

WGS analysis of *Salmonella* Welikade, a rarely described serovar involved in a foodborne outbreak in France in 2016

Cherchame Emeline¹, Lailier Renaud¹, Vignaud Marie-Leone¹, Jourdan-Da Silva Nathalie³, Le Hello Simon^{2#}, Guillier Laurent¹ and Cadet-Six Sabrina¹

¹: Université PARIS-EST; Anses, Maisons-Alfort Laboratory for Food Safety, France
²: National Reference Center of Salmonella, Pasteur Institute, Paris, France.
³: Santé publique France, Saint-Maurice, France
Current affiliation : CHU Hôpital Côte de Nacre (Caen), France.

#FoodSafetyWGS2019

INTRODUCTION

Salmonella enterica subsp. *enterica* serovar Welikade doesn't belong to the most frequent isolated serovars and is underrepresented in public databases. Besides, no complete genome is available on public database for this serovar. In France, since 1980, data from the French *Salmonella* Network (SN) shows that *S. Welikade* is principally isolated from laying hens and breeding flocks of *G. gallus* and turkey (n= 82/132 801 isolates). From 2012 to 2015, according to data of the National Reference Center for *Salmonella* (CNR), this serovar was responsible of only 3 cases. Astonishingly in 2016, 8 human cases, whose 5 reported in South of France (Midi-Pyrénées), were recorded by the French national public health agency. The aim of this study was to explore the diversity of this serovar, determining the genomic distance of the clones circulating within human and non-human isolates and, through WGS approach, to confirm the food source of the outbreak occurred in South of France in 2016.

MATERIAL AND METHODS

- ❖ DNA was prepared with the Wizard® Genomic DNA Purification Kit (Promega, France).
- ❖ Genome were sequenced with the Illumina paired-end technology.
- ❖ Pan genome analysis by using ROARY [1] : (investigation of the genomic distance of *S. Welikade* toward *Salmonella enterica*)
 - 129 genomes belonging to 42 serovars (including the European top 25) were analyzed.
 - Phylogenetic tree (figure 1) was established on the gene presence/absence.
- ❖ Core genome SNP analysis (figure 2) on strains collection (table 1):
 - 29 strains of *S. Welikade* collected by the French SN
 - 12 human strains from the CNR for *Salmonella*.
 - Genomes were analyzed by iVARCall2 [2] to generate the SNP matrix distance.
 - Phylogenetic inference based on the maximum-likelihood method implemented in RAxML [3].
 - MLST [7] profiles were obtained by CGE service [8].
- ❖ Genome accessory analysis: research of phages by using PHASTER [9].
- ❖ Annotation of trees: R [4] with the package ape [5] and ggtree [6].

Welikade strains collection	1905	1993	2012	2014	2015	2016	Total général
Animal		1		2	15	5	23
Gallus gallus				2	15	5	22
Animal - Gallus				2	15	5	22
Gull		1					1
Animal - Gull		1					1
Feed				2	1		3
By-products of sunflower				1			1
Feed				1			1
Final product for Gallus					1		1
Feed					1		1
Food consisted for poultry				1			1
Feed				1			1
Food				1	2		3
Cake				1			1
Food - Cake				1			1
Goat cheese					2		2
Food - Goat Cheese					2		2
Human	1		1		2	8	12
Blood						1	1
Human						1	1
Faeces	1		1		2	6	10
Human	1		1		2	6	10
Urine						1	1
Human						1	1
Total général	1	1	1	2	20	16	41

Table 1 : Panel of *Salmonella enterica* subsp. *enterica* Welikade analyzed in this study

Genomic distance of the serovar Welikade within the *Salmonella enterica*

- ❖ Pangenome analysis (figure 1):
 - *Salmonella enterica* subsp. *enterica* Bredeney CFSAN001080 (NCBI NZ_CP007533.1): best reference for SNP analysis (complete genome available).
 - Gaminara: the closest serovar. No complete genome available on public database.

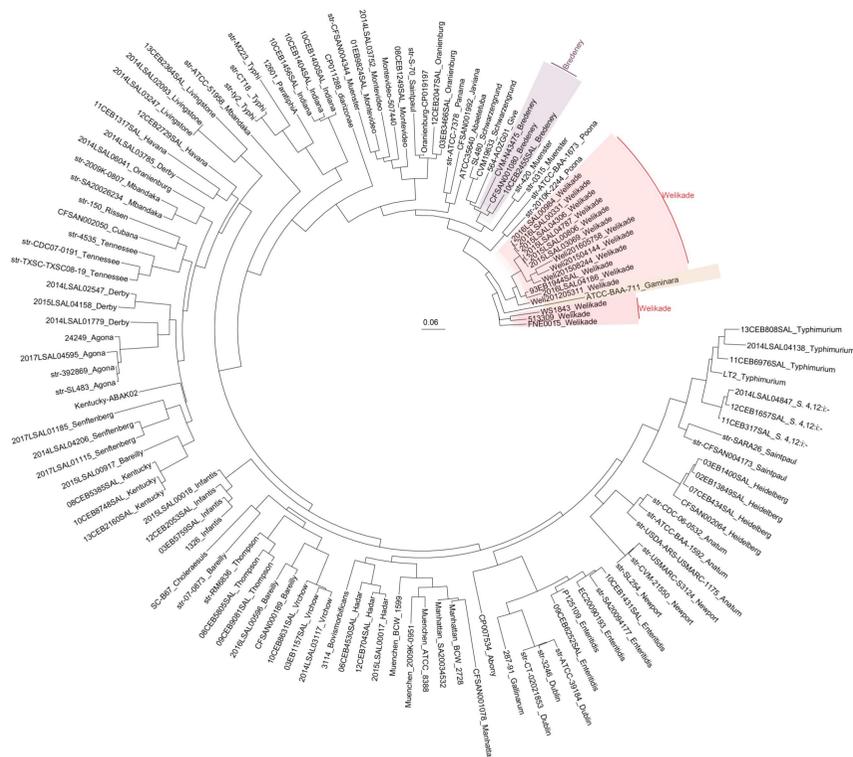
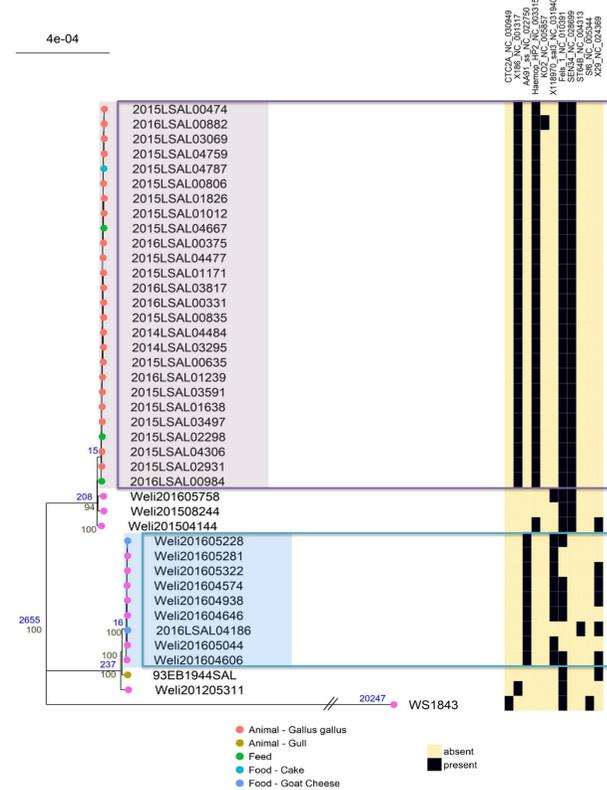


Figure 1 : Phylogenetic tree based on the binary presence and absence of accessory genes. The tree is rooted on the branch of the *Salmonella enterica* subsp. *enterica* Welikade strains.

Single Nucleotide Polymorphism (SNP) analysis

- ❖ SNP analysis using the *S. Bredeney* CFSAN001080 as reference (figure 2):
 - Breath coverage of the reference: 96% on average.
 - Historical strain WS1843, isolated in 1956 (MLST profile 2831): genetically distant of $20\,247 \pm 222$ SNPs to the 2 other groups (MLST profile 3300): genetically distant of $2\,655 \pm 12$ SNPs.



Whole strains from poultry sector (feed, animal and food) isolated between 2014 and 2016 group together : 15 ± 4 SNPs.

7 strains implied in infection occurred in 2016 group together with 2 strains isolated in goat cheese : 16 ± 4 SNPs.

Figure 2 : Phylogenetic core genome SNP tree constructed under the maximum likelihood criterion using the GTR-gamma model. The phylogenetic inference converged after 750 bootstraps replicates. The tree is rooted on the branch of the historical strain WS1843 with a trunk for readability. Bootstraps higher than 80% are noted in brown. Average SNP distance are noted in blue for each clade. On the right of the tree, a heatmap presenting the presence (in black) or absence (in beige) of phages.

CONCLUSION AND PERSPECTIVE

- SNP and phylogenetic analyses realized on the 41 *S. Welikade* genomes suggest that this serovar is subdivided by at least three very distant clades.
- The SNP analysis realized in this study has confirmed the source of contamination of human infection that occurred in 2016: goat cheese.
- The SNP analysis also demonstrates a food and poultry sector contamination by feed sources: products for poultry and *Gallus gallus*.
- This study improved the knowledge on this serovar, rarely described.

REFERENCES

- Page, A.J., et al., *Roary: rapid large-scale prokaryote pan genome analysis*. Bioinformatics, 2015. 31(22): p. 3691-3.
- Felten, A., et al., *First gene-ontology enrichment analysis based on bacterial coregenome variants: insights into adaptations of Salmonella serovars to mammalian- and avian-hosts*. BMC Microbiol, 2017. 17(1): p. 222.
- Stamatakis, A., *RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies*. Bioinformatics, 2014. 30(9): p. 1312-3.
- Gentleman, R.C., et al., *Bioconductor: open software development for computational biology and bioinformatics*. Genome Biology, 2004. Volume 5 (Issue 10): p. Article R80.
- Paradis, E., J. Claude, and K. Strimmer, *APE: Analyses of Phylogenetics and Evolution in R language*. Bioinformatics, 2004. 20(2): p. 289-290.
- Yu, G., et al., *ggtree: an R package for visualization of phylogenetic trees with their covariates and other associated data*. Methods in Ecology and Evolution, 2017. 8(1): p. 28-36.
- Achtman, M., et al., *Multilocus sequence typing as a replacement for serotyping in Salmonella enterica*. PLoS Pathog, 2012. 8(6): p. e1002776.
- <https://cge.cbs.dtu.dk/services/MLST/>
- Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S., *PHASTER: a better, faster version of the PHAST phage search tool*. Nucleic Acids Res, 2016. 44(W1): p. W16-21.

ACKNOWLEDGEMENTS

We gratefully acknowledge the Cire Midi-Pyrénées team for conducting field epidemiological investigation and the French *Salmonella* Network team for giving the access to their collection.