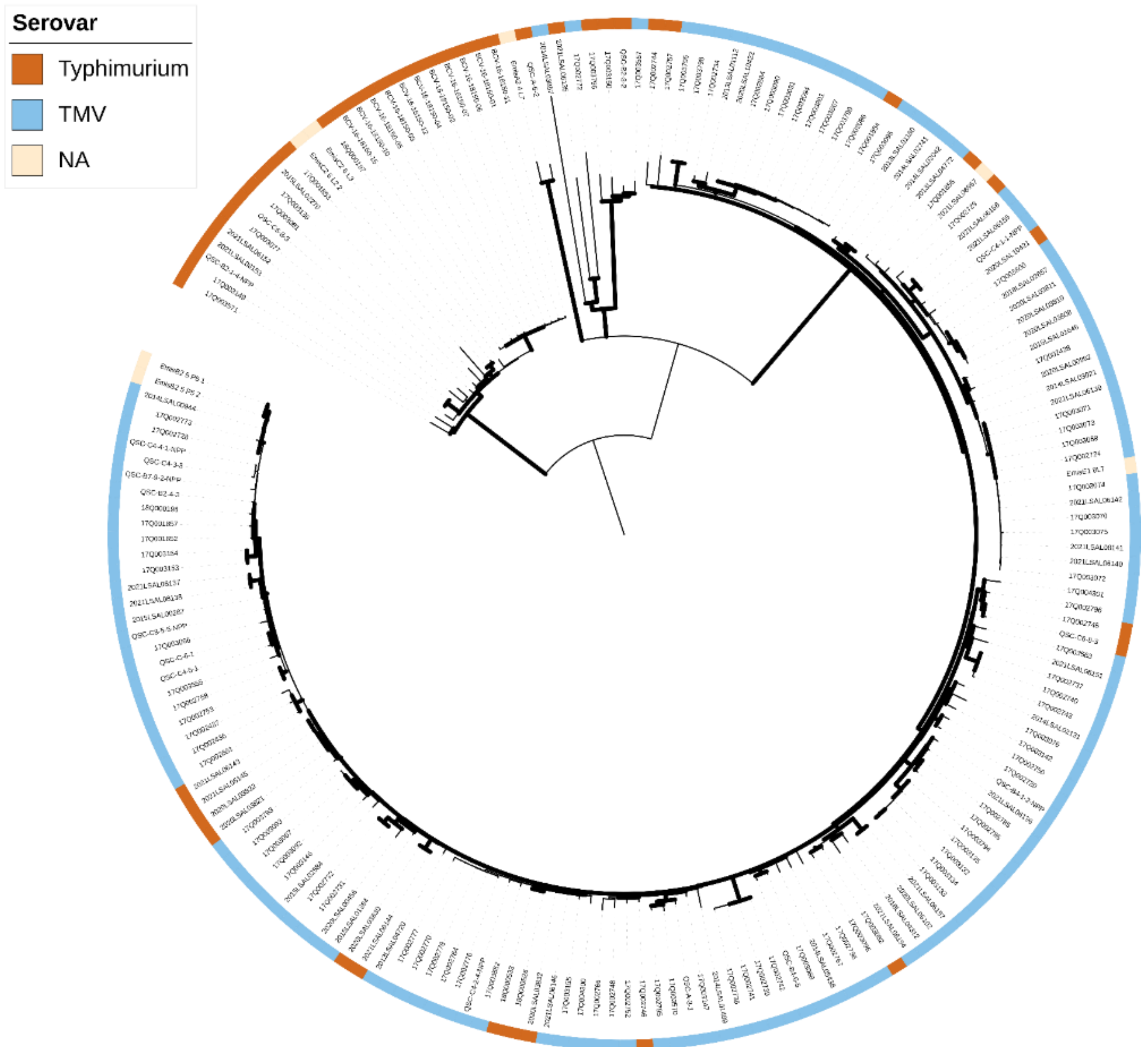


Figure 2. Phylogenomic tree of S. Typhimurium strains (orange) and its TMV (blue)

The industry's main question was whether there was a link between geographical diversity and genomic diversity in *S*. Typhimurium and TMV. To do this, 188 strains from 3 pork-producing regions were selected. Some strains came from pig farms, but most were strains from samples taken at the slaughterhouse, with identification of the animal's department of origin**. This dataset revealed a distribution of these strains throughout France, with no particular adaptation to geography** (Figure 3).

Tree scale: 0.0001

Matrices
Region 1
Region 2
Region 3
Others

Figure 3. Phylogenomic tree of S. Typhimurium strains and its TMV. The outer circle shows the location of the strains according to their region of isolation.

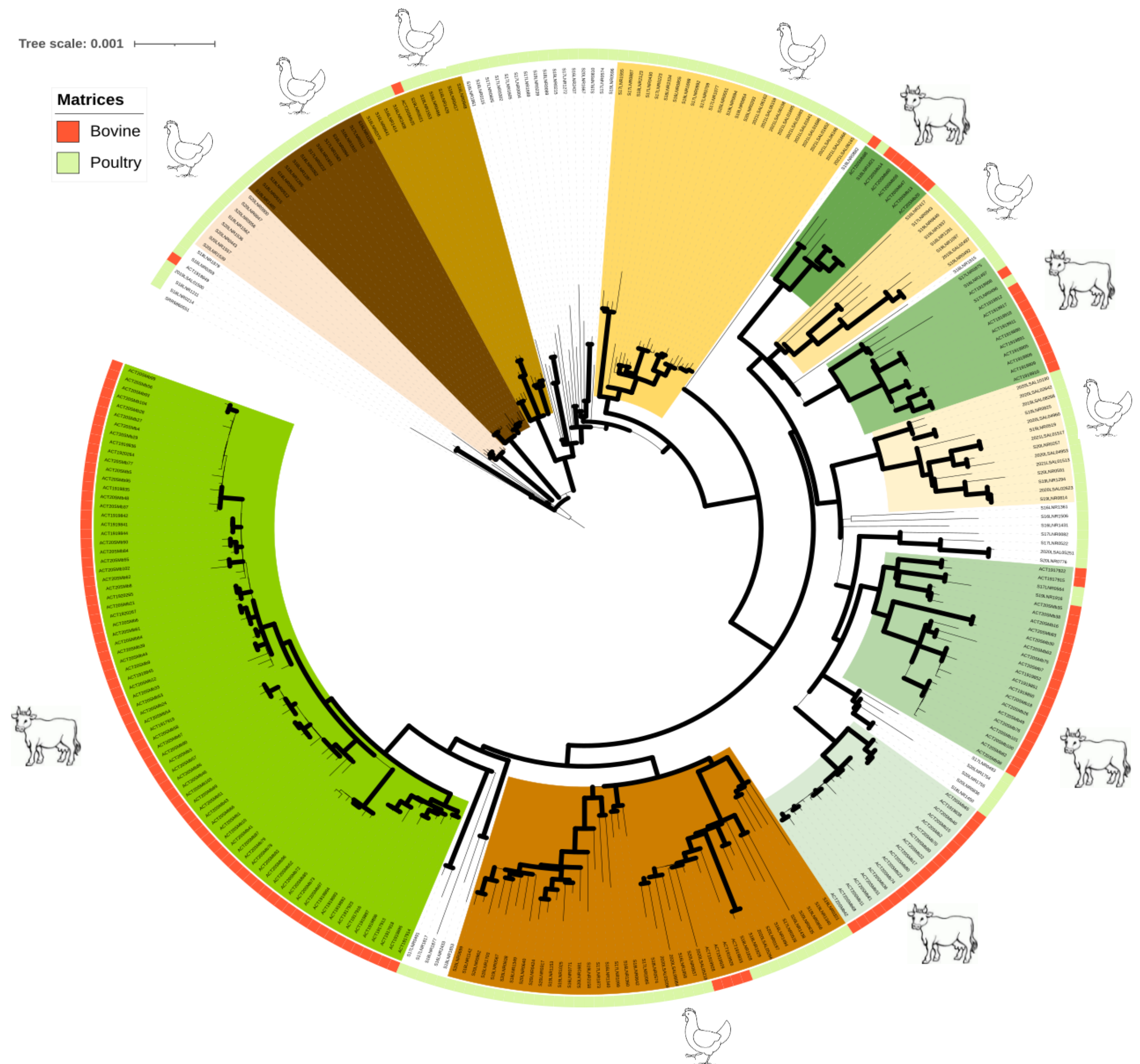
Analysis of variants or genes did not reveal any discrimination between samples from different regions. pgSNP was also applied to this dataset, in order to check whether geographical adaptation was localised in the accessory genome of these serovars. Thanks to the pipeline, the content of the accessory genome was rapidly identified, with the presence of phages and plasmids on certain strains. Analyses were carried out on 30% more bases to reconstruct the phylogenomic tree. These new reconciliations, obtained thanks to the accessory genome, could not be linked to geography.

2.4 Description of the genomic diversity of Salmonella Mbandaka strains

All the *S.* Mbandaka strains isolated in France belong to the same MLST profile: ST413. The phylogenomic tree inferred from the poultry and dairy strains showed that there is no host specificity, but rather avian clades interspersed with bovine clades (Figure 4). Across the **different host clades, there was no matrix specificity, either between poultry samples isolated from laying hens and broilers**

or matrices from the beef dairy sector. Turkey samples are scattered all around the tree, but with a unique branch length, and have therefore been identified as singletons. The meta-geographical data reveal that most of the time, poultry samples are clustered with isolates from the same or neighbouring regions. However, in the upper part of the phylogenomic tree, strains from different regions are grouped together, without any link of geographical proximity between the regions. This low level of diversity between strains from different geographical sources is almost exclusively true for poultry isolates, which allows us to hypothesise this as an outcome of a common source of reproduction. In fact, the parents are on the farms and take care of the births of the hens, which are then sorted into broilers and layers. These hens are then redistributed among the farms, and can spread the disease from the same source, which would explain this clade.

Figure 4. Phylogenomic tree of S. Mbandaka strains. The outer circle shows whether the strains are of dairy cattle (orange) or poultry (green) origin. Host clades are highlighted in colour.



The pgSNP bioinformatics tool was also applied to this dataset, and despite the number of additional bases analysed on the phylogenomic tree (+27%), very few extra differences on the host clades were

found compared to those already identified thanks to the coregenome. Finally, the possibility of shared contamination between the two sectors was investigated, based on the observations of dairy operators in the North-West region. The possibility of airborne transmission was proposed, based on public data on strains from wild American birds. **The proximity of these strains to strains from poultry close to the coast, and also to a strain from manure in the North-West, suggests possible contamination by wildlife**. Additional European data is needed to validate these hypotheses. A study of this pathway, in conjunction with neighbouring countries, could validate this possibility of contamination by wildlife, which has already been observed in *S.* Typhimurium carried by rodents.

2.5 Search for genetic markers associated with the geographical location of S. Typhimurium and its TMV strains

A comparative genomic analysis was carried out *in silico* in an attempt to identify genomic markers or pairs of genomic markers that would be specific either to the origin of the strains for *S.* Mbandaka or to their geographical origin for *S.* Typhimurium. Such markers have been identified, but they are not yet fully operational. In order to be able to deploy PCR-based screening tests, which are relatively simple and reasonably priced, it was decided that the number of discriminating markers should not exceed 3. For example, for the dairy sector, the dataset appeared to be split between 12 different clusters of strains isolated from poultry or cattle (Figure 4). There are markers specific to bovine or poultry strains, but no markers that are 100% discriminatory of the bovine or poultry origin. The combination of 2 or 3 markers never achieves specificity and sensitivity in excess of 95% for the two criteria, leading to predictive accuracy of between 80% and 95%, which is not good enough. The integration of additional markers, belonging to the accessory genome, did not make it possible to obtain discriminating markers with sufficient accuracy. The same approach was used for pig serovars in an attempt to associate genetic markers specific to a region or department. This French diversity was characterised using a combination of genes and variants, which is capable of discriminating between French strains with an accuracy of 86%. The performance criteria were not judged to be sufficiently discriminating to be applied in a routine setup. Further analyses of accessory variants should be conducted to improve accuracy.

In conclusion, we can say that it is **possible to define markers more or less associated with the discrimination criteria desired by the sectors. However, the conditions for their implementation to make them economically affordable and technically relevant (accuracy of prediction) have not been met**. PCR application tests have therefore not been developed.

Conclusion - Project outlook

The CASDAR-RT EMISSAGE project between the academic laboratory of the Anses in Maisons-Alfort, and two technical institutes (Ifip and ACTALIA) has enabled all the partners to gain a better understanding of 3 *Salmonella enterica subsp. enterica* serovars of major importance in the pork and dairy sectors. **The genomic diversity of strains circulating in pig slaughterhouses in France for the serovars Typhimurium and its monophasic variant, as well as that of bovine strains in the North-West region of raw milk cheese production and in the poultry industry, has been characterised for the first time. This work was carried out on a large number of strains, the result of a major effort to isolate them under production field conditions**. The pgSNP bioinformatics pipeline has been developed as a tool for processing genomic data, with the aim of maximising genetic information based on all genomic information (the pangenome) in order to compare strains. The relevance of the pgSNP pipeline has been validated on TIAC datasets and, by maximising the genetic information available, it has proved to be highly discriminating. This approach, which uses the pangenome to differentiate between strains, is currently in full development, and pgSNP is a major progress in this field.

However, **knowledge of genomic diversity and the proximity of strains within the same serovar (for the monophasic Typhimurium variant) or between bovine and avian strains are important factors in better investigating contamination in the field. The identification by pgSNP of a single TMV clone circulating in the pig industry, and specifically in France, is a major contribution to our knowledge of epidemiological investigations**. **The hypothesis of cross-species contamination between poultry and cattle is being considered for certain clusters of *S.* Mbandaka strains.** The provision of ecological data on these strains (exact context of sampling and isolation), unfortunately not available, would have enabled this hypothesis to be tested.

In terms of advances and perspectives, the genomic knowledge of these serovars has greatly progressed. The existence of a single clone of the French TMV raises questions about its origin (breeding or slaughterhouse). The persistence of certain clusters of strains of the 3 serovars in their respective sectors raises the question of their adaptation and/or resistance properties The initial genomic analyses carried out on *S.* Mbandaka and *S.* Typhimurium or its TMV indicate the presence of genetic resistance factors: for these three serovars, numerous genes for resistance to antibiotics, heavy metals and certain biocides have been identified. It would be interesting to check whether these genes are characteristic of this serovar or of these pathways, with regards to other serovars or pathways. Would it be the case, screening strategies to identify the most resistant strains could be developed to enable a better monitoring and prevention of contamination at farm level, curb spreading, improve damage control by avoiding cleaning methods serovars are resistant to and ultimately improve risk management.

Ethics

The authors declare that the experiments were carried out in compliance with the applicable national regulations.

Declaration on the availability of data and models

The data supporting the results presented in this article are available on request from the author of the article.

Declaration on Generative Artificial Intelligence and Artificial Intelligence Assisted Technologies in the Drafting Process.

The authors have used artificial intelligence-assisted technologies to translate from French to English.

Declaration of interest

The authors declare that they do not work for, advise, own shares in, or receive funds from any organisation that could benefit from this article, and declare no affiliation other than those listed at the beginning of the article.

References

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 Surveillance : 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'épidémio-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., Lailler 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.

This article is published under the Creative Commons licence (CC BY-NC-ND 4.0)
https://creativecommons.org/licenses/by-nc-nd/4.0/
When citing or reproducing this article, please include the title of the article, the names of all the authors, mention of its publication in the journal *Innovations Agronomiques* and its DOI, and the date of publication.