



Project Number: 1448

Project Acronym: ArtiSaneFood

Project title: Innovative Biointerventions and Risk Modelling Approaches for Ensuring Microbial Safety and Quality of Mediterranean Artisanal Fermented Foods



Deliverable 2.2 – Part B: Genome Analysis of pathogens (**R**eport, **P**ublic; Month 27)

Introduction

The objective of Task 2.3 was to submit the isolates collected in the artisanal foods investigated in the project to whole genome sequencing (WGS) to map each isolate to a strain level; and investigate their pathogenic, virulence and antibiotic resistance properties. All sequenced genomes will be made publicly available in databases, as ENA-EBI, NCBI, Enterobase, to be used to improve global surveillance of foodborne pathogens.

Summary of Materials and Methods

Isolate Collection

Bacterial pathogens were collected during the activities performed in Task 2.2 and reported in Deliverable 2.2 Part A.

In total 147 bacterial pathogens were isolated and confirmed by either biotyping, PCR or MALDI/TOF. Isolates were collected from PGI ‘Alheira de Mirandela’ ‘Alheira de Vinhais’ sausage (A), PDO ‘Queijo de Cabra Transmontano’ cheese (B), Spanish Iberian raw-cured sausage “salchichón” (C), PDO ‘Queso Manchego’ cheese (D), PDO ‘Squacquerone di Romagna’ cheese (E) and Emilia-Romagna salami (F).

For each isolate, minimum metadata were collected on isolation matrix, isolation date, latitude and longitude. A specific code was assigned to each isolate: “ArF” (indicating ARTISANEFood) followed by the letter related to the specific food product (A-F) followed by the pathogen (LM for *Listeria monocytogenes*, SE for *Salmonella enterica*, SA for *Staphylococcus aureus*, KO for *Klebsiella oxytoca* and KP for *Klebsiella pneumoniae*), followed by the progressive number of the isolate. As an example, isolate ArFALM01 corresponds to isolate 1 of *Listeria monocytogenes* collected from PGI ‘Alheira de Mirandela’ ‘Alheira de Vinhais’ sausage. Among these bacterial species, *Klebsiella spp.* was out of Artisanefood targets. By chance, on McConkey agar, several *Klebsiella spp.* isolates from Italian salami and soft cheese were collected. *Klebsiella spp.* is an ubiquitous bacterium which occasionally can be pathogenic for humans. In the literature several publications reported multiresistant *Klebsiella pneumoniae* and *K. oxytoca* strains associated to nosocomial infections. Partner UNIBO decided to sequence those isolates in order to get more insight on the potential pathogenicity and antimicrobial resistance of isolated foodborne *Klebsiella*. In particular foodborne *Klebsiella* were compared to public genomes of *Klebsiella spp.* isolated from nosocomial infections.

Table 1: Bacterial pathogens collection and associated metadata

Sample code	Sample	Isolation matrix	Isolation date	Latitude	Longitude
ArFASE01	A-MP4	Alheira Valpaços	26/05/2020	41,6071672	-7,311620098
ArFASE02	S1-1A	Alheira Braganza	28/01/2021	41,82749389	-6,620919116
ArFASE03	S1-1B	Alheira Braganza fair	10/11/2019	41,80750247	-6,763285234
ArFASE04	S1-2A	Alheira Braganza	28/01/2021	41,82749389	-6,620919116
ArFASE05	S2-1A	Alheira Braganza fair	10/11/2019	41,80750247	-6,763285234
ArFASE06	S2-2C	Alheira Braganza fair	10/11/2019	41,80750247	-6,763285234
ArFASE07	S2-3B	Alheira Braganza fair	10/11/2019	41,80750247	-6,763285234
ArFASE08	S25	Alheira Braganza fair	01/02/2020	41,80750247	-6,763285234
ArFASE09	S26	Alheira Braganza fair	01/02/2020	41,80750247	-6,763285234
ArFASE10	S4	Alheira Braganza fair	08/12/2019	41,80750247	-6,763285234
ArFASE11	S4-3C	Alheira Braganza fair	08/12/2019	41,80750247	-6,763285234
ArFASA01	AB3-13	Alheira Braganza	14/02/2020	41,82749389	-6,620919116
ArFASA02	AB3-14	Alheira Braganza	14/02/2020	41,82749389	-6,620919116

ArFASA03	AV2-1	Alheira Vinhais	21/01/2020	41,91848418	-7,08464047
ArFASA04	AV3-1	Alheira Vinhais	27/02/2020	41,91848418	-7,08464047
ArFASA05	AB3-C	Meat Braganza	14/02/2020	41,82749389	-6,620919116
ArFASA06	AV2-4	Alheira Vinhais	21/01/2020	41,91848418	-7,08464047
ArFASA07	AB3-5	Alheira Braganza	14/02/2020	41,82749389	-6,620919116
ArFASA08	AV3-2	Alheira Vinhais	27/02/2020	41,91848418	-7,08464047
ArFASA09	AV2-C	Meat Vinhais	21/01/2020	41,91848418	-7,08464047
ArFBSA01	Q2-20	Cheese	23/11/2020	41,50079164	-7,166431087
ArFBSA02	Q3-16	Cheese	30/11/2020	41,50079164	-7,166431087
ArFBSA03	Q3-17	Cheese	30/11/2020	41,50079164	-7,166431087
ArFBSA04	Q2-19	Cheese	23/11/2020	41,50079164	-7,166431087
ArFCLM01	LM1	raw-cured sausage "salchichón"	20/07/2020	371.643.700.291.461	4.124.746.311.594.110
ArFCLM02	LM2	raw-cured sausage "salchichón"	20/07/2020	371.643.700.291.461	4.124.746.311.594.110
ArFCLM03	LM3	meat	15/09/2020	37.186.668.021.662.50 0	403.635.661.159.368
ArFCLM04	LM5	Surface	15/09/2020	37.186.668.021.662.50 0	403.635.661.159.368
ArFCLM05	LM6	Surface	30/09/2020	37.186.668.021.662.50 0	403.635.661.159.368
ArFCLM06	LM7	Meat	20/10/2020	37.186.668.021.662.50 0	403.635.661.159.368
ArFCLM07	LM8	raw-cured sausage "salchichón"	20/10/2020	37.186.668.021.662.50 0	403.635.661.159.368
ArFCLM08	LM9	raw-cured sausage "salchichón"	20/10/2020	37.186.668.021.662.50 0	403.635.661.159.368
ArFCLM09	LM13	Surface	20/10/2020	37.186.668.021.662.50 0	403.635.661.159.368
ArFCLM10	LM14	Surface	20/10/2020	37.186.668.021.662.50 0	403.635.661.159.368
ArFCLM11	LM119	raw-cured sausage "salchichón"	30/09/2020	37.186.668.021.662.50 0	403.635.661.159.368
ArFCSA01	SA18	Meat	29/06/2020	371.643.700.291.461	4.124.746.311.594.110
ArFCSA02	SA19	Meat	29/06/2020	371.643.700.291.461	4.124.746.311.594.110
ArFCSA03	SA20	Meat	29/06/2020	371.643.700.291.461	4.124.746.311.594.110
ArFCSA04	SA25	raw-cured sausage "salchichón"	15/09/2020	37.186.668.021.662.50 0	403.635.661.159.368
ArFCSA05	SA33	raw-cured sausage "salchichón"	20/10/2020	37.186.668.021.662.50 0	403.635.661.159.368
ArFCSA06	SA34	raw-cured sausage "salchichón"	20/10/2020	37.186.668.021.662.50 0	403.635.661.159.368
ArFCSA07	SA48	Meat	31/08/2020	371.643.700.291.461	4.124.746.311.594.110
ArFCSA08	SA49	Meat	31/08/2020	371.643.700.291.461	4.124.746.311.594.110
ArFDSA01	SA101	Cheese	22/09/2020	3.691.626.154.673.720	34.967.237.140.598.500
ArFDSA02	SA102	Milk	14/10/2020	3.691.626.154.673.720	34.967.237.140.598.500
ArFDSA03	SA103	Cheese	22/09/2020	3.691.626.154.673.720	34.967.237.140.598.500
ArFDSA04	SA105	Milk	18/02/2020	3.690.827.643.389.530	4.369.039.521.390.980
ArFESE01	L6 CM2	maturation room - cheese	04/03/2021	1224515	4417392
ArFESA01	L2 CP1582	cheese - 15 days - 8 °C	05/06/2020	1224515	4417392
ArFESA02	L5 CP421	cheese - 4 days - 2 °C	01/02/2021	1224515	4417392

ArFESA03	L6 CP11285	cheese - 11 days - 2/8 °C	15/03/2021	1224515	4417392
ArFESA04	L5 CP122	cheese - 1 day - 2 °C	29/01/2021	1224515	4417392
ArFESA05	L5 CP822	cheese - 8 days - 2 °C	05/02/2021	1224515	4417392
ArFESA06	L5 CP1122	cheese - 11 days - 2 °C	08/02/2021	1224515	4417392
ArFEKO01	L6 SWW4	warm room - water drainage swab	01/03/2021	1224515	4417392
ArFEKO02	L6 CW2	warm room - cheese	01/03/2021	1224515	4417392
ArFEKO03	L6 CW3	warm room - cheese	01/03/2021	1224515	4417392
ArFEKO04	L6 CP123	cheese - 1 day - 2°C	05/03/2021	1224515	4417392
ArFEKO05	L6 CP124	cheese - 1 day - 2°C	05/03/2021	1224515	4417392
ArFEKO06	L6 CP4281	cheese - 4 days - 2/8 °C	08/03/2021	1224515	4417392
ArFEKO07	L6 CP825	cheese - 8 days - 2 °C	12/03/2021	1224515	4417392
ArFEKO08	L6 CP821	cheese - 8 days - 2 °C	12/03/2021	1224515	4417392
ArFEKO09	L6 CP422A	cheese - 4 days - 2 °C	08/03/2021	1224515	4417392
ArFEKO10	L6 CP823	cheese - 8 days - 2 °C	12/03/2021	1224515	4417392
ArFEKO11	L6 CP822	cheese - 8 days - 2 °C	12/03/2021	1224515	4417392
ArFEKO12	L6 CP883	cheese - 8 days - 8 °C	12/03/2021	1224515	4417392
ArFEKO13	L6 CP4285	cheese - 4 days - 2/8 °C	08/03/2021	1224515	4417392
ArFEKO14	L6 CP1525	cheese - 15 days - 2 °C	19/03/2021	1224515	4417392
ArFEKO15	L6 CP1184	cheese - 11 days - 8 °C	15/03/2021	1224515	4417392
ArFEKO16	L6 CP1523	cheese - 15 days - 2 °C	19/03/2021	1224515	4417392
ArFEKO17	L5 CP183B	cheese - 1 day - 8 °C	29/01/2021	1224515	4417392
ArFEKO18	L5 CP421	cheese - 4 days - 2 °C	01/02/2021	1224515	4417392
ArFEKO19	L5 CP4281	cheese - 4 days - 2/8 °C	01/02/2021	1224515	4417392
ArFEKO20	L5 CP1123	cheese - 11 days - 2 °C	08/02/2021	1224515	4417392
ArFEKO21	L2 CP1181	cheese - 11 days - 8 °C	08/02/2021	1224515	4417392
ArFEKO22	L5 CP1182	cheese - 11 days - 8 °C	08/02/2021	1224515	4417392
ArFEKO23	L2 CP2	cheese at packaging	21/05/2020	1224515	4417392
ArFEKO24	L2 CP482	cheese - 4 days - 8 °C	25/05/2020	1224515	4417392
ArFEKO25	L2 CP483	cheese - 4 days - 8 °C	25/05/2020	1224515	4417392
ArFEKO26	L2 CP821	cheese - 8 days - 2 °C	29/05/2020	1224515	4417392
ArFEKO27	L2 CP11282	cheese - 11 days - 2/8 °C	01/06/2020	1224515	4417392
ArFEKO28	L2 CP15282	cheese - 15 days - 2/8 °C	05/06/2020	1224515	4417392
ArFEKO29	L3 SEM3	maturation room - wall swab	16/07/2020	1224515	4417392
ArFEKO30	L3 CP2	cheese at packaging	16/07/2020	1224515	4417392
ArFEKO31	L4 SWM5	maturation room - water drainage swab	05/11/2020	1224515	4417392
ArFEKO32	L4 CM2	cheese in maturation room	05/11/2020	1224515	4417392

ArFEKO33	L4 SWP5	packaging area - water drainage swab	05/11/2020	1224515	4417392
ArFEKO34	L4 CP2	cheese at packaging	05/11/2020	1224515	4417392
ArFEKO35	L4 CW4	cheese in warm room	02/11/2020	1224515	4417392
ArFEKO36	L5 CP1	cheese at packaging	28/01/2021	1224515	4417392
ArFEKO37	L5 SWM1	maturation room - water drainage swab	28/01/2021	1224515	4417392
ArFEKO38	L5 CP125	cheese - 1 day - 2 °C	29/01/2021	1224515	4417392
ArFEKO39	L5	cheese - 1 day - 8°C	29/01/2021	1224515	4417392
ArFEKO40	CP183B L5 CM1	cheese in maturation room	28/01/2020	1224515	4417392
ArFEKP01	L5 CP1525	cheese - 15 days - 2°C	12/02/2021	1224515	4417392
ArFEKP02	L6 CP485A	cheese - 4 days - 8°C	08/03/2021	1224515	4417392
ArFEKP03	L3 CP425	cheese - 4 days - 2°C	20/07/2020	1224515	4417392
ArFFLM01	2SWD2A	drying room - water drainage swab	29/09/2020	1073497	4463913
ArFFLM02	2SWD2B	drying room - water drainage swab	29/09/2020	1073497	4463913
ArFFLM03	2SWD5A	drying room - water drainage swab	29/09/2020	1073497	4463913
ArFFLM04	2SWD5B	drying room - water drainage swab	29/09/2020	1073497	4463913
ArFFSA01	5SBD3	drying room - salami	10/11/2020	1073497	4463913
ArFFSA02	2SBD3	drying room - salami	29/09/2020	1073497	4463913
ArFFSA03	6SBR4	maturation room - salami 3 weeks of ripening	09/12/2020	1073497	4463913
ArFFSA04	3STM1	stuffing room - surface swab	07/10/2020	1073497	4463913
ArFFSA05	4MB1	stuffing room - meat mixture	21/10/2020	1073497	4463913
ArFFSA06	2MB3	stuffing room - meat mixture	23/09/2020	1073497	4463913
ArFFKO01	5STM4	stuffing room - surface swab	04/11/2020	1224515	4417392
ArFFKO02	6SBR105	storage room - salami 10 weeks of ripening	26/01/2021	1224515	4417392
ArFFKO03	SBD2	drying room - salami	07/07/2020	1224515	4417392
ArFFKO04	2SBD1	drying room - salami	29/09/2020	1224515	4417392
ArFFKO05	2SBD2	drying room - salami	29/09/2020	1224515	4417392
ArFFKO06	2SBR1	maturation room - salami at 3 weeks of ripening	13/10/2020	1224515	4417392
ArFFKO07	3SWM2	stuffing room - water drainage swab	07/10/2020	1224515	4417392
ArFFKO08	3SBD3	drying room - salami	13/10/2020	1224515	4417392
ArFFKO09	3SBD5	drying room - salami	13/10/2020	1224515	4417392
ArFFKO10	3SBR1	maturation room - salami at 3 weeks of ripening	27/10/2020	1224515	4417392
ArFFKO11	4SBR3	maturation room - salami at 3 weeks of ripening	10/11/2020	1224515	4417392
ArFFKO12	6MB4	stuffing room - meat	18/11/2020	1224515	4417392

		mixture			
ArFFKO13	6STM5	stuffing room - surface swab	18/11/2020	1224515	4417392
ArFFKP01	5SBR104	storage room - salami 10 weeks of ripening	12/01/2021	1224515	4417392
ArFFKP02	5MB2	stuffing room - meat mixture	04/11/2020	1224515	4417392
ArFFKP03	5SWM2	stuffing room - water drainage swab	04/11/2020	1224515	4417392
ArFFKP04	5SWM4	stuffing room - water drainage swab	04/11/2020	1224515	4417392
ArFFKP05	5SBR4	maturation room - salami at 3 weeks of ripening	24/11/2020	1224515	4417392
ArFFKP06	5SM1	stuffing room - filler stuffer machine swab	04/11/2020	1224515	4417392
ArFFKP07	MB1	stuffing room - meat mixture	01/07/2020	1224515	4417392
ArFFKP08	MB2	stuffing room - meat mixture	01/07/2020	1224515	4417392
ArFFKP09	MB4	stuffing room - meat mixture	01/07/2020	1224515	4417392
ArFFKP10	MB5	stuffing room - meat mixture	01/07/2020	1224515	4417392
ArFFKP11	SBD3	drying room - salami	07/07/2020	1224515	4417392
ArFFKP12	2MB2	stuffing room - meat mixture	23/09/2020	1224515	4417392
ArFFKP13	2MB5	stuffing room - meat mixture	23/09/2020	1224515	4417392
ArFFKP14	3MB2	stuffing room - meat mixture	07/10/2020	1224515	4417392
ArFFKP15	3SWD3	drying room - water drainage swab	13/10/2020	1224515	4417392
ArFFKP16	5SBD1	drying room - salami	10/11/2020	1224515	4417392
ArFFKP17	5SBD5	drying room - salami	10/11/2020	1224515	4417392
ArFFKP18	6MB2	stuffing room - meat mixture	18/11/2020	1224515	4417392
ArFFKP19	6SBR281	storage room - salami 28 weeks of ripening	01/06/2021	1224515	4417392
ArFFKP20	6SBR283	storage room - salami 28 weeks of ripening	01/06/2021	1224515	4417392
ArFMSE01	SALM1	merguez sausage	22/10/2019	30.394910286035035	-9.581101061366358
ArFMSE02	SALM10	merguez sausage	14/01/2020	30.4078848	-9.5752582
ArFMSE03	SALM2	merguez sausage	22/10/2019	30.394910286035035	-9.581101061366358
ArFMSE04	SALM24	merguez sausage	13/10/2020	30.4016729	-9.5498077
ArFMSE05	SALM25	merguez sausage	13/10/2020	30.399505200195353	-9.548577969316822
ArFMSE06	SALM3	merguez sausage	22/10/2019	30.394910286035035	-9.581101061366358
ArFMSEA01	ST.AU.3	merguez sausage	22/10/2019	30.394910286035035	-9.581101061366358

Whole genome sequencing

Genomic DNA was extracted and purified using MagAttract HMW DNA Kit (Qiagen, Hilden, Germany), following manufacture recommendations. A video tutorial was designed by UNIBO and shared among partners. Partners extracted DNA which was subsequently sent to UNIBO. UNIBO sequenced and analysed the genomes. Libraries were built using the Nextera® XT DNA Library Preparation Kit (Illumina, Milan, Italy) and sequenced on Illumina MiSeq Platform which generates tagged 250 bp paired

end reads. Unless otherwise indicated, all subsequent genomic analyses were performed on Galaxy Trakr platform (Gangiredla et al., 2021). After performing a quality cleaning through TRIMMOMATIC, reads were de novo assembled using Unicycler v0.4.8.0, a standardized and fully automatic open-source pipeline for bacterial genome assembly. The quality assessment of the assemblies was verified using QUAST.

Species confirmation

Species of the isolates were confirmed by biotyping (RapID™ ONE System, ThermoFisher Scientific, Milan, Italy) and/or PCR (Rahn et al., 1992 (*Salmonella*), Perelle et al., 2004 (*E. coli*), Saraiva et al., 2018 (*S. aureus*), Chander et al., 2011 (*Klebsiella spp.*)) or MALDI/TOF.

RefSeq Masher Matches v0.1.2 was used to confirm the species at the genome level. This pipeline finds the closest matching of the query genomes sequence to NCBI Ref seq genomes (Ondov et al., 2016). Exclusively for *Klebsiella* species, Kleborate v2.0.0 was used as bioinformatic pipeline for species confirmation (Lam et al., 2021).

Multi-Locus Sequence Typing

For the assignment of genomes to ST-types, de novo assemblies were submitted to MLST v2.19.0, an open-source software which scan contig files against traditional PubMLST typing schemes and perform in silico MLST (Maiden et al., 1998). Exclusively for *Klebsiella pneumoniae* genomes, Kleborate was used for MLST typing based on curated schemes hosted in the BIGSdb platform of Institut Pasteur (<http://bigsdb.pasteur.fr/klebsiella/klebsiella.html>). Moreover, *Klebsiella oxytoca* isolates were further submitted to PubMLST (Jolley et al., 2018) for new allele an ST-type assignment if no STs were previously identified.

SNP calling

SNP calling was performed using the open-source software snippy v4.5.0 applied on reads (Seeman, 2015). Snippy is a rapid haploid variant calling and core genome alignment open-source tool. The pipeline includes several tools that align reads or assemblies from each isolate to a reference genome and then identifies variants among the alignments. The following genomes were used as references for SNP calling: ArFCLM01 as *L. monocytogenes* reference genome; *Staphylococcus aureus* NCTC 3825 (GenBank accession N° GCF_000013425.1) as *Staphylococcus aureus* reference genome; ArFASE04 as *Salmonella spp.* reference genome and *Klebsiella oxytoca* ASM381292v1 (GenBank accession N° NZ_CP033844.1) as *Klebsiella spp.* reference genome. Based on the core SNP alignment, a high-resolution phylogeny tree was built including the conserved nucleotide variant sites shared by all genomes. IQ-TREE v2.1.2 (Schrempf et al., 2016) and PhyML v3.3.2 (Guidon et al., 2010) was used to analyse the SNP differences between isolates based on maximum likelihood algorithm and phylogenetic trees were visualized with iTOL (Letunic et al., 2021). Finally, a pairwise SNP distance matrix was built using snp-dists v0.6.3.

Resistome and virulome

Analyses of the resistome and virulome of all genomes were performed using ABRicate v1.0.1. This tool performs a BLAST search of genes using ResFinder (Zankar et al., 2012) and VirulenceFinder databases (VFBD, Chen *et al*, 2016). Genomes were considered positive for AMR or virulence genes when the gene showed > 80% of coverage and > 80% of identity.

Summary of results

Listeria monocytogenes

Among the ArtiSaneFood bacterial pathogens, *Listeria monocytogenes* was identified in 11 samples collected from Spanish Iberian raw-cured sausage “salchichón” production (raw ingredient – meat, surface and final food product) as well as in 4 samples collected from Italian Emilia-Romagna salami (water drainage swabs of the drying room) (Table 1). Isolates were confirmed as belonging to *L. monocytogenes* species by both MALDI/TOF and PCR. All strains but one (ArFCLM11 – *Enterococcus faecalis*) were confirmed by WGS based approaches (Masher Matches v0.1.2) and submitted to following WGS analyses. Statistics of *de novo* assemblies showed the high quality of sequenced genomes with number of contigs ranging from 25 to 32, N50 ranging from 262311 to 533374, and largest contig from 505969 to 1222483 (Table 2).

Table 2: Assembly statistics of *L. monocytogenes* genomes

Assembly	# contigs (≥ 0 bp)	# contigs (≥ 1000 bp)	Total length (≥ 0 bp)	Total length (≥ 1000 bp)	# contigs	Largest contig	Total length	GC (%)	N50	N75	L50	L75
ArFCLM01	31	15	2968902	2963776	18	744482	2965969	37.95	476695	253636	3	5
ArFCLM02	30	17	2973706	2968273	22	1222483	2971958	37.96	516329	253636	2	4
ArFCLM03	28	13	2914813	2907827	19	1203972	2912122	37.91	520718	264607	2	4
ArFCLM04	28	18	3007255	3003548	21	587682	3005519	37.82	438637	215721	3	6
ArFCLM05	29	21	2799091	2796007	24	754951	2798143	37.88	434699	228550	3	5
ArFCLM06	31	12	2915390	2907586	18	1203972	2911633	37.91	520718	264607	2	4
ArFCLM07	32	19	2951950	2947207	23	505969	2949800	37.84	408574	215721	4	6
ArFCLM08	29	14	2932553	2927010	17	1203972	2929150	37.91	533374	264607	2	4
ArFCLM09	25	18	3006740	3003554	21	587682	3005716	37.82	438643	215721	3	6
ArFCLM10	28	17	3006828	3002745	21	587683	3005338	37.82	438901	215721	3	6
ArFCLM11	29	18	2873281	2868405	22	1095809	2871502	37.45	262311	217204	3	6
ArFFLM01	26	15	2866457	2863169	17	745246	2864287	37.94	475136	253175	3	5
ArFFLM02	25	14	2860941	2857497	16	745246	2858775	37.93	475183	253001	3	5
ArFFLM03	25	14	2860650	2857206	16	745246	2858484	37.92	475183	252911	3	5
ArFFLM04	25	14	2860779	2857335	16	745246	2858613	37.93	475183	253043	3	5

Based on SNP calling, a phylogenetic tree of all *L. monocytogenes* genomes was inferred. As expected, genomes clustered according to the source (Italian salami or Spanish raw-cured sausage) and ST-Type. The 14 genomes were gathered in 5 clades: clade 1 including the 5 Spanish ST8 genomes, clade 2 with only one genome of ST451, clade 3 with 3 Spanish ST1 genomes, clade 4 with three Spanish ST3 genomes and clade 5 with 4 Italian ST489 genomes. Based on the phylogenetic tree clustering and the SNP distance matrix (Table 3), data suggest that different clones of *L. monocytogenes* were circulating in the Spanish raw-cured sausage plant from July to October 2020 (Figure 1, table 3). On the contrary, in the Italian salami plant, a single clone of *L. monocytogenes* was observed specifically isolated from water drainage swab samples in September 2020. The same Italian clone was not observed in following

samples, neither in the environment nor in raw ingredients and food samples. Regarding ST-types, ST1 and ST8 are frequently identified in clinical settings, suggesting the concern for the potential transfer of these isolates from the Spanish raw-cured sausage plant to humans through food consumption.



Figure 1: Phylogenetic tree of *L. monocytogenes* genomes

Table 3: SNPs distance matrix of *L. monocytogenes* genomes

snp-dists 0.	ArFCLM09	ArFCLM10	ArFCLM01	ArFCLM02	ArFCLM03	ArFCLM04	ArFCLM05	ArFCLM06	ArFCLM07	ArFCLM08	ArFFLM01	ArFFLM02	ArFFLM03	ArFFLM04	Reference
ArFCLM09	0	644	114127	114126	114100	732	23763	114096	599	114095	114009	114022	114024	114029	114127
ArFCLM10	644	0	114067	114066	114040	726	23844	114040	625	114051	113961	113962	113966	113977	114067
ArFCLM01	114127	114067	0	5	9140	114018	113106	9144	114146	9157	8619	8591	8602	8584	0
ArFCLM02	114126	114066	5	0	9139	114017	113105	9143	114145	9156	8618	8590	8601	8583	5
ArFCLM03	114100	114040	9140	9139	0	113996	113170	65	114109	73	9045	9029	9050	9036	9140
ArFCLM04	732	726	114018	114017	113996	0	23858	113999	739	114002	113896	113910	113909	113918	114018
ArFCLM05	23763	23844	113106	113105	113170	23858	0	113161	23785	113160	112937	112955	112953	112959	113106
ArFCLM06	114096	114040	9144	9143	65	113999	113161	0	114109	68	9035	9025	9034	9032	9144
ArFCLM07	599	625	114146	114145	114109	739	23785	114109	0	114110	114022	114031	114035	114043	114146
ArFCLM08	114095	114051	9157	9156	73	114002	113160	68	114110	0	9032	9038	9049	9029	9157
ArFFLM01	114009	113961	8619	8618	9045	113896	112937	9035	114022	9032	0	62	51	65	8619
ArFFLM02	114022	113962	8591	8590	9029	113910	112955	9025	114031	9038	62	0	55	79	8591
ArFFLM03	114024	113966	8602	8601	9050	113909	112953	9034	114035	9049	51	55	0	58	8602
ArFFLM04	114029	113977	8584	8583	9036	113918	112959	9032	114043	9029	65	79	58	0	8584
Reference	114127	114067	0	5	9140	114018	113106	9144	114146	9157	8619	8591	8602	8584	0

Regarding antimicrobial resistance prediction, as it is often the case for *L. monocytogenes*, few AMR associated genes were detected. In particular, all genomes but one (ArFCLM11) carried exclusively the *fosX* gene associated to fosfomycin resistance. ArFCLM11 carried the *lsa(A)* and *tet(M)* genes associated to lincosamides/streptogramin A and tetracycline resistance respectively.

Regarding virulence prediction, all 14 *Listeria monocytogenes* genomes but one (ArFCLM05, ST451) showed a full length *inlA* gene (Figure 2). *Listeria* Pathogenic Islands 1 (LIPI 1) was detected in all genomes, however ST1 genomes lacked *actA* gene generally located in LIPI 1 along with *prfA*, *actA*, *hly*, *mpl*, *plcA*, *plcB*, and *iap*. In addition, ST1, ST3 and ST489 genomes included LIPI 3 (*llsA*, *llsG*, *llsH*, *llsX*, *llsB*, *llsY*, *llsD*, *llsP*), which has been associated, together with LIPI 1, to increased virulence and specifically to invasiveness (*llsX* gene of LIPI 3) (Vilchis-Rangel et al., 2019).

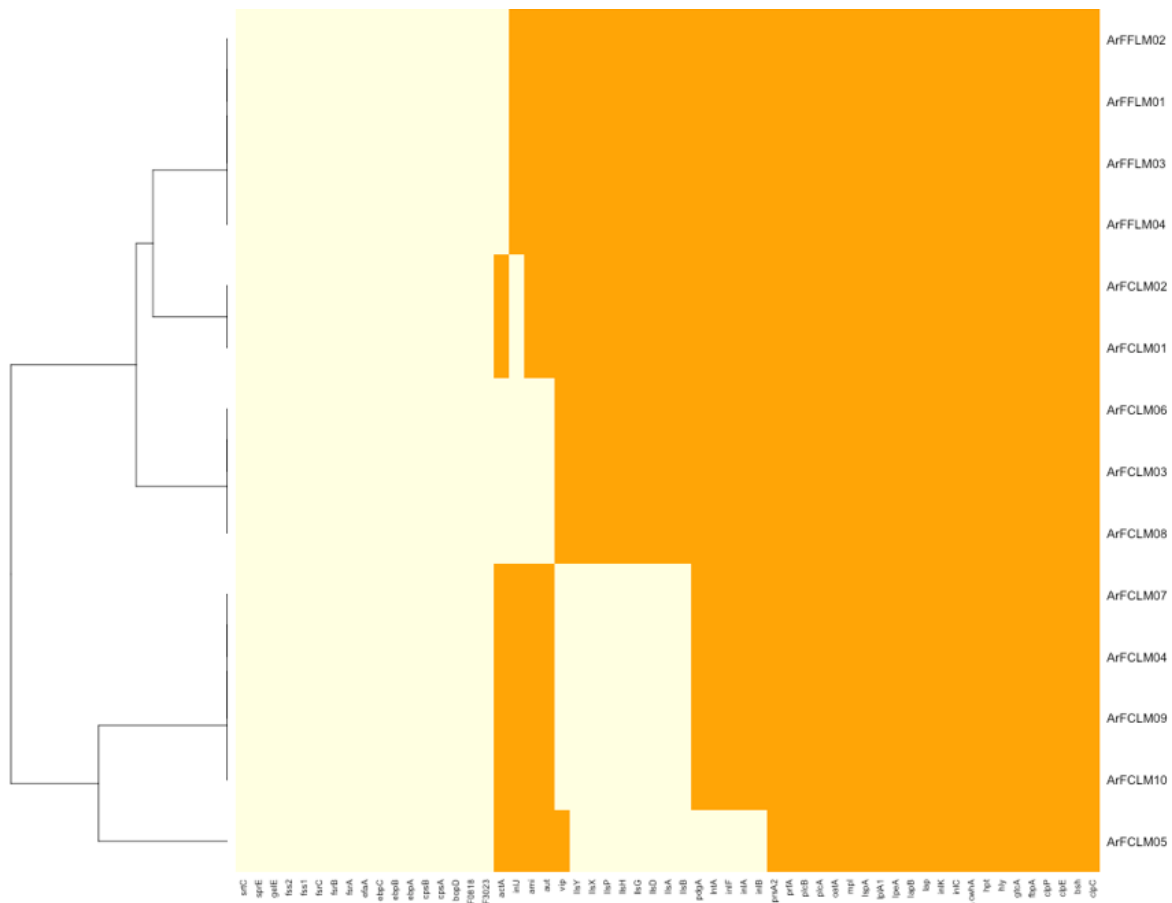


Figure 2: Heatmap of virulome of *L. monocytogenes* isolates (yellow: absence (<80% of sequence identity), orange: presence (>80% of sequence identity))

Salmonella enterica

Among the Artisanefood bacterial pathogens, *Salmonella enterica* was identified in 11 samples collected from Portuguese Alheira Graganza sausage production (raw ingredient – meat, final food product and the environment - fair), as well as one isolate, collected from Italian soft cheese production (maturation room), and in 6 Moroccan merguez sausages (Table 1). Isolates were confirmed as belonging to *Salmonella* species by MALDI/TOF and/or biochemical tests. All strains but four (ArFASE01 – *Hafnia paralvei*; ArFASE08 – *Proteus mirabilis*; ArFASE09 – *Escherichia coli*; ArFASE10 - *Proteus mirabilis*) were confirmed as belonging to *Salmonella enterica* by WGS based approaches (Masher Matches v0.1.2) and submitted to following WGS analyses. In particular, the Italian isolate belongs to *Salmonella enterica subsp. enterica* serovar Gallinarum/pullorum; Portuguese isolates to *Salmonella enterica subsp. enterica* serovar Paratyphi B and Moroccan isolates to *Salmonella enterica subsp. enterica* serovars Hadar (2 isolates, ArFMSE01 and ArFMSE04), Kentucky (ArFMSE02), Montevideo (ArFMSE03), Albany (ArFMSE05), Seftemberg (ArFMSE06).

Statistics of *de novo* assemblies showed the high quality of sequenced genomes (except for ArFESE01 which was omitted from further WGS analyses) with number of contigs ranging from 23 to 56, N50 ranging from 231038 to 614135, and largest contig from 521807 to 1308384 (Table 4).

Table 4: Assembly statistics of *S. enterica* genomes

Assembly	# contigs (>= 0 bp)	# contigs (>= 1000 bp)	Total length (>= 0 bp)	Total length (>= 1000 bp)	# contigs	Largest contig	Total length	GC (%)	N50	N75	L50	L75
ArFASE02	77	46	4871281	4860097	55	551422	486703	52.13	319689	133461	6	13
ArFASE03	79	47	4871099	4859817	56	551424	486670	52.13	319689	145402	6	13
ArFASE04	75	46	4871712	4860646	55	644777	486758	52.13	231038	145402	6	12
ArFASE05	76	45	4870999	4859815	54	551383	486674	52.13	364216	145402	6	12
ArFASE06	73	44	4871497	4860431	53	644777	486736	52.13	399502	145402	5	11
ArFASE07	76	45	4870999	4859817	54	551385	486675	52.13	364216	145402	6	12
ArFASE11	76	45	4871003	4859820	54	551422	486675	52.13	364216	145402	6	12
ArFESE01	825	777	3879638	3858127	794	51427	387032	50.85	7066	3891	154	339
ArFMSE01	46	31	4673259	4667316	36	735081	467102	52.22	356366	262576	5	8
ArFMSE02	43	19	4669097	4661849	23	728786	466483	52.13	518238	322975	4	7
ArFMSE03	44	17	4705824	4696630	25	113457	470218	52.24	614135	312808	3	6
ArFMSE04	53	30	4703834	4695907	35	130838	469994	52.23	613464	262534	3	6
ArFMSE05	62	30	4636149	4626924	35	521807	463064	52.18	323876	192517	6	10
ArFMSE06	54	28	4914160	4904511	37	166146	491137	52.09	557385	237646	3	6
ArFASE02	77	46	4871281	4860097	55	551422	486703	52.13	319689	133461	6	13

Based on SNP calling, a phylogenetic tree of all *Salmonella enterica* genomes was inferred. As expected, genomes clustered according to the serovar and ST-Type. The 13 genomes were gathered in 7 clades: clade 1 including the 7 Portuguese *S. Gallinarum*/Pullorum ST43 genomes, clade 2 with the two Moroccan *S. Hadar* genomes of ST33, clades 3, 4, 5 and 6 each including one Moroccan isolate: *S. Kentucky* (ST314), *S. Montevideo* (ST3667), *Albany* (ST1818), *Seftemberg* (ST198). Based on the phylogenetic tree clustering and the SNP distance matrix, data suggest that one clone was circulating among the Portuguese sausage fair from November 2019 to May 2020 with SNP differences from 0 to 1. (Figure 2, Table 5). On the contrary, in the Moroccan sausage plant, different *Salmonella* serovars were observed. Regarding ST-types, ST43 of *S. Paratyphi* B of the Portuguese genomes has been already described in human infections in Europe, UK and South America. *S. Hadar* ST33 has been already described in animals and food of animal origin.

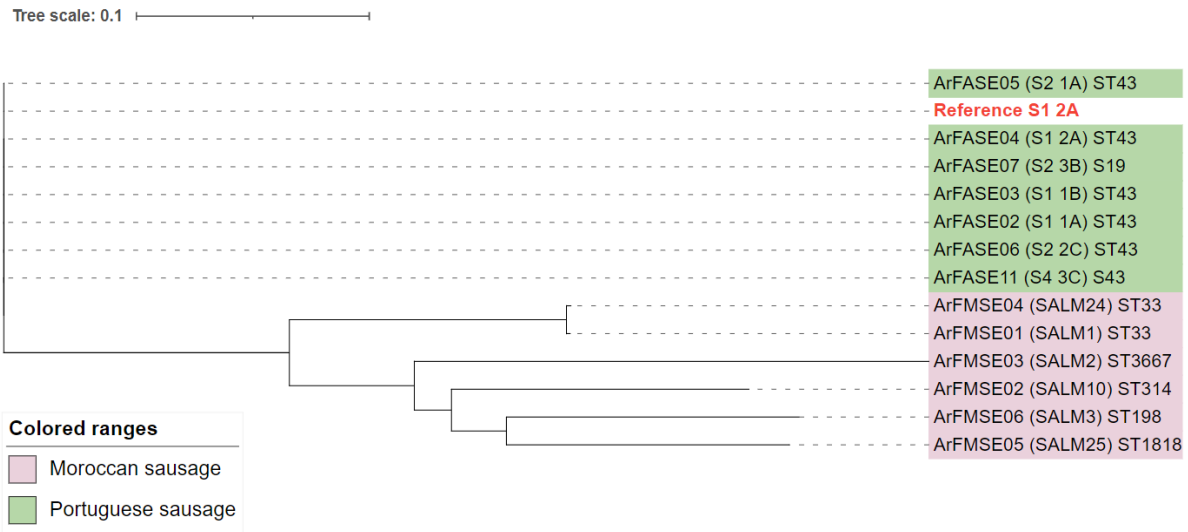


Figure 3: Phylogenetic tree of *Salmonella enterica* genomes

Table 5: SNPs distance matrix of *S. enterica* genomes

snp-dists 0.6.3	Reference	ArFASE02	ArFASE03	ArFASE04	ArFASE05	ArFASE06	ArFASE07	ArFASE11	ArFMSE02	ArFMSE01	ArFMSE04	ArFMSE05	ArFMSE03	ArFMSE06
Reference	0	0	0	0	0	1	0	0	31373	26328	26387	32591	37737	32700
ArFASE02	0	0	0	0	0	1	0	0	31373	26328	26387	32591	37737	32700
ArFASE03	0	0	0	0	0	1	0	0	31373	26328	26387	32591	37737	32700
ArFASE04	0	0	0	0	0	1	0	0	31373	26328	26387	32591	37737	32700
ArFASE05	0	0	0	0	0	1	0	0	31373	26328	26387	32591	37737	32700
ArFASE06	1	1	1	1	1	0	1	1	31374	26329	26388	32592	37738	32701
ArFASE07	0	0	0	0	0	1	0	0	31373	26328	26387	32591	37737	32700
ArFASE11	0	0	0	0	0	1	0	0	31373	26328	26387	32591	37737	32700
ArFMSE02	31373	31373	31373	31373	31373	31374	31373	31373	0	31196	31193	27336	36308	28344
ArFMSE01	26328	26328	26328	26328	26328	26329	26328	26328	31196	0	411	32233	38427	32577
ArFMSE04	26387	26387	26387	26387	26387	26388	26387	26387	31193	411	0	32211	38405	32555
ArFMSE05	32591	32591	32591	32591	32591	32592	32591	32591	27336	32233	32211	0	35678	26926
ArFMSE03	37737	37737	37737	37737	37737	37738	37737	37737	36308	38427	38405	35678	0	35125
ArFMSE06	32700	32700	32700	32700	32700	32701	32700	32700	28344	32577	32555	26926	35125	0

Regarding antimicrobial resistance prediction in *Salmonella* isolates, all isolates carried the *aac(6')-Iaa* gene conferring aminoglycoside resistance. Portuguese *Salmonella* genomes do not carry additional AMR genes, whereas both genomes of Moroccan *S. Hadar* (ArFMSE01 and ArFMSE04) additionally carried *aph(3'')-Ib*, *aph(6)-Id* (both associated to aminoglycoside resistance) and *tet(A)* (tetracycline resistance). Along with *S. Hadar*, Moroccan *S. Seftemberg* showed a multiresistance predicted profile with additional genes *aadA2* (aminoglycoside resistance), *blaTEM1-B* (beta-lactam resistance), *dfra12* (trimethoprim resistance), *sul1* (sulphonamide resistance) and *tet(A)*.

Regarding virulence prediction, all 13 *Salmonella* genomes carried from 100 to 107 virulence genes (Figure 4). Genes of virulence plasmid pSV were not detected. However, virulence genes of *Salmonella* pathogenicity islands SP1 (*orgABC*, *prgHIJK*, *sipABCD*, *sicAP*, *spaOPQRS*, *invABCEFGHIJ*), SP2 (*ssaGHIJKLMNOPQRSTUV*, *sscAB*, *sseABCDEFGHIJK1K2L*, *ssaCDE*,) and SP3 (*misL*, *mgtBC*) were detected along with *sopE2* gene described in the literature as virulence marker of UK and Italian *S. Typhimurium* monophasic variant clades (Marcus et al., 2000; Palma et al., 2019)

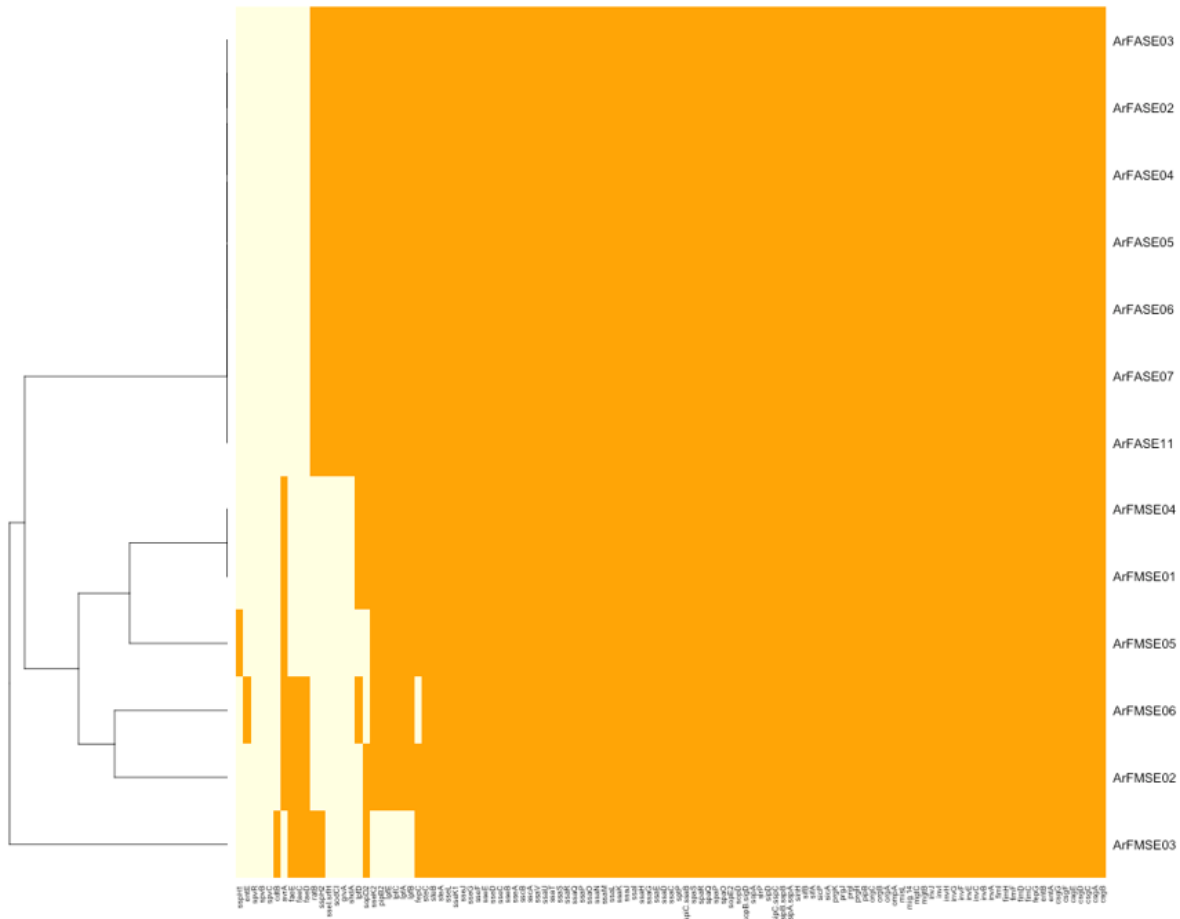


Figure 4: Heatmap of virulome of *Salmonella enterica* isolates (yellow: absence (<80% of sequence identity), orange: presence (>80% of sequence identity))

Staphylococcus aureus

Among the Artisanefood bacterial pathogens, *Staphylococcus aureus* was identified in 37 samples: 9 samples collected from Portuguese Alheira Graganza and Alheira Vinhais sausage production (raw ingredient – meat, final food product), as well as in 4 samples collected from Portuguese ‘Queijo de Cabra Transmontano’ cheese, 8 (raw ingredient – meat, final sausage product) and 4 (raw ingredient – milk, final cheese product) samples from Spanish Iberian raw-cured sausage “salchichón” and ‘Queso Manchego’ cheese respectively; 6 samples each from Italian Squaquerone di Romagna cheese (final cheese product) and Italian salami (raw material-meat, processing environment, salami during ripening) and one sample from Moroccan merguez sausage (Table 1). Isolates were confirmed as belonging to *Staphylococcus aureus* species by MALDI/TOF and/or biochemical tests and PCR. Of the 38 *S. aureus* isolates, 12 were not sequenced due to low DNA quality. On the 25 isolates sequenced, 10 were not confirmed by WGS analyses (Masher Matches v0.1.2) being identified as *S. warneri* or *S. epidermidis*. Overall, 15 *S. aureus* genomes were retained for further WGS analyses: ArFCSA01 to ArFCSA08 from Spanish sausage, ArFDSA01, ArFDSA03 and ArFDSA04 from Spanish cheese, ArFESA01, ArFESA03 and ArFESA04 from Italian cheese, ArFFSA03 from Italian salami and ArFMSA01 from Moroccan sausage.

Statistics of *de novo* assemblies showed the high quality of sequenced genomes (except for ArFESA04 which was omitted by subsequent WGS analyses) with number of contigs ranging from 15 to 39, N50 ranging from 230321 to 815868, and largest contig from 372180 to 1017875 (Table 6).

Table 6: Assembly statistics of *S. aureus* genomes

Assembly	# contigs (≥ 0 bp)	# contigs (≥ 1000 bp)	Total length (≥ 0 bp)	Total length (≥ 1000 bp)	# contigs	Largest contig	Total length	GC (%)	N50	N75	L50	L75
ArFCSA01	35	20	2664558	2660759	22	564837	2662322	32.68	306107	151399	3	6
ArFCSA02	38	22	2718693	2714306	25	406882	2716501	32.69	289679	160598	4	7
ArFCSA03	47	27	2758691	2752904	31	556881	2755786	32.70	306107	87424	3	7
ArFCSA04	31	15	2670475	2666015	18	539965	2668175	32.71	397333	315969	3	5
ArFCSA05	38	18	2669035	2663209	21	556767	2665322	32.72	315969	248098	4	6
ArFCSA06	24	12	2671095	2667009	15	873106	2669287	32.72	603808	397288	2	4
ArFCSA07	39	21	2731332	2725158	25	865930	2728126	32.69	815868	172601	2	4
ArFCSA08	39	23	2729924	2724468	27	851790	2727569	32.69	815867	172601	2	4
ArFDSA01	43	35	2743255	2741507	36	432694	2742101	32.70	181824	93410	5	10
ArFDSA03	49	37	2700499	2697908	39	432694	2698930	32.69	230321	70196	4	10
ArFDSA04	36	18	2739656	2735845	19	609892	2736652	32.80	339956	151591	3	6
ArFESA01	32	19	2738019	2735258	20	1010428	2735825	32.71	493801	127524	2	5
ArFESA03	40	20	2752746	2747700	22	889547	2749123	32.65	304438	175520	3	5
ArFESA04	2	2	3191	3191	2	2046	3191	29.74	2046	1145	1	2
ArFFSA03	29	26	2780891	2780488	26	1017875	2780488	32.77	469582	167730	2	5
ArFMSA01	39	23	2689124	2684331	26	372180	2686581	32.66	243624	158919	5	8

Based on SNP calling, a phylogenetic tree of all *S. aureus* genomes was inferred. As expected, genomes clustered according to the serovar, ST-Type and country. The 15 genomes were gathered in 7 clades. Clade 1 included the 2 *S. aureus* ST121 genomes originating from Spanish cheese and collected the same day in September 2020 (40 SNPs of difference). Clades 2, 3 and 4 gathered only one genome each belonging to ST398 (Spanish cheese), ST5 (Italian salami) and ST1 (Italian soft cheese) respectively. Clade 5 included 3 *S. aureus* ST7 strains collected from Spanish sausage in the time frame of one month from September to October 2020 (from 10 to 37 SNPs difference). Clade 6 gathered three strains belonging to ST15 and one *S. aureus* strain to which the ST-Type was not attributed. In clade 6 one strain (SA3) originated from Moroccan sausage, whereas the other three (ArFCSA01, ArFCSA02, ArFCSA03) were all from Spanish meat used for sausage production collected on the same day in September 2020 (28 to 36 SNPs of difference). Finally, clade 7 gathered three ST8 *S. aureus*, one from Italian soft cheese and two from Spanish meat for sausage production. The two strains from Spanish meat (ArFCSA07, ArFCSA08) were isolated the same day in August 2020 and showed only one SNP of difference.

Based on the phylogenetic tree clustering and the SNP distance matrix, data suggest that different clones of *S. aureus* were co-existing among Spanish sausage production as well as Italian soft cheese (Figure 3, Table 7). Regarding ST-types, ST121 is a *S. aureus* globally disseminated hypervirulent clone (Rao et al.,

2015). ST5 and ST8 have been associated to hospital acquired Methicillin Resistant *Staphylococcus aureus* (HA-MRSA) and ST398 was found both in humans and pig/pork meat (Deurenberg et al., 2007; van Belkum et al., 2008). ST1 and ST5 and ST15 have also been described in humans, food and wildlife (Lv et al., 2021; Heaton et al., 2020).

Figure 5: Phylogenetic tree of *Staphylococcus aureus* genomes

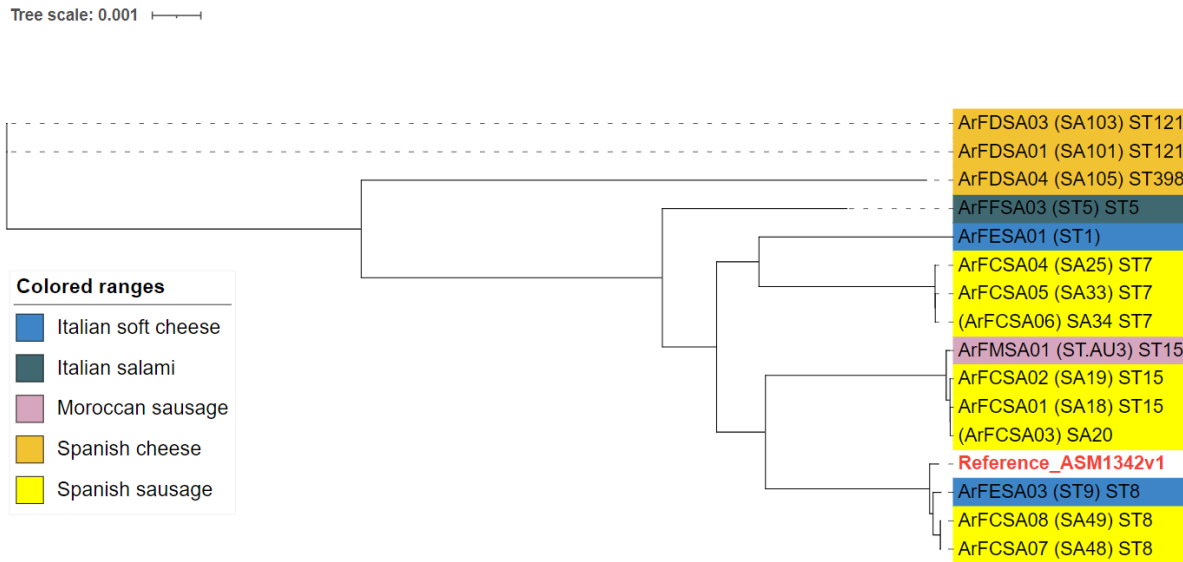


Table 7: SNPs distance matrix of *S. aureus* genomes

snp-dists 0.6	Reference	ArFDSA01	ArFDSA03	ArFDSA04	ArFCSA01	ArFCSA02	ArFCSA03	ArFCSA04	ArFCSA05	ArFCSA06	ArFMSA01	ArFCSA07	ArFCSA08	ArFESA01	ArFFSA03	ArFESA03
Reference	0	39067	38693	47105	16230	17163	17348	17550	17523	17461	15979	903	910	18885	19880	963
ArFDSA01	39067	0	40	40344	36982	37132	37357	37076	37028	36979	36880	38801	38825	37374	37650	38749
ArFDSA03	38693	40	0	40315	36888	37018	37076	36957	36909	36859	36811	38457	38469	37069	37310	38468
ArFDSA04	47105	40344	40315	0	44900	45168	45206	45032	45013	44986	44847	45895	45897	46007	45576	45935
ArFCSA01	16230	36982	36888	44900	0	36	28	17099	17100	17070	348	16142	16142	17468	18041	16189
ArFCSA02	17163	37132	37018	45168	36	0	34	17176	17176	17159	347	16386	16381	17627	18212	16409
ArFCSA03	17348	37357	37076	45206	28	34	0	17227	17223	17193	342	16405	16403	17848	18364	16623
ArFCSA04	17550	37076	36957	45032	17099	17176	17227	0	17	23	16970	17506	17514	17359	18269	17578
ArFCSA05	17523	37028	36909	45013	17100	17176	17223	17	0	10	16976	17492	17501	17364	18247	17554
ArFCSA06	17461	36979	36859	44986	17070	17159	17193	23	10	0	16949	17428	17434	17315	18227	17501
ArFMSA01	15979	36880	36811	44847	348	347	342	16970	16976	16949	0	15974	15972	17374	17958	16009
ArFCSA07	903	38801	38457	45895	16142	16386	16405	17506	17492	17428	15974	0	1	18099	19052	701
ArFCSA08	910	38825	38469	45897	16142	16381	16403	17514	17501	17434	15972	1	0	18118	19062	701
ArFESA01	18885	37374	37069	46007	17468	17627	17848	17359	17364	17315	17374	18099	18118	0	19890	18215
ArFFSA03	19880	37650	37310	45576	18041	18212	18364	18269	18247	18227	17958	19052	19062	19890	0	19106
ArFESA03	963	38749	38468	45935	16189	16409	16623	17578	17554	17501	16009	701	701	18215	19106	0

Regarding antimicrobial resistance prediction in *S. aureus*, isolates ArFDSA01, ArFDSA03, ArFDSA04, ArFCSA01, ArFCSA02, ArFCSA03, ArFFSA03, ArFESA03, ArFMSA01 carried *blaZ* gene associated to beta-lactamase production and beta-lactam resistance. Although generally located close to *blaZ*, no *mec* genes were identified predicting all *S. aureus* strains as methicillin susceptible (Hiramatsu et al., 2013). ArFDSA01 and ArFDSA03 additionally carried *str* gene associated to streptomycin resistance. ArFDSA04 additionally carried *tet(M)* and ArFCSA01, ArFCSA02, ArFCSA03 and ArFMSA01 additionally carried *tet(K)* gene both associated to tetracycline resistance. No resistance associated genes were detected in genomes ArFCSA04, ArFCSA05, ArFCSA06, ArFCSA07, ArFCSA08, ArFESA01. None isolate was predicted as multiresistant.

Regarding virulence prediction, all 15 *S. aureus* genomes carried from 54 to 67 virulence genes (Figure 6). ST121 genomes, described as associated to hypervirulent strains, carried the characteristic *lukS-lukF* genes coding for proteins LukS–PV and LukF–PV responsible of the assembly of PVL a bicomponent pore-forming cytotoxin closely related to the development of *S. aureus* infection (Hu et al.,

2015). The other 13 genomes but one (ArFDSA04) carried *lukF* but not *lukS*. Additionally, haemolysin related genes were found in all genomes (*hly*, *hld*, *hlgA*, *hlgBx*, *hlgC*). Enterotoxin related gene *seb* was found exclusively in ST121 (ArFDSA01, ArFDSA03) and one ST15 (ArFMSA01) genome (Rao et al., 2015).

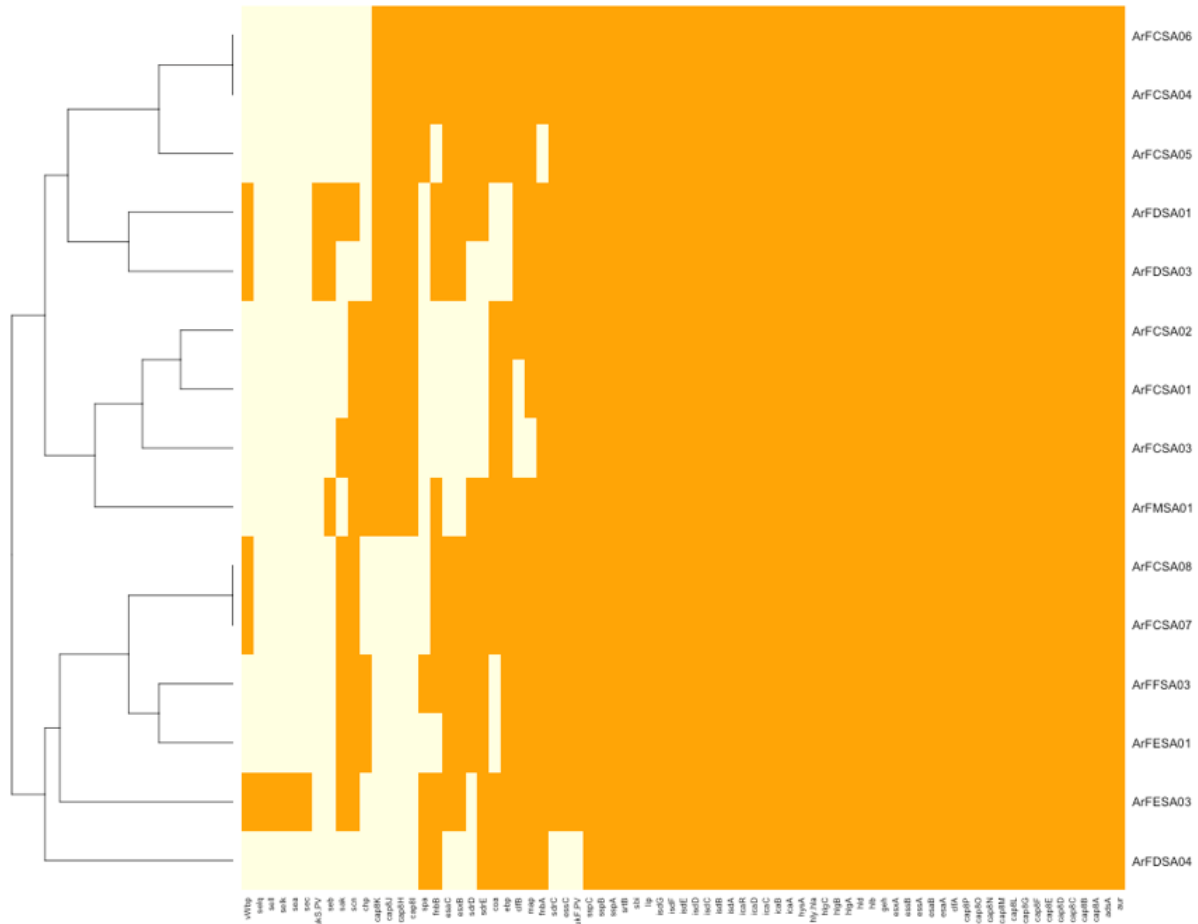


Figure 6: Heatmap of virulome of *Staphylococcus aureus* isolates (yellow: absence (<80% of sequence identity), orange: presence (>80% of sequence identity))

Klebsiella spp.

In parallel foodborne *Klebsiella* spp. was identified in 76 samples: 40 *K. oxytoca* and 3 *K. pneumoniae* from Italian soft cheese production (from the environment, cheese during production and cheese during storage) as well as 12 *K. oxytoca* and 20 *K. pneumoniae* from Italian salami production (from the environment, raw material-meat, salami during ripening) (Table 1). Isolates were confirmed as belonging to *K. oxytoca* and *K.pneumoniae* by biochemical tests and PCR. Of the 76 genomes, two were identified by WGS as *Citrobacter freundii* and were omitted from further analyses (ArFFO13 and ArFEKP02). Of the remaining 74 isolates were confirmed at species level by WGS analyses being identified as *K. michiganensis* (15), *K. planticola* (5), *K. ornithinolytica* (13) and *K. varicola subspecies varicola* (1), *K. grimontii* (7), *K.pasteurii* (5), *K.oxytoca* (11), *K pneumoniae* (16). This observation is in line with recent findings suggesting the limits of biochemical tests and PCR on the identification of *Klebsiella* species belonging to the same complex. In particular, only WGS based approaches can distinguish among *K.*

oxytoca, *K. huaxiensis*, *K. michiganensis*, *K. pasteurii* and *K. grimontii* all belonging to the ***K. oxytoca* species complex (KoSC)** and indistinguishable by morphology, biochemical tests or rRNA 16S sequence (Cosic et al., 2021; Merla et al., 2019; Wyres et al., 2020). Similarly, two other species complexes gather *Klebsiella* species otherwise indistinguishable. ***K. pneumoniae* species complex (KpSC)** gathers *K. pneumoniae*, *K. quasipneumoniae*, *K. variicola*, *K. quasivariicola* and *K. africana*. Three additional *Klebsiella* species, namely *K. ornithinolytica*, *K. planticola* and *K. electrica* were attributed to *Raoultella* complex based on the *gyrB* gene sequence although this is still under debate since it renders *Klebsiella* non-monophyletic (Figure 4).

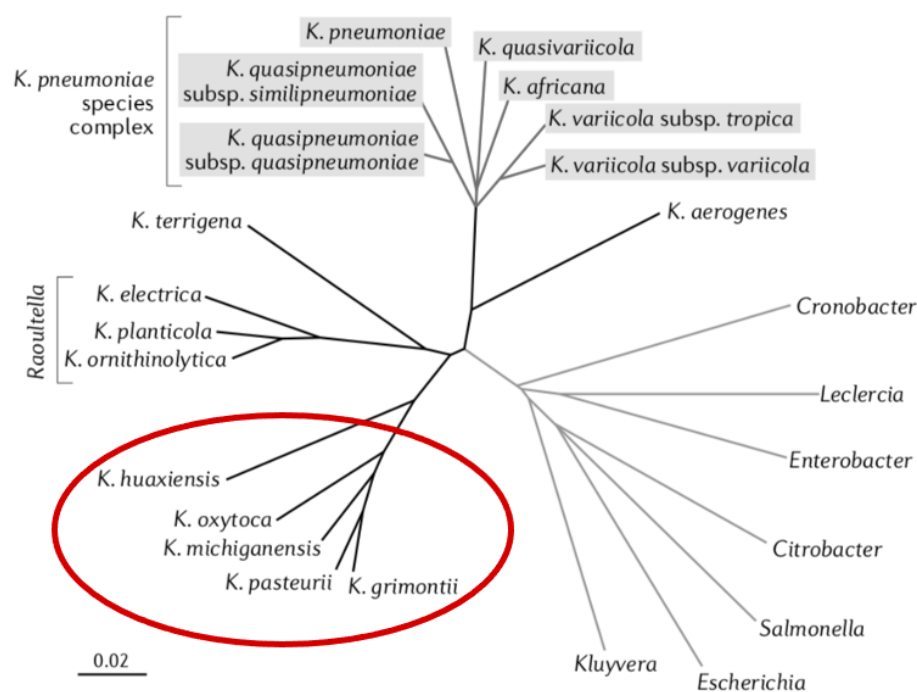


Figure 7: Whole-genome-based tree showing the phylogenetic relationships between *K. pneumoniae* and its close relatives (from Wyres et al., 2020). In red the *K. oxytoca* species complex (KoSC).

Along with foodborne *Klebsiella* of Italian soft cheese and salami, 10 publicly available sequences of *Klebsiella* spp. isolates collected from humans in Europe, were included for comparison (Table 8). The major purpose was to get insights into the potential risk for human health of foodborne *Klebsiella* in terms of virulence and antimicrobial resistance potentials.

Table 8: Public *Klebsiella* spp. clinical genomes (paired-end reads)

<i>Klebsiella</i> species	Country	Bioproject	SRA Code
<i>K pneumoniae</i>	Italy	PRJNA746285	SRR15123351
			SRR15123353
			SRR17799538
<i>K. oxytoca</i>	Italy	PRJNA801490	SRR17858119
	Spain	PRJNA626430	SRR11585239
<i>K variicola</i>	Italy	PRJEB10018	ERR1204816
			ERR1228222
<i>K. michiganensis</i>	Italy	PRJNA592166	SRR10579999
<i>K. planticola</i>	Denmark	PRJEB37711	ERR4013551
<i>K. ornithinolytica</i>	Denmark	PRJEB37711	ERR4013963

Statistics showed the high quality of *de novo* assemblies of Artisanefood genomes as well as public genomes with number of contigs ranging from 39 to 237, N50 ranging from 82701 to 522597, and largest contig from 688401 to 1296437 (Table 8).

Table 9: Assembly statistics of *Klebsiella* spp. genomes (new assigned ST-types are underlined)

Assembly	<i>Klebsiella</i> species (confirmed by WGS)	ST-types	# contigs (≥ 0 bp)	# contigs (≥ 1000 bp)	Total length (≥ 0 bp)	Total length (≥ 1000 bp)	# contigs	Largest contig	Total length	GC (%)	N50	N75	L50
ArFEKO01	<i>K.grimontii</i>	<u>378</u>	192	121	6024580	5999657	137	345658	6010925	55.75	152861	81762	15
ArFEKO02	<i>K.ornithinolytica</i>		89	61	5619865	5608986	70	460694	5616196	55.74	256407	118705	9
ArFEKO03	<i>K.ornithinolytica</i>		114	72	5618538	5603061	85	364082	5613019	55.74	200537	98049	11
ArFEKO04	<i>K.grimontii</i>	<u>377</u>	75	48	5909377	5899850	52	458158	5903045	55.84	233706	134976	9
ArFEKO05	<i>K.michiganensis</i>	95	91	71	5932762	5927966	72	653543	5928772	55.66	223609	104714	9
ArFEKO06	<i>K.pasteurii</i>	<u>387</u>	171	104	6356220	6328088	127	516195	6345575	55.08	173766	107795	12
ArFEKO07	<i>K.ornithinolytica</i>		89	61	5619699	5608797	70	460694	5616006	55.74	256407	115799	9
ArFEKO08	<i>K.ornithinolytica</i>		86	61	5616504	5606832	69	509566	5613099	55.74	256509	118706	9
ArFEKO09	<i>K.ornithinolytica</i>		85	58	5620016	5609290	67	705012	5616499	55.74	256374	118706	8
ArFEKO10	<i>K.michiganensis</i>	95	82	62	5953802	5948900	63	864770	5949832	55.65	331041	175196	5
ArFEKO11	<i>K.ornithinolytica</i>		104	78	5832696	5824972	83	641230	5828484	55.59	237411	144704	8
ArFEKO12	<i>K.michiganensis</i>	95	78	64	5953535	5949475	65	959902	5950407	55.65	290829	139915	7
ArFEKO13	<i>K.grimontii</i>	<u>377</u>	48	27	5801307	5794216	31	938773	5797411	55.86	447749	357337	5
ArFEKO14	<i>K.michiganensis</i>	95	78	61	5948934	5944581	62	1059595	5945513	55.65	301049	173064	6
ArFEKO15	<i>K.michiganensis</i>	183	144	95	6434707	6414765	111	915433	6426439	55.23	243590	98555	7
ArFEKO16	<i>K.michiganensis</i>	95	80	61	5952329	5947686	62	1059532	5948492	55.65	331041	139915	5
ArFEKO17	<i>K.michiganensis</i>	<u>380</u>	105	66	6429032	6415117	75	488477	6421182	55.50	244595	156197	9
ArFEKO18	<i>K.michiganensis</i>	<u>380</u>	105	67	6431785	6418108	76	623058	6424174	55.50	230407	135159	9
ArFEKO19	<i>K.ornithinolytica</i>		136	85	5851017	5835216	93	879833	5841058	55.60	338022	160092	6
ArFEKO20	<i>K.michiganensis</i>	<u>380</u>	111	68	6432091	6417313	77	481622	6423378	55.50	244595	135159	9
ArFEKO21	<i>K.michiganensis</i>	50	109	55	6148222	6126784	77	831376	6141081	55.78	380765	178586	6
ArFEKO22	<i>K.grimontii</i>	<u>378</u>	237	137	6178468	6143695	158	319825	6159396	55.69	152188	75198	16
ArFEKO23	<i>K.pasteurii</i>	<u>386</u>	139	78	6136164	6115301	94	593242	6127235	55.34	363880	176779	7
ArFEKO24	<i>K.ornithinolytica</i>		180	104	6100919	6070265	129	846848	6088270	54.99	272415	156271	6
ArFEKO25	<i>K.grimontii</i>	<u>377</u>	52	25	5875319	5865735	31	1296437	5870053	55.90	522597	407861	4
ArFEKO26	<i>K.michiganensis</i>	183	177	111	6480492	6453253	133	904803	6469547	55.19	271170	92845	8
ArFEKO27	<i>K.oxytoca</i>	37	131	64	6051166	6027663	80	562706	6038101	55.07	249384	164448	9

ArFEKO28	<i>K.oxytoca</i>	37	125	61	6050549	6026999	77	562706	6037422	55.07	231692	184297	9
ArFEKO29	<i>K.pasteurii</i>	<u>387</u>	183	106	6376422	6344813	131	512903	6363435	55.06	186214	110162	11
ArFEKO30	<i>K.grimontii</i>	<u>378</u>	169	115	6141014	6121760	130	345716	6132300	55.69	164646	93015	14
ArFEKO31	<i>K.michiganensis</i>	210	169	93	6455489	6429418	110	445191	6441149	55.21	209506	107972	10
ArFEKO32	<i>K.ornithinolytica</i>		175	105	6101112	6073282	128	846776	6089798	54.99	272415	153504	7
ArFEKO33	<i>K.michiganensis</i>	210	133	80	6288290	6265525	99	741781	6279616	55.30	219262	117017	7
ArFEKO34	<i>K.ornithinolytica</i>		117	76	5890814	5877530	85	650472	5883331	55.61	281669	177258	7
ArFEKO35	<i>K.michiganensis</i>	202	114	72	6115655	6103378	79	977352	6108204	55.94	182457	92718	9
ArFEKO36	<i>K.pasteurii</i>	<u>311</u>	93	53	6082934	6068915	63	923373	6076551	55.27	306791	216417	6
ArFEKO37	<i>K.ornithinolytica</i>		127	80	5856590	5842777	87	879833	5847686	55.60	316370	160092	6
ArFEKO38	<i>K.pasteurii</i>	<u>386</u>	124	79	6128280	6115921	85	593130	6120539	55.35	310074	125486	7
ArFEKO39	<i>K.michiganensis</i>	<u>380</u>	128	87	6374074	6359784	96	371804	6365871	55.47	177824	113663	13
ArFEKO40	<i>K.ornithinolytica</i>		87	59	5620056	5609031	68	704941	5616241	55.74	256407	118706	8
ArFEKP01	<i>K.ornithinolytica</i>		107	61	5804764	5789184	70	942102	5794976	55.58	353177	215389	6
ArFEKP03	<i>K.pneumoniae</i>	3254	70	43	5299532	5292202	45	1049275	5293581	57.42	385008	216052	5
ArFFKO01	<i>K.oxytoca</i>	36	87	53	5882107	5871591	58	560224	5875276	55.00	293185	186379	8
ArFFKO02	<i>K.pneumoniae</i>	305	71	38	5405024	5395835	43	1027747	5399021	57.27	369412	178308	5
ArFFKO03	<i>K.grimontii</i>	<u>379</u>	54	26	5684935	5675237	32	1198196	5678969	55.99	507116	354747	4
ArFFKO04	<i>K.oxytoca</i>	18	59	34	5731629	5724295	40	1141663	5727936	55.10	370615	193820	4
ArFFKO05	<i>K.oxytoca</i>	18	64	32	5742133	5733023	36	1236631	5735537	55.10	495855	201869	3
ArFFKO06	<i>K.oxytoca</i>	18	68	33	5730927	5720561	39	1031512	5724202	55.10	370615	201869	5
ArFFKO07	<i>K.oxytoca</i>	36	94	56	5885837	5874078	61	560452	5877763	55.00	264851	162476	8
ArFFKO08	<i>K.oxytoca</i>	18	73	35	5734036	5722487	42	1141891	5726698	55.11	370615	202722	5
ArFFKO09	<i>K.oxytoca</i>	18	62	37	5736564	5729858	41	1141891	5732258	55.10	370615	194178	4
ArFFKO10	<i>K.oxytoca</i>	18	61	29	5725366	5715600	36	1141891	5719853	55.11	370615	202748	4
ArFFKO11	<i>K.oxytoca</i>	18	74	33	5739726	5726545	40	1141954	5730890	55.10	370615	197218	4
ArFFKO12	<i>K.michiganensis</i>	<u>381</u>	87	54	6144425	6133536	61	532850	6138721	55.83	306745	163097	8
ArFFKP01	<i>K.pneumoniae</i>	3254	74	45	5299678	5290846	50	942354	5294199	57.42	378229	205987	5
ArFFKP02	<i>K.pneumoniae</i>	3254	67	42	5300180	5293423	44	1236462	5294802	57.42	378229	304639	5
ArFFKP03	<i>K.pneumoniae</i>	3254	71	45	5299381	5292185	47	549090	5293564	57.42	378229	196086	6
ArFFKP04	<i>K.pneumoniae</i>	3254	73	43	5299563	5290527	47	1049584	5293168	57.42	373006	205987	5
ArFFKP05	<i>K.planticola</i>		65	39	5798197	5788938	46	1041534	5793999	55.65	318047	233793	6
ArFFKP06	<i>K.pneumoniae</i>	3254	66	41	5299981	5293226	43	1129720	5294605	57.42	385008	304639	5
ArFFKP07	<i>K.pneumoniae</i>	4242	88	66	5507879	5502524	68	418300	5504241	57.18	171642	100464	11
ArFFKP08	<i>K.pneumoniae</i>	4242	71	43	5526297	5519361	45	582941	5521077	57.19	339249	174648	7
ArFFKP09	<i>K.pneumoniae</i>	4242	68	42	5526525	5520210	44	742829	5521926	57.19	339249	174648	7
ArFFKP10	<i>K.pneumoniae</i>	4242	65	41	5523804	5517514	44	582941	5519868	57.19	349911	174648	7
ArFFKP11	<i>K.planticola</i>		61	39	5798380	5790747	44	1041525	5794760	55.65	320571	207623	6
ArFFKP12	<i>K.pneumoniae</i>	4242	62	39	5519190	5513483	41	742654	5515200	57.20	349911	206669	6
ArFFKP13	<i>K.pneumoniae</i>	107	84	53	5485451	5473645	62	451602	5479832	57.14	346725	189846	7
ArFFKP14	<i>K.pneumoniae</i>	3254	72	46	5299154	5291116	50	487023	5294101	57.42	373006	187145	7
ArFFKP15	<i>K.variicola_sub</i>	3525	60	32	5554738	5546750	35	778411	5549113	57.45	474834	204769	5

<i>sp._variicola</i>													
ArFFKP16	<i>K.pneumoniae</i>	3254	73	44	5299517	5290736	48	942445	5293377	57.42	372914	186843	6
ArFFKP17	<i>K.planticola</i>		73	43	5797824	5787864	50	1041525	5793121	55.65	266365	178720	7
ArFFKP18	<i>K.pneumoniae</i>	4242	63	40	5518976	5513304	42	582941	5515021	57.20	339249	206669	7
ArFFKP19	<i>K.planticola</i>		56	38	5791003	5784986	43	672246	5788667	55.66	357711	217595	6
ArFFKP20	<i>K.planticola</i>		60	42	5789918	5783622	46	704203	5786798	55.65	357711	207563	6
ERR1228222	<i>K.variicola</i>	1250	39	35	5821486	5820812	35	1174277	5820812	57.05	385211	295599	5
ERR4013551	<i>K.planticola</i>		57	38	5440420	5435823	40	613727	5437440	55.91	338878	139287	6
SRR10579999	<i>K.michiganensis</i>		197	178	6044434	6037889	182	239660	6040893	55.71	82701	40788	22
SRR17799538	<i>K.pneumoniae</i>	1565	141	77	5625364	5596463	106	475475	5616727	57.16	277032	174033	8
SRR17858119	<i>K.pneumoniae</i>	785	111	54	5749023	5724892	71	1190727	5737792	56.97	348409	216782	6
ERR4013963	<i>K.ornithinolytica</i>		52	28	5299263	5291686	32	1037016	5294425	55.93	401494	288518	4
SRR15123351	<i>K.pneumoniae</i>	2502	107	66	5619662	5606350	75	735443	5612801	57.08	322829	176112	6
ERR1204816	<i>K.variicola</i>	1148	134	83	5864144	5844493	98	723108	5856204	56.99	296929	134240	7
SRR15123353	<i>K.pneumoniae</i>	2502	105	65	5597709	5586040	72	688401	5590575	57.10	293680	143727	7
SRR11585239	<i>K.oxytoca</i>	2	112	48	5955604	5935380	59	776058	5943313	55.06	246407	155243	8

ST-types were not assigned to *K. ornithinolytica* and *K. planticola* due to the lack of an appropriate MLST allelic scheme. Sixteen *Klebsiella* genomes were attributed to six new ST-Types never found before (Table 9, underlined STs). Interestingly, a higher species diversity was observed in Italian soft cheese rather than in Italian salami production. Most frequently identified foodborne *Klebsiella* were *K. michiganensis* ST380 and ST95, as well as *K.oxytoca* ST37 in Italian soft cheese production. In Italian salami production, *K. pneumoniae* ST3254 and ST4242 were identified along with *K. oxytoca* ST18. One isolate of *K. pneumoniae* ST3254 was found also in Italian soft cheese production.

Based on SNP calling, a phylogenetic tree of all *Klebsiella* genomes (foodborne and human) was inferred (Figure 8). As expected, genomes clustered according to the species, ST-Types (when available) and food matrix. The 74 genomes were gathered in 10 clades: clades 1 to 7 gathering genomes of *K. ornithinolytica*, *K. planticola*, *K. variicola*, *K. pneumoniae*, *K. oxytoca*, *K. pasteurii* and *K. grimontii* respectively. Clades 8 and 9 included genomes of *K. michiganensis*. One additional clade includes only one genome namely the public genome of human origin of *K. michiganensis*.

Based on the phylogenetic tree clustering and the SNP distance matrix, data suggest that different species and clones of *Klebsiella* were co-existing in Italian soft cheese and salami production (Figure 8). Only few genomes showed genetic similarities based on SNP differences. ArFFKP08, ArFFKP09 and ArFFKP10 belonged to *K. pneumoniae* ST4242 and showed from 68 to 81 SNPs difference suggesting their close genetic relationship. All three isolates were collected from meat mixture in the staffing room the same day of sampling (01/07/2020). More interestingly ArFFKO04, ArFFKO05, ArFFKO06, ArFFKO08, ArFFKO09, ArFFKO10, ArFFKO11 belonged to *K. oxytoca* ST18 and showed from 6 to 13 SNPs difference suggesting they belong to the same strain. All 7 isolates were collected from salami during ripening at drying room and maturation room in three subsequent sampling of three different batches (September, October and November 2020) suggesting the persistence of this strain over time. Although this strain was not found in environmental samples, further studies would be mandatory to identify the source of re-contamination which could have been the environment as well as the raw ingredients (meat mixture and swine).

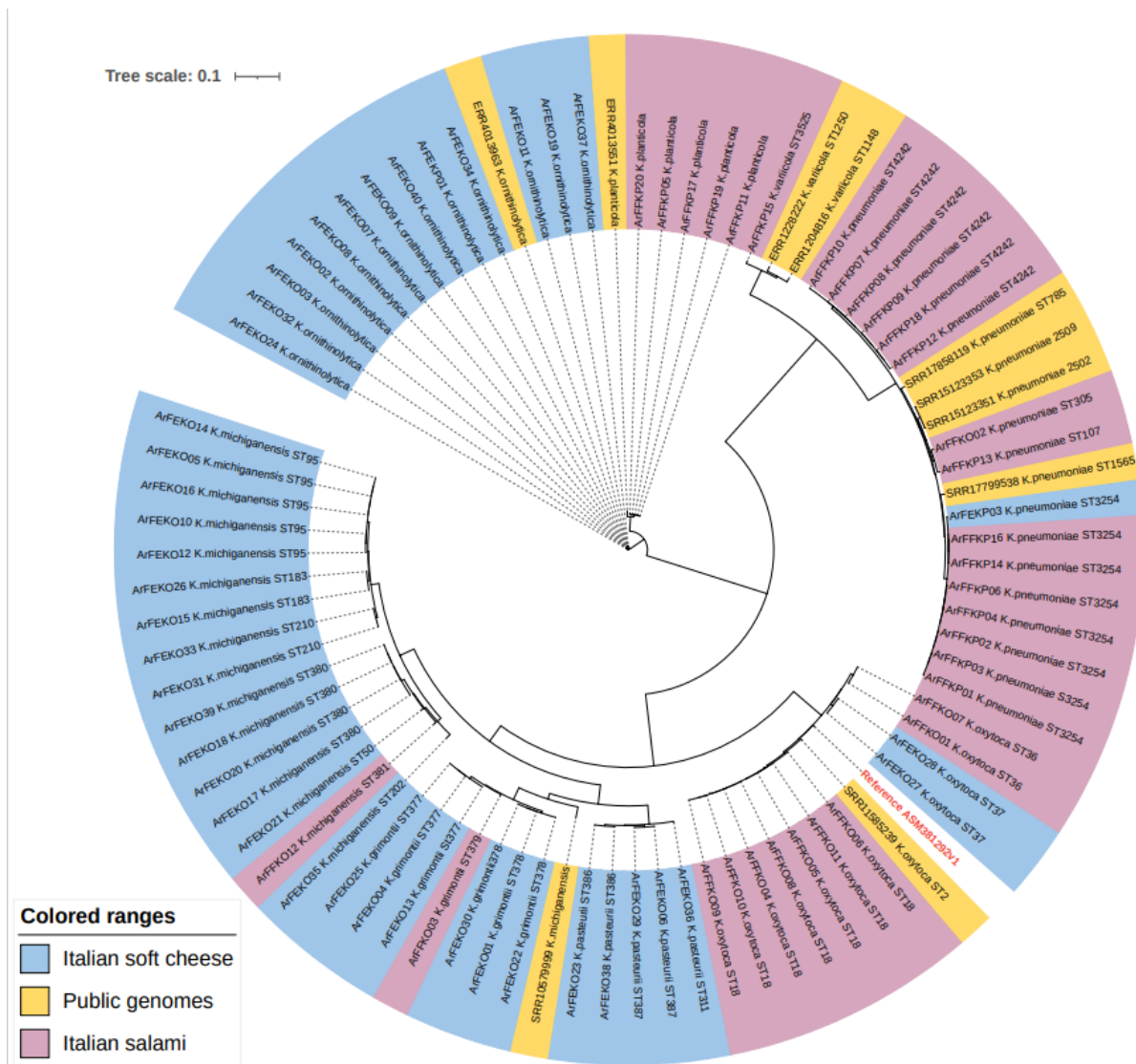


Figure 8: Phylogenetic tree of *Klebsiella* spp. genomes

Regarding antimicrobial resistance prediction in *Klebsiella* spp., isolates showed number of associated resistance genes from 4 to 7 in foodborne *Klebsiella* and from 4 to 15 in public human genomes. Genes predicted resistance in foodborne *Klebsiella* to beta-lactams (*blaOXA*, *blaSHV*, *blaPLA*), fosfomycin (*fosA*) and reduced susceptibility to fluoroquinolones (*oqxAB*). *Klebsiella* species with the highest number of AMR genes were *K. michiganensis* (5 genes) (ArFEKO17, ArFEKO18, ArFEKO20, ArFEKO21, ArFEKO31, ArFEKO33, ArFEKO35, ArFFKO12) and *K. pasteurii* (6-7 genes) (ArFEKO29, ArFEKO06) which both additionally showed aminoglycosides resistance genes (*aph*).

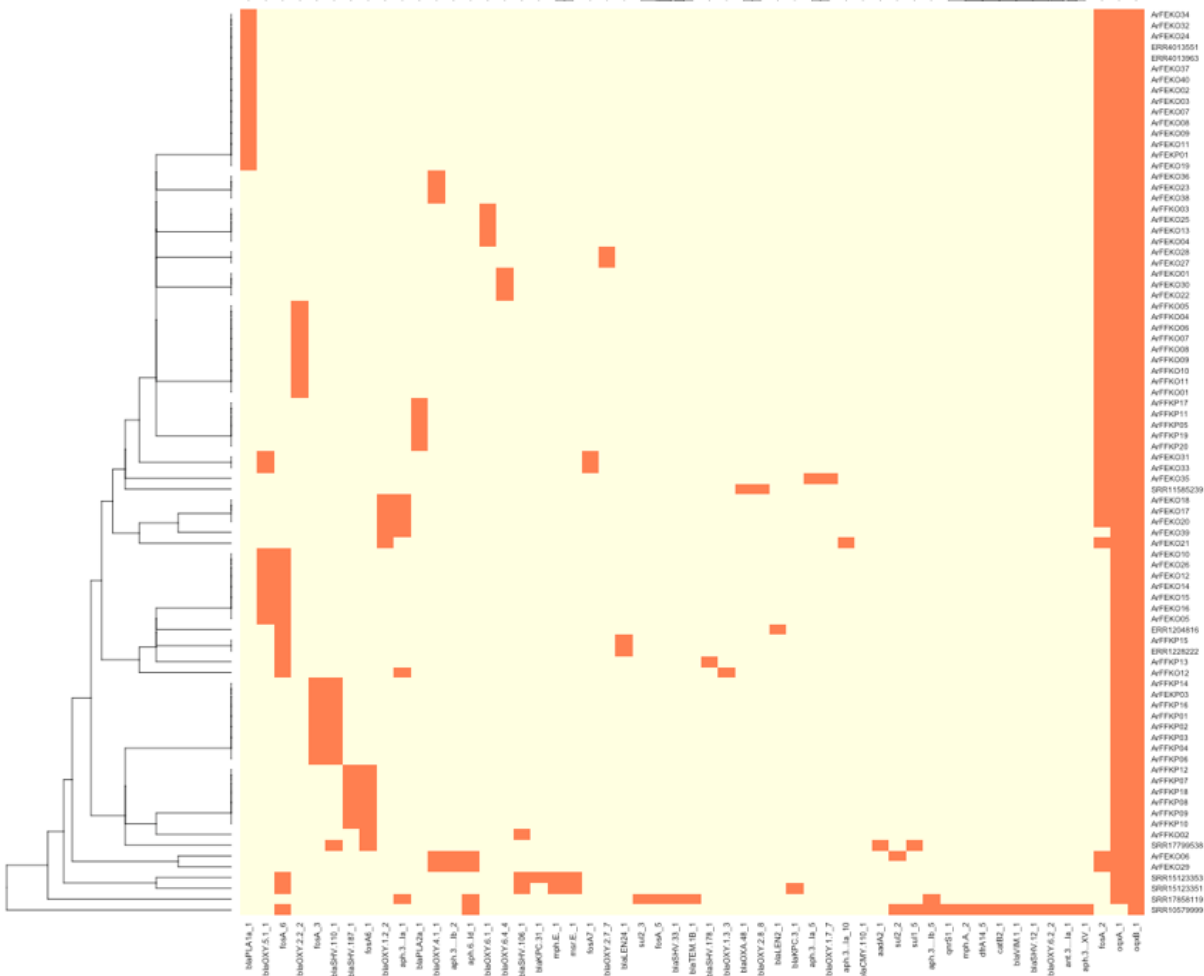


Figure 9: Heatmap of resistome of foodborne and human *Klebsiella* genomes (yellow: absence (<80% of sequence identity), orange: presence (>80% of sequence identity))

Regarding virulence prediction, all 74 *Klebsiella* genomes carried from 3 to 25 virulence genes (Figure 10). Those genomes with the highest number of virulence genes belonged to 7 *K. pneumoniae* (ArFFKP07, ArFFKP08, ArFFKP09, ArFFKP10, ArFFKP12, ArFFKP13, ArFFKP18) isolates collected from Italian salami. These genomes carried both the genes encoding for aerobactin (*iucABC* and *iutA*) and the genes encoding for yersiniabactin (*ybtAEPQSTUX*): two siderophores important for iron carriage which are crucial for bacterium survival within the host. In general, aerobactin genes were found in foodborne isolates in association or not with yersiniabactin genes. These genes were not found in human isolates which carried *irp1* and *irp2* genes related to an alternative iron transport system to siderophores. Additionally, these seven genomes carried *yagVWXYZ* (*ecpABCD*) cluster. This cluster of virulence genes was found in all *K. pneumoniae* species complex genomes (*K. pneumoniae* and *K. variicola*) irrespective of the origin (human or food). The *ecp* gene cluster is associated to pili formation and it is therefore essential in adherence and biofilm formation (Muraya et al., 2022).

Regarding KoSC and *Raoultella* species complex genomes, the most frequently identified virulence genes were the *ybt* genes encoding for yersiniabactin along with *futA*, *irp1* and *irp2*. In those genomes *ybt* and *irp* genes were found as alternatives to *iuc* and *ecp* genes.

