International Journal of Food Microbiology 291 (2019) 181–188

Contents lists available at ScienceDirect

International Journal of Food Microbiology

journal homepage: www.elsevier.com/locate/ijfoodmicro

Insights from genome-wide approaches to identify variants associated to phenotypes at pan-genome scale: Application to *L. monocytogenes*' ability to grow in cold conditions

Lena Fritsch[a], Arnaud Felten[a], Federica Palma[a], Jean-François Mariet[a], Nicolas Radomski[a], Michel-Yves Mistou[a], Jean-Christophe Augustin[a,b], Laurent Guillier[a,*]

[a] *French Agency for Food, Environmental and Occupational Health & Safety (Anses), Laboratory for Food Safety, Université Paris-Est, Maisons-Alfort F-94701, France*
[b] *Ecole Nationale Vétérinaire d'Alfort, Maisons-Alfort F-94704, France*

A R T I C L E I N F O

Keywords:
Whole genome sequencing
Genome wide association study
Cold stress

A B S T R A C T

Intraspecific variability of the behavior of most foodborne pathogens is well described and taken into account in Quantitative Microbial Risk Assessment (QMRA), but factors (strain origin, serotype, …) explaining these differences are scarce or contradictory between studies. Nowadays, Whole Genome Sequencing (WGS) offers new opportunities to explain intraspecific variability of food pathogens, based on various recently published bioinformatics tools.

The objective of this study is to get a better insight into different existing bioinformatics approaches to associate bacterial phenotype(s) and genotype(s). Therefore, a dataset of 51 *L. monocytogenes* strains, isolated from multiple sources (i.e. different food matrices and environments) and belonging to 17 clonal complexes (CC), were selected to represent large population diversity. Furthermore, the phenotypic variability of growth at low temperature was determined (i.e. qualitative phenotype), and the whole genomes of selected strains were sequenced. The almost exhaustive gene content, as well as the core genome SNPs based phylogenetic reconstruction, were derived from the whole sequenced genomes. A Bayesian inference method was applied to identify the branches on which the phenotype distribution evolves within sub-lineages. Two different Genome Wide Association Studies (i.e. gene- and SNP-based GWAS) were independently performed in order to link genetic mutations to the phenotype of interest.

The genomic analyses presented in this study were successfully applied on the selected dataset. The Bayesian phylogenetic approach emphasized an association with "slow" growth ability at 2 °C of the lineage I, as well as CC9 of the lineage II. Moreover, both gene- and SNP-GWAS approaches displayed significant statistical associations with the tested phenotype. A list of 114 significantly associated genes, including genes already known to be involved in the cold adaption mechanism of *L. monocytogenes* and genes associated to mobile genetic elements (MGE), resulted from the gene-GWAS. On the other hand, a group of 184 highly associated SNPs were highlighted by SNP-GWAS, including SNPs detected in genes which were already likely involved in cold adaption; hypothetical proteins; and intergenic regions where for example promotors and regulators can be located.

The successful application of combined bioinformatics approaches associating WGS-genotypes and specific phenotypes, could contribute to improve prediction of microbial behaviors in food. The implementation of this information in hazard identification and exposure assessment processes will open new possibilities to feed QMRA-models.

1. Introduction

In the field of food microbiology, phenotypic strain variability is a well-known phenomenon (Koutsoumanis and Lianou, 2013). While studying the behavior of bacterial pathogens in food, it is quite common to include from several to large collection of strains (Haberbeck et al., 2015; Van Der Veen et al., 2008). Concerning the main foodborne pathogens, some significant discrepancies have been enlightened while observing and modelling the strain behaviour. For instance, Aryani et al. (2015) studied the growth variability of 20 *Listeria monocytogenes*

---

[*] Corresponding author at: Salmonella and Listeria Unit, Laboratory for Food Safety, 14 rue Pierre et Marie Curie, F-94701 Maisons-Alfort, France.
E-mail address: laurent.guillier@anses.fr (L. Guillier).

https://doi.org/10.1016/j.ijfoodmicro.2018.11.028
Received 28 February 2018; Received in revised form 9 October 2018; Accepted 28 November 2018

strains in different conditions of pH, water activity (a_w), NaCl concentration, undissociated lactic acid concentration and temperature (Aryani et al., 2015). The variability was quantified through model parameters as minimal growth temperature, T_{min} (°C), with values ranging from −3.3 °C to −1.1 °C. With regard to *Bacillus cereus*, large differences for cardinal growth parameters, including minimal, optimal and maximal growth temperatures, were also described observing the parameter T_{opt} (°C) from 31.0 °C to 43.1 °C for the 12 studied strains (Carlin et al., 2013). Haberbeck et al. (2015) investigated the variation in growth/no growth boundaries of 188 *Escherichia coli* strains with a range of minimal pH for growth varying between 3.8 and 4.3 according to the tested strains (Haberbeck et al., 2015). The inclusion of several strains is also common when studying other phenotypes which are not related to predictive modelling. Numerous other examples can be found concerning the bacterial adhesion or biofilm formation (Lianou and Koutsoumanis, 2012; Piercey et al., 2017), ability to survive to a stress (Zoz et al., 2017) or more loosely defined phenotype, such as the ability to persist in food processing environment (Nowak et al., 2017).

The description of strain variability can by itself be sufficient to identify the appropriate control measure that should be applied to guarantee food safety. In such a situation, the strain with higher ability to survive/grow is the benchmark reference strain. However, characterization of phenotypic variability is usually followed by statistical association studies on qualitative data characterizing the strains and the observed phenotypic differences. According to the pathogen, several factors of interest may explain the observed variability. They usually rely on the isolation origin of the strain, on some sub-typing characteristics (i.e. serotype) or even on any relevant presence/absence of markers (Carlin et al., 2013; Shabala et al., 2008; Van Der Veen et al., 2008). For example, Carlin et al. (2013) demonstrated a clear correlation between different phylogenetic groups of *B. cereus* and the adaptation to temperature, pH and a_w. In contrast, other studies focusing on phenotypic strain variability failed to detect associations between factors as the strain serotype or origin (Shabala et al., 2008; Van Der Veen et al., 2008).

These examples demonstrate that there is a need of new analytical approaches in order to identify genetic markers predicting bacterial behaviors. Based on the expansion of WGS technologies, as well as the recent development of open source bioinformatics tools associating mutations and bacterial phenotypes, it is now possible to detect genetic elements involved in given binary or continuous phenotypes (Lees and Bentley, 2016; Power et al., 2017).

Almost all information about the phenotypic behaviors is encoded in the microbial DNA (Chen and Shapiro, 2015). Association studies between this genetic information and phenotypic traits of interest can be performed at different levels of genetic elements, such as the presence/absence of specific genes (Brynildsrud et al., 2016), k-mers (Earle et al., 2016; Lees et al., 2016; Marinier et al., 2017) or single nucleotide polymorphisms (SNPs) (Collins and Didelot, 2018; Earle et al., 2016).

Genome Wide Association Studies (GWAS) have been recently developed and applied on bacterial genomes, mainly focusing on clinically relevant phenotypes such as virulence (Buchanan et al., 2017; Kremer et al., 2017; Laabei et al., 2014; Pielaat et al., 2015) and antibiotic resistance (Earle et al., 2016; Salipante et al., 2015). Up to our knowledge, a limited number of GWAS (Hingston et al., 2017b; Kremer et al., 2017; Pielaat et al., 2015) assessed de novo associations of phenotypes relevant to industry and/or useful for risk assessment studies.

Even if bacterial GWAS aims to mitigate the prejudicial effects of population structure, methods exploring the relationship between genotypes and phenotypes may rely on phylogenetic reconstruction. The relationship between a genotype and a phenotype can indeed be investigated by observing how the phenotype is distributed according to genotypic values (Ansari and Didelot, 2016).

The successful application of GWAS and phylogeny-based methods has been extensively proven for phenotypes with clinical importance (i.e. antibiotic resistance or virulence). So far, the application of these methods for investigating phenotypes of interest for food microbiology is still limited. Therefore, the objective of this article is to present and to assess the ability of different approaches to identify genetic markers associated to a phenotype of interest in order to predict the bacterial behavior in foods. These methods were applied to a selected dataset of newly sequenced *L. monocytogenes* strains with well-characterized phenotypes at low temperature.

2. Materials and methods

2.1. Strains

From the Anses collection stored at −80 °C, a set of 51 *L. monocytogenes* strains was constituted in order to achieve large population diversity in terms of genotypes (i.e. CC-types) and food-related origins (i.e. meat products, dairy, seafood and environment). The lineages, clonal complexes (CC) based on 7 loci MLST (Ragon et al., 2008), as well as the strain origin are detailed in Supplementary material 1.

2.2. Whole genome sequencing

Firstly, cryo bead-based strains were inoculated onto tryptone soya agar (TSA, AES Chemunex, Bruz, France) and incubated for 24 h at 37 °C to check isolates purity. A single colony of each strain isolated on TSA was picked to inoculate different brain heart infusion broth (BHI, Oxoid, Dardilly, France) tubes (10 ml) incubated overnight (12–18 h) at 37 °C. The Wizard® Genomic DNA Purification Kit (Promega, France) was used in order to isolate genomic DNA. Nanodrop® Spectrophotometer and Qbit® fluorimeter were used to assess the quantity of the extracted DNA. Global integrity of DNA (200 ng) was assessed via horizontal agarose electrophoresis (Seakem GTG Agarose gel at 0.8% migration 3 h with an electrical field at 120 V/105 mA in TBE 1× gel) as a quality control measure. Paired-end sequencing (i.e. 2 × 150 bp) were performed by the 'Institut du Cerveau et de la Moëlle' (ICM, France) using TruSeq automated library preparation (Illumina) and Nextseq500 sequencing system (Illumina). The paired-end reads, isolates information and genome assemblies are available under the PRJEB24673 (ERP106517) ENA bio-project.

2.3. Phenotyping in cold conditions

Qualitative phenotypes based on the ability to reach turbidity in a given time (i.e. two months) at 2 °C were established for each strain and a minimum of three replicates were performed for all experiments.

One cryo bead of each strain was put in 10 ml TSB-YE and incubated for 7 h at 37 °C. Then, a 1/10 dilution was prepared and further incubated at 37 °C for 17 h. In order to achieve a density of 5 log cfu·ml^{-1}, overnight cultures were diluted in pre-chilled tryptone salt (TS, Oxoid, Hampshire, UK) and TSB-YE (2 °C), successively. Enumerations ensuring the concentration of the obtained suspensions were performed onto TSA. Microtiter plates were inoculated with 200 µl per well (Overney et al., 2016) of the diluted suspensions and incubated at 2 °C. The microplates were covered with parafilm in order to avoid dehydration. The turbidity of wells was visually inspected each day for each strain, so that the strains can then be divided into two groups based on time to observe turbidity. The two groups include strains that showed visible turbidity before (hereafter called "fast") and later (hereafter called "slow") than 21 days, respectively.

2.4. Genomic approaches

In order to link mutations to qualitative phenotypes, two different SNP-based phylogenetic reconstructions (i.e. including and excluding recombination events; § 2.4.1), two different GWAS methods (i.e. gene- and SNP-based GWAS; § 2.4.2), as well as a Bayesian inference method

(§ 2.4.3), were performed.

2.4.1. Phylogenetic reconstructions

2.4.1.1. SNPs calling. A variant calling analysis was performed using iVARCall2, a recently published workflow based on the GATK HaplotypeCaller algorithm (Felten et al., 2017). Briefly, this pipeline aims to align paired-end reads against a reference genome to identify SNPs and InDels by local de novo assembly. The complete sequence of *L. monocytogenes* strain EGD-e (accession number: NC_003210) was used as reference genome. The pipeline produces also a multi-fasta file, called "pseudogenomes", including pseudo reference genomes obtained by substituting into the reference all the genotypes of SNPs for each strain included in the analysis.

2.4.1.2. SNP-based phylogenetic inference. Phylogenetic reconstruction based on core genome SNPs was carried out with RAxML (Randomized Axelerated Maximum Likelihood) (Stamatakis, 2014). The pseudogenome fasta file generated by iVARCall2 was used as input for RAxML to build a phylogenetic tree in newick format. The phylogenetic inference was performed with bootstrap analysis and searching for the best-scoring Maximum Likelihood (ML) tree with General Time-Reversible (GTR) model of substitution and the secondary structure 16-state model. One thousand bootstraps were applied and convergence was checked. The resulting newick trees were further used in the following analyses.

In order to detect recombination events, ClonalFrameML software package (Didelot and Wilson, 2015) was applied using as input the phylogenetic tree from RaxML and the pseudogenome fasta file from iVARCall2. Then, the SNPs corresponding to recombination events were filtered out from the pseudogenomes using the Clonal_VCFilter script (Felten et al., 2017) in order to obtain a phylogenetic tree where variants linked to recombination events were excluded.

2.4.2. GWAS methods

2.4.2.1. Gene-based GWAS. A GWAS based on the accessory gene content of the 51 *L. monocytogenes* de novo assemblies was performed. First, draft genomes were obtained based on the SPAdes algorithm (Bankevich et al., 2012) after quality check of Illumina reads using FASTQC (Andrews, 2010) and trimming low quality reads with Trimmomatic (Bolger et al., 2014).

Finally, the qualitative evaluation of assemblies was performed by computing various metrics with QUAST (Gurevich et al., 2013). The whole genome annotation was carried out using Prokka (Seemann, 2014) with default parameters. Prokka uses the assemblies as input and produces GFF3-files, including sequences and annotations, which were used to extract the pan-genome of the 51 *L. monocytogenes* isolates with the software Roary (Page et al., 2015). Finally, gene-based GWAS was performed using Scoary (Brynildsrud et al., 2016). A file including phenotypic traits, the phylogenetic tree based on SNPs and the gene presence/absence matrix from Roary were thus used as input-data to Scoary.

2.4.2.2. SNP-based GWAS. A SNP-based GWAS was performed based on the workflow published by Earle et al., 2016 (Earle et al., 2016). The R scripts developed by Earle et al. and the related external programs, such as GEMMA (Zhou and Stephens, 2012), were applied to create SNPs-GWAS outputs. The pseudogenome fasta sequences produced by iVARCall2 (Felten et al., 2017) and the associated binary phenotypic trait were supplied as input of the workflow developed by Earle et al. (Earle et al., 2016). In the SNPs-GWAS, correction for multiple testing was accounted for by applying Bonferroni adjustments. The consideration of population structure effect provides estimates of the proportion of variance in phenotypes explained by "SNP heritability". For associated SNPs, the related information are provided including annotation, associated p-values, positions, types and bases of SNPs, as well as reference and non-reference codons with the corresponding amino acid.

2.4.3. Bayesian inference method

With a view to assessing the evolution of qualitative phenotypes distribution along the branches of the phylogenetic tree, a recently developed Bayesian inference method (Ansari and Didelot, 2016) was applied. The R package called TreeBreaker was used in order to infer the evolution of a discrete phenotype distribution across a phylogenetic tree and divide the tree into segments where their distributions are constant (Ansari and Didelot, 2016). Therefore, SNP-based newick trees obtained at core genome and gene (*inlA*) levels excluding recombination events and the discrete phenotypic traits ("fast" = 0, "slow" = 1) were supplied as input to this program.

3. Results and discussion

Out of the 51 *L. monocytogenes* strains included in the study, 43 were classified as "fast" growing strains and 8 as "slow" growing strains at 2 °C (Supplementary material 1). All results presented below are supported by high-quality reads-mapping and draft genomes based on depth (median of 328 ×) and breadth (median of 87.39%) coverages, as well as total contigs length (i.e. cumulated length of 2,984,694 kb with N50 of 2,950,739), respectively (Supplementary material 2).

3.1. GWAS methods to explore the genetic basis of growth capacity at 2 °C

3.1.1. GWAS based on presence-absence of genes

A gene-based GWAS was carried out exploring the different analytical steps and corresponding statistical parameters computed by Roary (Page et al., 2015) and Scoary (Brynildsrud et al., 2016). The pangenome of the 51 *Listeria monocytogenes* strains (Fig. 1) is composed of a total of 6612 genes, including 2014 core genes (present in the 99% of isolates) and 4598 accessory genes (Supplementary material 3). A similar number of core genes has been already reported in previous studies on *L. monocytogenes*: 2032 (den Bakker et al., 2010) and 2354 (Kuenne et al., 2013). In order to select a curated set of genes associated to the "fast" or "slow" growth phenotype, the 4598 accessory genes detected by Roary were therefore considered for the GWAS analysis performed with Scoary. A detailed description of the different steps applied in the GWAS analysis is reported in Supplementary material 3. It is worth mentioning that in the results obtained with this analysis, the population structure correction is taken into account based on the produced SNP-based ML-tree (Supplementary material 4).

A list of a total of 114 genes with strong statistical associations with the growth ability at 2 °C (Supplementary material 4) was finally listed. The output table includes a couple of genes already reported to be involved in the cold adaption mechanism of *L. monocytogenes* such as the genes coding for RNA helicase (Markkula et al., 2012) and for the precursor of internalin A (Kovacevic et al., 2013). Interestingly, 15 out of 114 genes (~13%) of reported genes, corresponding to mobile genetic elements (such as phage capsid family protein or transposase from transposon Tn916), were found to be associated with the measured phenotype. Despite, 70 out of 114 genes (~61%) reported in the output table correspond to hypothetical proteins.

3.1.2. SNP-based GWAS

The objective of the SNP-based GWAS analysis was to detect the phenotype-associated genetic variants in the core genome. When performing microbial GWAS, confounding factors (i.e. bacterial population structure and recombination events) may affect the data analysis and lead to spurious associations. The approach proposed by Earle et al. (2016) was applied including mixed models which take into account for phylogenetic relatedness and signals of lineage associations. Therefore, all SNPs located in the 2014 core genome genes and in the intergenic regions which can also be involved in phenotype associations, especially the 5′-untranscribed regions which may be regulated by small

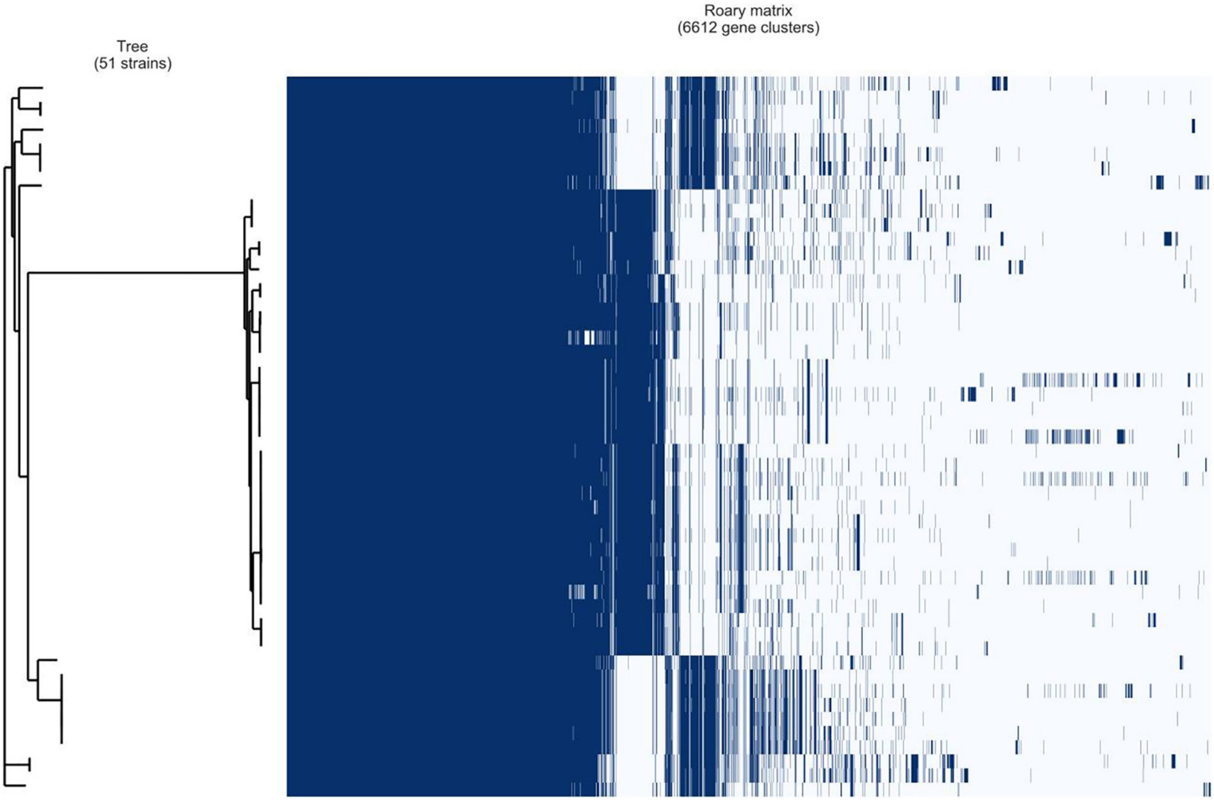

Fig. 1. Phylogenetic tree of 51 *L. monocytogenes* isolates compared to a matrix of presence and absence of core and accessory genes in dark blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

RNAs (Cerutti et al., 2017), were included in the analysis (i.e. 104,160 core genome SNPs). Fig. 2 presents the association for all these SNPs to the studied phenotype. In this Figure, the positions of the SNPs in the genome are on the x-axis while the corresponding $-\log_{10}$ (p-values) quantifying the association are displayed on the y-axis. The majority of the SNPs (n = 101,685) shows no significant association (p-value $>$ 0.05), whereas slight (p-value $<$ 0.05) and strong significance associations (p-value $<$ 0.01) were observed for 2291 and 184 SNPs, respectively. Table 1 shows the first 40 SNPs with the highest p-values of association and the complete table of SNPs with p-value $<$ 0.01 is reported in Supplementary material 5. This group of highly associated SNPs are harbored by genes which were previously described as likely involved in cold adaption such as lmo0135 ABC-transporter (Cabrita et al., 2013), lmo033 internalin (Hingston et al., 2017b; Kovacevic et al., 2013) and lmo0316 hydroxyethylthiazole (Mattila et al., 2012). The other highly associated SNPs are harbored by hypothetical proteins and in intergenic regions, where for example promotors and regulators can be located (Thorpe et al., 2017a).

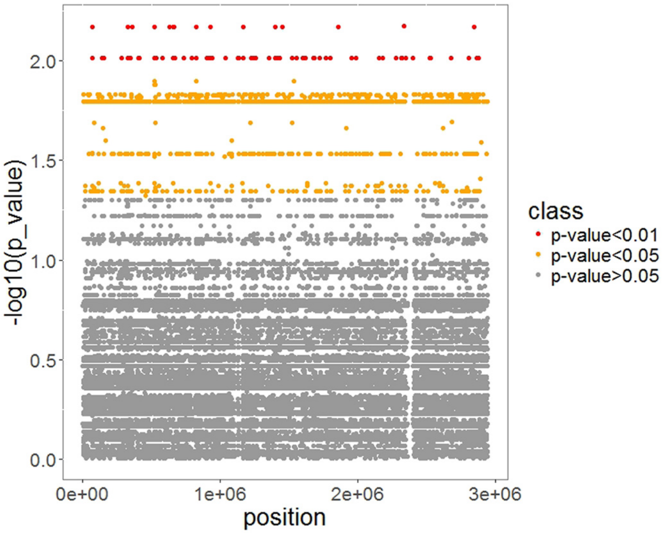

Fig. 2. Manhattan plot of core genome SNPs which show associations to the given phenotypic traits of the 51 *L. monocytogenes* strains. Each point presents a SNP with its position and corresponding –log10(p-value), which shows the significance of the geno-phenotype association. The color represents the extent of this significance, red (p-value $<$ 0.01), orange (p-value $<$ 0.05) and grey (p-value $>$ 0.05). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Although SNP-based GWAS has been recently applied in bacterial genomics, it is worth to notice that the SNPs associated to cold growth phenotype were supported by lower p-values in comparison to those reported by already published GWAS-based studies investigating antibiotic resistance (Chen and Shapiro, 2015; Earle et al., 2016). At least two reasons can be advanced to explain this discrepancy. The first is related to the limited number of genomes related to "slow" growing strains. A more balanced dataset might provide lower p-values supporting stronger phenotypic association. The other reason is related to the selected phenotype of interest. Contrary to other studies, the analyzed phenotype in this study is related to multiple compartments of the bacterial cell (Chan and Wiedmann, 2008). As previously reported, cold adaptation of *L. monocytogenes* may involve a variety of mechanisms, like the uptake of compatible solutes and oligopeptides (Chan and Wiedmann, 2008), the membrane fluidity (Saunders et al., 2016) and/or the general stress response factors (Cacace et al., 2010; Chan et al., 2008).

Other methods to perform bacterial GWAS were published these last years. Some of them rely on kmer presence/absence (Earle et al., 2016; Lees et al., 2016; Marinier et al., 2017), these approaches aim to associate traits with substitutions in the core genome, and the accessory

Table 1

Extract of SNPs with strong evidences (p-value < 0.01) for genome-wide association to the tested phenotypic trait of *L. monocytogenes* at 2 °C. All associated SNPs at 0.05 significance p-value levels are listed in Supplementary material 5.

Refposition	p-Value	Allele0	Allele1	A	C	G	T	Type	Refcodon	Non-refcodon	Refaa	Non-refaa	Name	Start	End	Product
2335431	0.0067	A	T	42	0	0	9	Synonymous	CCA	CCT	Pro	Pro	lmo2244	2334586	2335455	Ribosomal large subunit pseudouridine synthase
67812	0.0068	G	A	9	0	42	0	Synonymous	GTG	GTA	Val	Val	lmo0061	66160	70656	Hypothetical protein
326545	0.0068	A	G	42	0	9	0	Synonymous	AGT	AGC	Ser	Ser	lmo0300	325729	327120	Phospho-beta-galactosidase
361961	0.0068	G	A	9	0	42	0	Synonymous	GGG	GGA	Gly	Gly	lmo0333	360936	366272	Internalin
361964	0.0068	T	A	9	0	0	42	Synonymous	CTT	CTA	Leu	Leu	lmo0333	360936	366272	Internalin
361966	0.0068	C	T	0	42	0	9	Non-synonymous	TCG	TTG	Ser	Leu	lmo0333	360936	366272	Internalin
522745	0.0068	G	A	9	0	42	0	Synonymous	AAC	AAT	Asn	Asn	lmo0488	522628	523521	LysR family transcriptional regulator
632556	0.0068	T	C	0	9	0	42	Synonymous	GAT	GAC	Asp	Asp	lmo0590	631045	632811	Hypothetical protein
659121	0.0068	T	C	0	9	0	42	Synonymous	AGT	AGC	Ser	Ser	lmo0620	659083	659475	Hypothetical protein
659133	0.0068	C	T	0	42	0	9	Synonymous	GCC	GCT	Ala	Ala	lmo0620	659083	659475	Hypothetical protein
659139	0.0068	T	C	0	9	0	42	Synonymous	GAT	GAC	Asp	Asp	lmo0620	659083	659475	Hypothetical protein
659158	0.0068	A	G	42	0	9	0	Non-synonymous	AGT	GGT	Ser	Gly	lmo0620	659083	659475	Hypothetical protein
821163	0.0068	T	G	0	0	9	42	Synonymous	CTA	CTC	Leu	Leu	lmo0793	821070	821771	Hypothetical protein
924279	0.0068	A	T	42	0	0	9	Synonymous	TCA	TCT	Ser	Ser	lmo0884	923161	924540	Protoporphyrinogen oxidase
924290	0.0068	G	C	0	9	42	0	Non-synonymous	AGA	ACA	Arg	Thr	lmo0884	923161	924540	Protoporphyrinogen oxidase
924366	0.0068	T	A	9	0	0	42	Synonymous	CCT	CCA	Pro	Pro	lmo0884	923161	924540	Protoporphyrinogen oxidase
924378	0.0068	T	C	0	9	0	42	Synonymous	GTT	GTC	Val	Val	lmo0884	923161	924540	Protoporphyrinogen oxidase
924387	0.0068	G	A	9	0	42	0	Synonymous	CAG	CAA	Gln	Gln	lmo0884	923161	924540	Protoporphyrinogen oxidase
924393	0.0068	A	G	42	0	9	0	Synonymous	CGA	CGG	Arg	Arg	lmo0884	923161	924540	Protoporphyrinogen Oxidase
924415	0.0068	A	G	42	0	9	0	Non-synonymous	ATC	GTC	Ile	Val	lmo0884	923161	924540	Protoporphyrinogen oxidase
924420	0.0068	A	G	42	0	9	0	Synonymous	AAA	AAG	Lys	Lys	lmo0884	923161	924540	Protoporphyrinogen oxidase
924438	0.0068	T	C	0	9	0	42	Synonymous	GTT	GTC	Val	Val	lmo0884	923161	924540	Protoporphyrinogen oxidase
924456	0.0068	C	T	0	42	0	9	Synonymous	AGC	AGT	Ser	Ser	lmo0884	923161	924540	Protoporphyrinogen Oxidase
1400214	0.0068	A	T	42	0	0	9	Synonymous	GCA	GCT	Ala	Ala	lmo1375	1399699	1400796	Aminotripeptidase
1448590	0.0068	C	T	0	42	0	9	Non-synonymous	GCT	ACT	Ala	Thr	lmo1418	1448315	1449583	Hypothetical protein
1855515	0.0068	G	A	9	0	42	0	Synonymous	CCG	CCA	Pro	Pro	lmo1778	1854748	1855521	ABC transporter ATP-binding protein
2844257	0.0068	A	G	42	0	9	0	Non-synonymous	ATA	ATG	Ile	Met	lmo2763	2844249	2845601	PTS cellbiose transporter subunit IIC
1163023	0.0068	T	A	9	0	0	42	Synonymous	CTA	CTT	Leu	Leu	lmo1129	1162666	1163280	Hypothetical protein
1163026	0.0068	C	T	0	42	0	9	Synonymous	AAG	AAA	Lys	Lys	lmo1129	1162666	1163280	Hypothetical protein
1163043	0.0068	T	C	0	9	0	42	Non-synonymous	AGT	GGT	Ser	Gly	lmo1129	1162666	1163280	Hypothetical protein
1163044	0.0068	C	T	0	42	0	9	Synonymous	AAG	AAA	Lys	Lys	lmo1129	1162666	1163280	Hypothetical protein
1163055	0.0068	C	A	9	42	0	0	Non-synonymous	GTA	TTA	Val	Leu	lmo1129	1162666	1163280	Hypothetical protein
1163059	0.0068	C	T	0	42	0	9	Synonymous	TTG	TTA	Leu	Leu	lmo1129	1162666	1163280	Hypothetical protein
66276	0.0097	T	C	0	3	0	48	Synonymous	CCT	CCC	Pro	Pro	lmo0061	66160	70656	Hypothetical protein
137562	0.0097	G	T	0	0	48	3	Synonymous	GGG	GGT	Gly	Gly	lmo0135	137323	138897	Peptide ABC transporter substrate-binding protein
139157	0.0097	A	G	48	0	3	0	Synonymous	CAA	CAG	Gln	Gln	lmo0136	138999	139949	Peptide ABC transporter permease
139178	0.0097	A	G	48	0	3	0	Synonymous	CCA	CCG	Pro	Pro	lmo0136	138999	139949	Peptide ABC transporter permease
139247	0.0097	T	C	0	3	0	48	Synonymous	AAT	AAC	Asn	Asn	lmo0136	138999	139949	Peptide ABC transporter permease
150185	0.0097	T	C	0	3	0	48	Intergenic	–	–			lmo0152:lmo0153	148354:150232	150009:151173	Peptide ABC transporter substrate-binding protein: zinc ABC transporter substrate-binding protein

genome (Earle et al., 2016). Another tool was recently proposed to specifically analyses presence/absence of intergenic regions (Thorpe et al., 2017b). Herein, SNPs in genes, intergenic regions, as well as presence/absence of genes can be explored. A novel SNP-based GWAS method was also recently proposed for continuous phenotype (Collins and Didelot, 2018). An additional kmer-based GWAS approach dealing with phenotype-associated genetic variants (i.e. SNPs, InDels, intergenic regions and recombination events) without prior knowledge is still under evaluation (Jaillard et al., 2018).

Notwithstanding, a promising alternative to GWAS which rely on the combination of high-throughput gene function assays with mechanistic based models prioritizing genetic variants (i.e. synonymous variants, variants in non-coding regions and accessory genome) has been recently refined (Galardini et al., 2017). Such method has been successfully applied on a large collection of *E. coli* strains with known genotype and growth phenotype to deliver growth predictions, considering the variant's probability of affecting gene functionality, and genetic intervention strategies (Galardini et al., 2017).

3.2. Analysis of the growth ability at 2 °C through a Bayesian inference method

The distribution of the observed qualitative phenotype was evaluated with phylogenetic trees based on core genome (Fig. 3) and *inlA* gene (Fig. 4). In general, if all individuals included in a phylogenetic analysis have the same phenotype distribution, measured phenotypes would be randomly distributed on the tree leaves. Treebreaker can assess if the phenotype distribution evolves over time. The model is based on Poisson process of evolving phenotype distribution on the branches of a phylogeny. It allows checking if that distribution is different from ancestral state reconstruction. Hence, this Bayesian inference method aims at identifying the steadiness distribution of phenotype within a tree segment. When a segment has a significant distribution of a specific phenotype compared to other segments, the corresponding branch is flagged. The thickness of the branches is proportional to the posterior probability of phenotype change on the given branch. To perform this analysis a phylogenetic tree was built based on core genome SNPs detected on the 51 *L. monocytogenes* strains excluding recombination events because high rates of recombination events can affect reconstruction of *L. monocytogenes* lineages (Moura et al., 2017). Recombination events are distinct from evolutionary mechanisms related single substitutions and can consequently bias ancestral sequence reconstruction from where phylogenetic reconstructions are inferred with regards to evolutionary models of substitutions (Yang and Rannala, 2012). The variants from recombination events were thus removed in order to improve the accuracy of phylogeny reconstruction. The two Bayesian inference branches presenting the highest associations with "slow" growth ability at 2 °C correspond to the lineage I and CC9 of the lineage II (Fig. 3). These findings are in accordance with the study of Hingston et al. (2017a), where the CC1, CC2 and CC6 belonging to lineage I as also CC9 (lineage II) presented the slowest growth rates at 4 °C in broth (Hingston et al., 2017b).

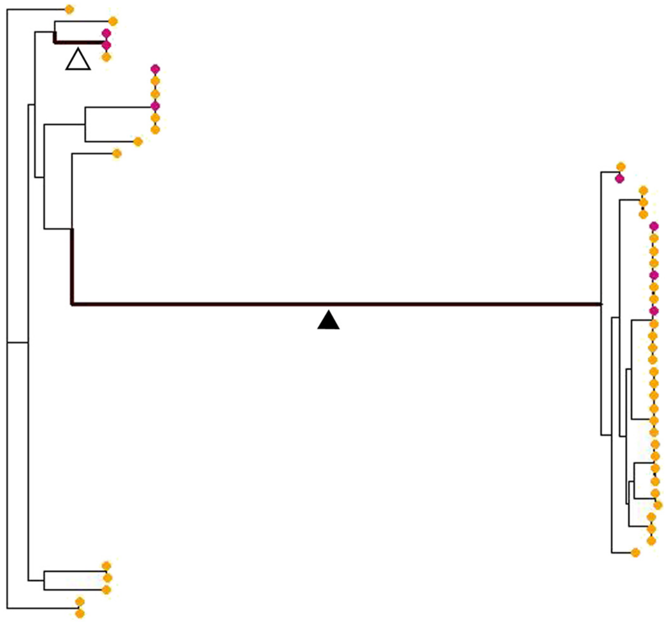

Fig. 3. Evolution of a discrete phenotype distribution (red dot = "slow", orange dot = "fast") on a phylogenetic tree including 51 analyzed genomes. The thickness of the branches is proportional to the posterior probability of having a change point. The branch "▲" corresponds to lineage I and the branch "Δ" to the CC9. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

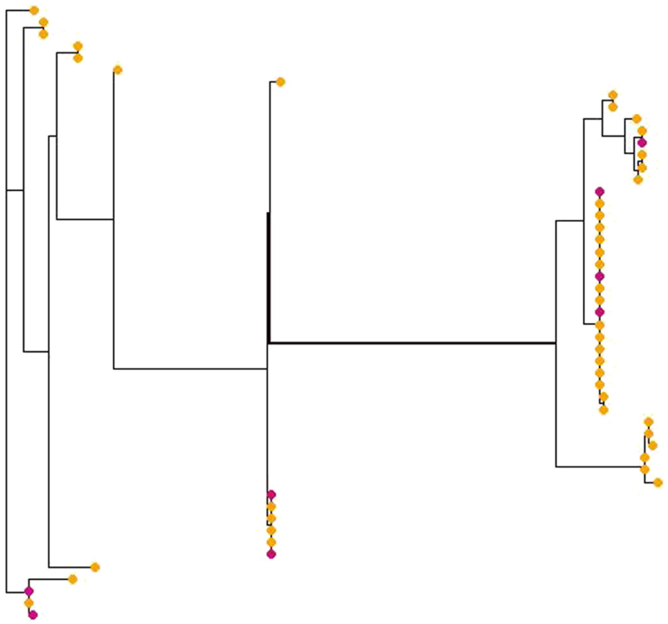

Fig. 4. Evolution of a discrete phenotype distribution (red dot = "slow", orange dot = "fast") on a phylogenetic tree based on DNA sequences of *inlA* gene. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

A similar approach to that herein presented, has been already applied to find association between antimicrobial resistance along branches of a phylogenetic tree (Suzuki et al., 2016). Nevertheless, in a study on *Streptococcus pyogenes*, it has been shown that phylogenetic inferences based on the alignment of specific SNP loci associated to specific phenotype may exhibit topological deviation compared to core genome-based phylogeny (Bao et al., 2016). Moreover, the categorical inferences established at the accessory gene level (i.e. wgMLST) can also exhibit topology distortion in comparison to core genome phylogenetic inferences based on SNPs (Henri et al., 2017). Accordingly, focusing the analysis on specific genes could be more relevant for investigating the evolution of the phenotypic trait of interest. For example, Møretrø et al. (2017) succeeded in finding relevant association between Benzalkonium Chloride (BC) susceptibility differences and a specific variant (i.e. cysteine/serine difference at amino acid 42) detected in *qacH* gene in *L. monocytognes* strains carrying or not the Quaternary Ammonium Compound (QAC) resistance genes (i.e. *bcrABC* or *qacH*) (Møretrø et al., 2017).

Therefore, a candidate gene approach focusing on a pre-specified gene of interest (i.e. *inlA*) already associated to cold adaptation (Hingston et al., 2017b; Kovacevic et al., 2013) was applied. A phylogenetic tree (Fig. 4) was thus established based on the sequences alignment of *inlA* gene and used as input for Treebreaker analysis. In Fig. 4, a pheno-genotype association (thicker branch) is shown, although the null-hypothesis model (i.e. no association of binary phenotype with the branch) is not rejected at 0.05 p-value level. Even though, the thicker branch reported in Fig. 4 is the same than that obtained based on core genome SNPs excluding recombination events and corresponding to the strains of lineage I.

4. Conclusion

A number of genes and SNPs, as well as specific phylogenetic sublineages were identified as associated to *L. monocytogenes'* growth at low temperature (2 °C). The three approaches of phenotype/genotype association were applied in the present study on a dataset of 51 strains. The application on a larger genome data set of accurately selected strains with contrasting phenotypes would likely increase the power to detect the associations.

The development of various innovative methods, as well as the increasing availability of bacterial genomes, opens a new area for food microbiology studies. The identification of genetic markers associated to phenotype of interest along the food chain should provide sensitive and accurate methods to predict and then control the behavior of foodborne pathogens (Cocolin et al., 2018b; den Besten et al., 2017). The use of GWAS could help to perform risk assessments especially targeting the subpopulations that pose the greatest risk linked to their higher ability to survive and/or grow in the food chain (Franz et al., 2016).

A rapid detection of strains harboring genetic markers related to the high ability to persist in the processing environment or to high virulence potential could guide reinforced control measures and optimize interventions strategies (Fritsch et al., 2018). The identification of such biomarkers at the genomic level is only one aspect. The expansion of alternative omics approaches based on transcriptomic or proteomics is opening new relevant perspectives for a full understanding of pathogen behavior in foods as well as in the whole food chain (Cocolin et al., 2018a; Hingston et al., 2017a; Renier et al., 2012).

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijfoodmicro.2018.11.028.

Acknowledgements

This work received funds from The French National Research Agency under the project ANR-15-CE21-0011 OPTICOLD.

Author contributions

LF and JFM performed the DNA extraction; AF and NR developed the bioinformatics workflows for genome assembly and variant calling; NR and FP contributed to manuscript writing and reviewing; MYM and JCA contributed to manuscript reviewing; LF performed strains phenotyping and genotyping; LF and LG drafted the manuscript; performed bioinformatics analyses, and designed the experimental study.

References

Andrews, S., 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data.

Ansari, M.A., Didelot, X., 2016. Bayesian inference of the evolution of a phenotype distribution on a phylogenetic tree. Genetics 204, 89–98.

Aryani, D., Den Besten, H., Hazeleger, W., Zwietering, M., 2015. Quantifying strain variability in modeling growth of *Listeria monocytogenes*. Int. J. Food Microbiol. 208, 19–29.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477.

Bao, Y.-J., Shapiro, B.J., Lee, S.W., Ploplis, V.A., Castellino, F.J., 2016. Phenotypic differentiation of Streptococcus pyogenes populations is induced by recombination-driven gene-specific sweeps. Sci. Rep. 6.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Brynildsrud, O., Bohlin, J., Scheffer, L., Eldholm, V., 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol. 17, 238.

Buchanan, C.J., Webb, A.L., Mutschall, S.K., Kruczkiewicz, P., Barker, D.O., Hetman, B.M., Gannon, V.P., Abbott, D.W., Thomas, J.E., Inglis, G.D., 2017. A genome-wide association study to identify diagnostic markers for human pathogenic *Campylobacter jejuni* strains. Front. Microbiol. 8.

Cabrita, P., Batista, S., Machado, H., Moes, S., Jenö, P., Manadas, B., Trigo, M.J., Monteiro, S., Ferreira, R.B., Brito, L., 2013. Comparative analysis of the exoproteomes of *Listeria monocytogenes* strains grown at low temperatures. Foodborne Pathog. Dis. 10, 428–434.

Cacace, G., Mazzeo, M.F., Sorrentino, A., Spada, V., Malorni, A., Siciliano, R.A., 2010. Proteomics for the elucidation of cold adaptation mechanisms in *Listeria monocytogenes*. J. Proteome 73, 2021–2030.

Carlin, F., Albagnac, C., Rida, A., Guinebretière, M.-H., Couvert, O., Nguyen-the, C., 2013. Variation of cardinal growth parameters and growth limits according to phylogenetic affiliation in the *Bacillus cereus* group. Consequences for risk assessment. Food Microbiol. 33, 69–76.

Cerutti, F., Mallet, L., Painset, A., Hoede, C., Moisan, A., Bécavin, C., Duval, M., Dussurget, O., Cossart, P., Gaspin, C., 2017. Unraveling the evolution and coevolution of small regulatory RNAs and coding genes in Listeria. BMC Genomics 18, 882.

Chan, Y.C., Wiedmann, M., 2008. Physiology and genetics of *Listeria monocytogenes* survival and growth at cold temperatures. Crit. Rev. Food Sci. Nutr. 49, 237–253.

Chan, Y.C., Hu, Y., Chaturongakul, S., Files, K.D., Bowen, B.M., Boor, K.J., Wiedmann, M., 2008. Contributions of two-component regulatory systems, alternative σ factors, and negative regulators to *Listeria monocytogenes* cold adaptation and cold growth. J. Food Prot. 71, 420–425.

Chen, P.E., Shapiro, B.J., 2015. The advent of genome-wide association studies for bacteria. Curr. Opin. Microbiol. 25, 17–24.

Cocolin, L., Mataragas, M., Bourdichon, F., Doulgeraki, A., Pilet, M.-F., Jagadeesan, B., Rantsiou, K., Phister, T., 2018a. Next generation microbiological risk assessment meta-omics: The next need for integration. Int. J. Food Microbiol. 287, 10–17.

Cocolin, L., Membré, J.-M., Zwietering, M., 2018b. Integration of omics into MRA. Int. J. Food Microbiol. 287, 1–2.

Collins, C., Didelot, X., 2018. A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. PLoS Comput. Biol. 14, e1005958.

den Bakker, H.C., Cummings, C.A., Ferreira, V., Vatta, P., Orsi, R.H., Degoricija, L., Barker, M., Petrauskene, O., Furtado, M.R., Wiedmann, M., 2010. Comparative genomics of the bacterial genus Listeria: genome evolution is characterized by limited gene acquisition and limited gene loss. BMC Genomics 11, 688.

den Besten, H.M., Amézquita, A., Bover-Cid, S., Dagnas, S., Ellouze, M., Guillou, S., Nychas, G., O'Mahony, C., Pérez-Rodriguez, F., Membré, J.-M., 2017. Next generation of microbiological risk assessment: potential of omics data for exposure assessment. Int. J. Food Microbiol. 287, 18–27.

Didelot, X., Wilson, D.J., 2015. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput. Biol. 11, e1004041.

Earle, S.G., Wu, C.-H., Charlesworth, J., Stoesser, N., Gordon, N.C., Walker, T.M., Spencer, C.C.A., Iqbal, Z., Clifton, D.A., Hopkins, K.L., Woodford, N., Smith, E.G., Ismail, N., Llewelyn, M.J., Peto, T.E., Crook, D.W., McVean, G., Walker, A.S., Wilson, D.J., 2016. Identifying lineage effects when controlling for population structure improves power in bacterial association studies. Nat. Microbiol. 1, 16041.

Felten, A., Nova, M.V., Durimel, K., Guillier, L., Mistou, M.-Y., Radomski, N., 2017. First gene-ontology enrichment analysis based on bacterial coregenome variants: insights into adaptations of Salmonella serovars to mammalian-and avian-hosts. BMC Microbiol. 17, 222.

Franz, E., Gras, L.M., Dallman, T., 2016. Significance of whole genome sequencing for surveillance, source attribution and microbial risk assessment of foodborne pathogens. Curr. Opin. Food Sci. 8, 74–79.

Fritsch, L., Guillier, L., Augustin, J.-C., 2018. Next generation quantitative microbiological risk assessment: Refinement of the cold smoked salmon-related listeriosis risk model by integrating genomic data. Microb. Risk Anal. 10, 20–27.

Galardini, M., Koumoutsi, A., Herrera-Dominguez, L., Varela, J.A.C., Telzerow, A., Wagih, O., Wartel, M., Clermont, O., Denamur, E., Typas, A., 2017. Phenotype inference in an *Escherichia coli* strain panel. elife 6, e31035.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075.

Haberbeck, L., Oliveira, R., Vivijs, B., Wenseleers, T., Aertsen, A., Michiels, C., Geeraerd, A., 2015. Variability in growth/no growth boundaries of 188 different *Escherichia coli* strains reveals that approximately 75% have a higher growth probability under low pH conditions than *E. coli* O157: H7 strain ATCC 43888. Food Microbiol. 45, 222–230.

Henri, C., Leekitcharoenphon, P., Carleton, H.A., Radomski, N., Kaas, R.S., Mariet, J.-F., Felten, A., Aarestrup, F.M., Gerner Smidt, P., Roussel, S., Guillier, L., Mistou, M.-Y., Hendriksen, R.S., 2017. An assessment of different genomic approaches for inferring phylogeny of *Listeria monocytogenes*. Front. Microbiol. 8.

Hingston, P., Chen, J., Allen, K., Truelstrup, H.L., Wang, S., 2017a. Strand specific RNA-sequencing and membrane lipid profiling reveals growth phase-dependent cold stress response mechanisms in *Listeria monocytogenes*. PLoS One 12, e0180123.

Hingston, P., Chen, J., Dhillon, B.K., Laing, C., Bertelli, C., Gannon, V., Tasara, T., Allen, K., Brinkman, F.S., Hansen, L.T., 2017b. Genotypes associated with *Listeria monocytogenes* isolates displaying impaired or enhanced tolerances to cold, salt, acid, or desiccation stress. Front. Microbiol. 8.

Jaillard, M., Lima, L., Tournoud, M., Mahé, P., van Belkum, A., Lacroix, V., Jacob, L., 2018. A Fast and Agnostic Method for Bacterial Genome-wide Association Studies: Bridging the Gap Between Kmers and Genetic Events. bioRxiv, pp. 297754.

Koutsoumanis, K.P., Lianou, A., 2013. Stochasticity in colonial growth dynamics of individual bacterial cells. Appl. Environ. Microbiol. 79, 2294–2301.

Kovacevic, J., Arguedas-Villa, C., Wozniak, A., Tasara, T., Allen, K.J., 2013. Examination of food chain-derived *Listeria monocytogenes* strains of different serotypes reveals considerable diversity in inlA genotypes, mutability, and adaptation to cold temperatures. Appl. Environ. Microbiol. 79, 1915–1922.

Kremer, P.H., Lees, J.A., Koopmans, M.M., Ferwerda, B., Arends, A.W., ... Feller, M.M., 2017. Benzalkonium tolerance genes and outcome in Listeria monocytogenes meningitis. Clin. Microbiol. Infect. 23 (4), 265–e1.

Kuenne, C., Billion, A., Mraheil, M.A., Strittmatter, A., Daniel, R., Goesmann, A., Barbuddhe, S., Hain, T., Chakraborty, T., 2013. Reassessment of the *Listeria monocytogenes* pan-genome reveals dynamic integration hotspots and mobile genetic elements as major components of the accessory genome. BMC Genomics 14, 47.

Laabei, M., Recker, M., Rudkin, J.K., Aldeljawi, M., Gulay, Z., Sloan, T.J., Williams, P., Endres, J.L., Bayles, K.W., Fey, P.D., 2014. Predicting the virulence of MRSA from its genome sequence. Genome Res. 24, 839–849.

Lees, J.A., Bentley, S.D., 2016. Bacterial GWAS: not just gilding the lily. Nat. Rev. Microbiol. 14, 406.

Lees, J.A., Vehkala, M., Välimäki, N., Harris, S.R., Chewapreecha, C., Croucher, N.J., Marttinen, P., Davies, M.R., Steer, A.C., Tong, S.Y., 2016. Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. Nat. Commun. 7, 12797.

Lianou, A., Koutsoumanis, K.P., 2012. Strain variability of the biofilm-forming ability of *Salmonella enterica* under various environmental conditions. Int. J. Food Microbiol. 160, 171–178.

Marinier, E., Zaheer, R., Berry, C., Weedmark, K.A., Domaratzki, M., Mabon, P., Knox, N.C., Reimer, A.R., Graham, M.R., Chui, L., 2017. Neptune: a bioinformatics tool for rapid discovery of genomic variation in bacterial populations. Nucleic Acids Res. 45 (18), e159.

Markkula, A., Mattila, M., Lindström, M., Korkeala, H., 2012. Genes encoding putative DEAD-box RNA helicases in *Listeria monocytogenes* EGD-e are needed for growth and motility at 3 °C. Environ. Microbiol. 14, 2223–2232.

Mattila, M., Somervuo, P., Rattei, T., Korkeala, H., Stephan, R., Tasara, T., 2012. Phenotypic and transcriptomic analyses of Sigma L-dependent characteristics in *Listeria monocytogenes* EGD-e. Food Microbiol. 32, 152–164.

Møretrø, T., Schirmer, B.C., Heir, E., Fagerlund, A., Hjemli, P., Langsrud, S., 2017. Tolerance to quaternary ammonium compound disinfectants may enhance growth of *Listeria monocytogenes* in the food industry. Int. J. Food Microbiol. 241, 215–224.

Moura, A., Criscuolo, A., Pouseele, H., Maury, M.M., Leclercq, A., Tarr, C., Björkman, J.T., Dallman, T., Reimer, A., Enouf, V., 2017. Whole genome-based population biology and epidemiological surveillance of *Listeria monocytogenes*. Nat. Microbiol. 2, 16185.

Nowak, J., Cruz, C.D., Tempelaars, M., Abee, T., van Vliet, A.H.M., Fletcher, G.C., Hedderley, D., Palmer, J., Flint, S., 2017. Persistent *Listeria monocytogenes* strains isolated from mussel production facilities form more biofilm but are not linked to specific genetic markers. Int. J. Food Microbiol. 256, 45–53.

Overney, A., Chassaing, D., Carpentier, B., Guillier, L., Firmesse, O., 2016. Development of synthetic media mimicking food soils to study the behaviour of *Listeria monocytogenes* on stainless steel surfaces. Int. J. Food Microbiol. 238, 7–14.

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T., Fookes, M., Falush, D., Keane, J.A., Parkhill, J., 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31, 3691–3693.

Pielaat, A., Boer, M.P., Wijnands, L.M., van Hoek, A.H., Bouw, E., Barker, G.C., Teunis, P.F., Aarts, H.J., Franz, E., 2015. First step in using molecular data for microbial food safety risk assessment; hazard identification of *Escherichia coli* O157: H7 by coupling genomic data with in vitro adherence to human epithelial cells. Int. J. Food Microbiol. 213, 130–138.

Piercey, M.J., Ells, T.C., Macintosh, A.J., Truelstrup Hansen, L., 2017. Variations in biofilm formation, desiccation resistance and Benzalkonium chloride susceptibility among *Listeria monocytogenes* strains isolated in Canada. Int. J. Food Microbiol. 257, 254–261.

Power, R.A., Parkhill, J., de Oliveira, T., 2017. Microbial genome-wide association studies: lessons from human GWAS. Nat. Rev. Genet. 18, 41.

Ragon, M., Wirth, T., Hollandt, F., Lavenir, R., Lecuit, M., Le Monnier, A., Brisse, S., 2008. A new perspective on *Listeria monocytogenes* evolution. PLoS Pathog. 4, e1000146.

Renier, S., Micheau, P., Talon, R., Hébraud, M., Desvaux, M., 2012. Subcellular localization of extracytoplasmic proteins in monoderm bacteria: rational secretomics-based strategy for genomic and proteomic analyses. PLoS One 7, e42982.

Salipante, S.J., Roach, D.J., Kitzman, J.O., Snyder, M.W., Stackhouse, B., Butler-Wu, S.M., Lee, C., Cookson, B.T., Shendure, J., 2015. Large-scale genomic sequencing of extraintestinal pathogenic *Escherichia coli* strains. Genome Res. 25, 119–128.

Saunders, L.P., Sen, S., Wilkinson, B.J., Gatto, C., 2016. Insights into the mechanism of homeoviscous adaptation to low temperature in branched-chain fatty acid-containing bacteria through modeling FabH kinetics from the foodborne pathogen *Listeria monocytogenes*. Front. Microbiol. 7.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069.

Shabala, L., Lee, S.H., Cannesson, P., Ross, T., 2008. Acid and NaCl limits to growth of *Listeria monocytogenes* and influence of sequence of inimical acid and NaCl levels on inactivation kinetics. J. Food Prot. 71, 1169–1177.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Suzuki, M., Shibayama, K., Yahara, K., 2016. A genome-wide association study identifies a horizontally transferred bacterial surface adhesin gene associated with antimicrobial resistant strains. Sci. Rep. 6, 37811.

Thorpe, H.A., Bayliss, S.C., Hurst, L.D., Feil, E.J., 2017a. Comparative analyses of selection operating on non-translated intergenic regions of diverse bacterial species. Genet. Genet. 116, 195784.

Thorpe, H.A., Bayliss, S.C., Sheppard, S.K., Feil, E.J., 2017b. Piggy: A Rapid, Large-scale Pan-genome Analysis Tool for Intergenic Regions in Bacteria. bioRxiv, pp. 179515.

Van Der Veen, S., Moezelaar, R., Abee, T., Wells-Bennik, M.H., 2008. The growth limits of a large number of *Listeria monocytogenes* strains at combinations of stresses show serotype-and niche-specific traits. J. Appl. Microbiol. 105, 1246–1258.

Yang, Z., Rannala, B., 2012. Molecular phylogenetics: principles and practice. Nat. Rev. Genet. 13, 303.

Zhou, X., Stephens, M., 2012. Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. 44, 821.

Zoz, F., Grandvalet, C., Lang, E., Iaconelli, C., Gervais, P., Firmesse, O., Guyot, S., Beney, L., 2017. *Listeria monocytogenes* ability to survive desiccation: influence of serotype, origin, virulence, and genotype. Int. J. Food Microbiol. 248, 82–89.