Two copies of the *ail* gene found in *Yersinia enterocolitica* and *Yersinia kristensenii*

Suvi Joutsen[a,b], Per Johansson[a], Riikka Laukkanen-Ninios[a,c], Johanna Björkroth[a] and Maria Fredriksson-Ahomaa[a]

[a]Department of Food Hygiene and Environmental Health, Faculty of Veterinary Medicine, P.O.Box 66 (Agnes Sjöbergin katu 2), 00014 University of Helsinki, Finland

[b]Risk Assessment Unit, Finnish Food Authority, Helsinki, Finland

[c]Food Safety Unit, Finnish Food Authority, Helsinki, Finland

Corresponding author:

E-mail address: maria.fredriksson-ahomaa@helsinki.fi

Abstract

Yersinia enterocolitica is the most common *Yersinia* species causing foodborne infections in humans. Pathogenic strains carry the chromosomal *ail* gene, which is essential for bacterial attachment to and invasion into host cells and for serum resistance. This gene is commonly amplified in several PCR assays detecting pathogenic *Y. enterocolitica* in food samples and discriminating pathogenic isolates from non-pathogenic ones. We have isolated several non-pathogenic *ail*-positive *Yersinia* strains from various sources in Finland. For this study, we selected 16 *ail*-positive *Yersinia* strains, which were phenotypically and genotypically characterised. Eleven strains were confirmed to belong to *Y. enterocolitica* and five strains to *Yersinia kristensenii* using whole-genome alignment, Parsnp and the SNP phylogenetic tree. All *Y. enterocolitica* strains belonged to non-pathogenic biotype 1A. We found two copies of the *ail* gene (*ail*1 and *ail*2) in all five *Y. kristensenii* strains and in one *Y. enterocolitica* biotype 1A strain. All 16 *Yersinia* strains carried the *ail*1 gene consisting of three different sequence patterns (A6-A8), which were highly similar with the *ail* gene found in high-pathogenic *Y. enterocolitica* biotype 1B strains (A2). The Ail protein encoded by the *ail*1 gene was highly conserved compared to the Ail protein encoded by the *ail*2 gene. Multiple sequence alignment of the *ail* gene and Ail protein were conducted with MAFF. In total, 10 *ail* sequence variations have been identified, of which 8 conserved ones belonged to the *ail*1 gene. According to our results, the detection of *ail* alone is not sufficient to predict the pathogenicity of *Yersinia* isolates.

Keywords: *Yersinia* spp.; *ail*; PCR detection; identification; pathogenicity

1. Introduction

The highly diverse genus *Yersinia*, including pathogenic and non-pathogenic species, has quite recently been classified in the *Yersiniaceae* family of the order *Enterobacteriales* (Adeolu et al. 2016). At the time of writing, it includes 19 species (Nguyen et al. 2019). Three species, *Yersinia enterocolitica, Yersinia pseudotuberculosis* and *Yersinia pestis*, are human pathogens. *Y. enterocolitica* is the most relevant species for human yersiniosis, which was the fourth most commonly reported enteric disease in Europe in 2018 (EFSA and ECDC 2019). This pathogen spreads typically through contaminated food or water but also through blood transfusion (Fredriksson-Ahomaa et al. 2012). *Y. enterocolitica* is a very heterogeneous species both biochemically and pathogenically (Fredriksson-Ahomaa et al. 2018). It can be divided into six biotypes, biotypes 1B (phylogroup 2) and 2–5 (phylogroups 3-6) of which are associated with yersiniosis and biotype 1A (phylogroup 1) is considered non-pathogenic due to lack of the virulence plasmid and important chromosomal virulence genes (Reuter et al. 2015). *Y. enterocolitica* biotype 1A strains mostly lack the classical chromosomal virulence genes *ail* and *yst*A. However, they usually carry the virulence-associated genes *inv*A and *yst*B (Batzilla et al. 2011; Hunter et al. 2019).

Enteropathogenic *Yersinia* invades the intestinal mucosa and proliferates in the Peyer's patches, i.e. lymphoid follicles of the small intestine. The chromosomal *ail* gene (attachment and invasion locus) of pathogenic *Yersinia* spp. encodes the small (17 kDa) outer membrane protein Ail, which is composed of eight transmembrane β-strands and four extracellular loops (1–4) of 10–21 amino acids (Miller et al. 2001). The Ail surface protein of *Y. enterocolitica* has many functions: it promotes attachment to and invasion into host cells and is critical for providing serum resistance (Miller et al. 2001; Bohn et al. 2019). Mutations in loops 2 and 3

62 of the Ail may lead to elimination of invasion and serum resistance of *Y. enterocolitica*
63 (Miller et al. 2001).
64
65 The *ail* gene has been shown to be highly conserved among *Y. enterocolitica* strains of the
66 same biotypes (Huang et al. 2010). Three sequence patterns (A1–A3) of the complete coding
67 sequence (CDS) of *ail* have been reported among pathogenic *Y. enterocolitica* strains: pattern
68 A1 is found in low-pathogenic strains belonging to biotypes 2–4, pattern A2 in highly
69 pathogenic strains of biotype 1B and pattern A3 has been found from a Chinese strain of
70 biotype 2. Huang et al. (2010) presumed that the *ail* gene of pathogenic *Y. enterocolitica*
71 strains have two original sequence patterns (A1 and A2), differing from each other with 21
72 mutations. Nine of these are missense mutations, which may have an effect on the function of
73 Ail and the virulence of the different biotypes. Three different *ail* sequences (named A4-A6
74 in our study) have been identified among non-pathogenic *Y. enterocolitica* biotype 1A strains
75 in earlier studies (Kraushaar et al. 2011, Liang et al. 2014, Platt-Samoraj et al. 2017).
76
77 Isolation and identification methods of *Y. enterocolitica* from clinical and food samples are
78 laborious and time-consuming, and require tests to differentiate pathogenic and non-
79 pathogenic isolates (Fredriksson-Ahomaa et al. 2018). The *ail* gene has widely been used as a
80 target gene in several PCR assays to quickly detect and identify pathogenic *Y. enterocolitica*
81 (Mäde et al. 2008; Thisted Lambertz et al. 2008; Petsios et al. 2016). It is also used in
82 validated standards for detecting pathogenic *Y. enterocolitica* directly from food or
83 environmental samples (ISO 2015) or for discriminating pathogenic *Yersinia* isolates from
84 non-pathogenic isolates (ISO 2017). The *ail* gene has sporadically been detected in *Y.*
85 *enterocolitica* strains belonging to non-pathogenic biotype 1A from humans (Sihvonen et al.
86 2011; Fredriksson-Ahomaa et al. 2012) and animals (Liang et al. 2014; Platt-Samoraj et al.

2017). We have quite recently detected the *ail* gene in non-pathogenic *Yersinia* strains isolated from wildlife (Joutsen et al. 2017; Sauvala et al. 2019), sheep (Joutsen et al. 2016) and lettuce (Nousiainen et al. 2016) in Finland.

In this study, we characterised a random collection of *ail*-positive non-pathogenic *Yersinia* strains isolated from various sources in Finland. The polymorphisms in the *ail* gene and Ail protein were explored using whole-genome sequence data.

2. Materials and Methods

2.1 Strains

In total, 16 *ail*-positive *Yersinia* strains were selected for characterisation and whole-genome sequencing. These strains have been isolated from different sources in Finland (Table 1). They were found from samples screened by PCR targeting the *ail* gene and regarded as non-pathogenic if they belonged to biotype 1A (Joutsen et al. 2017, Nousiainen et al. 2016, Sauvala et al. 2019). Seven *ail* sequences of *Y. enterocolitica* strains belonging to different biotypes, which have been reported in earlier studies, were included in the alignment analysis (Table 3).

Table 1

Origin of 16 *ail*-positive *Yersinia* strains isolated in Finland.

Species (Number of strains)	Source		Isolation year	Strain ID	References
Y. enterocolitica (1)	Mouse	Intestine	2005	F528D1	(Joutsen et al. 2017)
Y. enterocolitica (2)	Vole	Intestine	2002, 2005	M29A27, M34	(Joutsen et al. 2017)
Y. kristensenii (3)	Vole	Intestine	2005	M47, M70, M73	(Joutsen et al. 2017)
Y. kristensenii (1)	Shrew	Intestine	2005	M75	(Joutsen et al. 2017)
Y. enterocolitica (1)	Sheep	Faeces	2013	LAS383	(Joutsen et al. 2016)
Y. enterocolitica (3)	Deer	Carcass	2013	PR4, PR18, PR20	(Sauvala et al. 2019)
Y. enterocolitica (1)	Moose	Carcass	2013	HR88	(Sauvala et al. 2019)
Y. kristensenii (1)	Moose	Carcass	2013	HR100	(Sauvala et al. 2019)
Y. enterocolitica (1)	Mallard	Faeces	2013	SO16	(Sauvala et al. manuscript)
Y. enterocolitica (1)	Lettuce	Packaged	2013	PS23	(Nousiainen et al. 2016)
Y. enterocolitica (1)	Human	Faeces	1999	IHI111299	(Unpublished)

2.2 Strain characterisation

Identification of 16 *ail*-positive *Yersinia* strains was conducted with PCR targeting the 16S rRNA gene of *Y. enterocolitica* (Neubauer et al. 2000). The strains were characterised with API20E (BioMérieux, Marcy-l'Etoile, France) and biotyping (Joutsen et al. 2016). The presence of two virulence genes (*vir*F and *yad*A) on the virulence plasmid (pYV) and four virulence genes (*inv*A, *yst*A, *yst*B and *myf*) in the chromosome were studied by PCR (Joutsen et al. 2016).

2.3 Whole-genome sequencing and sequence analyses

Total DNA of *Yersinia* strains was purified using PureLink Genomic DNA Mini Kit (Invitrogen, Carlsbad, USA). The DNA library for Illumina sequencing was constructed using Nextera XT DNA Library Prep Kit (Illumina, CA, USA). Illumina NovaSeq6000 platform was used to generate 100 bp paired-end reads with 100x coverage (CeGaT, Center

for Genomics and Transcriptomics, Tuebingen, Germany). The raw reads were assembled with Spades (Bankevich et al. 2012).

The phylogenetic analysis of *Yersinia* genus was based on whole-genome alignment with Parsnp from Harvest bioinformatics suite (Treangen et al. 2014). Sixteen *ail*-positive *Yersinia* strains from our study along with 21 reference strains of 19 *Yersinia* spp. were used to construct the phylogenetic tree based on conserved core sequences (Figure S1). The average nucleotide identity (ANI) values of the 37 strains were calculated according to Richter and Rossello-Mora (2009) using PyANI (https://github.com/widdowquinn/pyani). The presence of the *ail, inv* and *yst* sequences were studied using BLAST (Altschul et al. 1990). Multiple sequence alignment of the *ail* sequences was conducted with MAFFT (Katoh et al. 2019). Phylogenetic trees based on the maximum likelihood principal were constructed with PhyML (Lefort et al. 2017) using the HKY85+I model for the whole CDSs (537 bp and 543 bp) and the HKY85 for the partial sequence (339 bp) of the *ail* gene. Multiple sequence alignment of the Ail protein was conducted with MAFFT. Sequences originating from plasmids among the sequence contigs were predicted using PLANETw (Vielva et al. 2017)

2.4 Data submission

The draft genomes of 11 *Y. enterocolitica* and 5 *Yersinia kristensenii* strains have been deposited in NCBI under BioProject ID: PRJNA636668.

3. Results

Eleven out of 16 *ail*-positive *Yersinia* strains were identified as *Y. enterocolitica* and 5 as *kristensenii* by PCR targeting the 16S rRNA gene of *Y. enterocolitica*. Nine out of 11 *Y. enterocolitica* strains, which were all utilizing sucrose (sucrose-positive strains), were correctly identified with API20E V5.0 (APIWEB™, BioMérieux) showing a %ID of 98.3. The *Y. kristensenii* strains had a %ID of 92.5 or 99.4. The two sucrose-negative *Y. enterocolitica* strains were incorrectly identified as *Y. kristensenii* (%ID=79.0). All *Y. enterocolitica* strains belonged to biotype 1A (Table 2).

Table 2
Characteristics of the 16 sequenced *ail*-positive *Yersinia* strains.

Species (Nr. of strains)	Strain ID	API 20E	Bio-type	PCR positive for					
				*vir*F	*yad*A	*inv*A	*yst*A	*yst*B	*myf*
Y. enterocolitica (9)	F528D1, HR88, PR4, PR18, PR20, SO16, PS23, LAS383, IHI111299	1155523 1155723	1A	-	-	+	-	+	-
Y. enterocolitica (2)	M29A27, M34	1155503	1A	-	-	+	-	+	-
Y. kristensenii (5)	M47, M70, M73, M75, HR100	1154503 1354503	NT	-	-	-	-	+	-

NT=not typable

All 16 *ail*-positive *Yersinia* strains were negative for the *vir*F and *yad*A genes located on the virulence plasmid and negative for the *yst*A and *myf* genes located in the chromosome. These genes are associated with the pathogenicity and typically found only in pathogenic *Y. enterocolitica* strains. All *Y. enterocolitica* strains carried the *inv*A and *yst*B genes, while all *Y. kristensenii* strains were negative for the *inv*A gene and positive for *yst*B gene.

In the phylogenetic analysis based on aligned core sequences, all our 11 *ail*-positive *Y. entercolitica* biotype 1A strains clustered together with *Y. enterocolitica* reference strains and 5 *ail*-positive *Y. kristensenii* strains together with the *Y. kristensenii* reference strain (Figure

S1). The average nucleotide identity (ANI) values between our 11 *ail*-positive *Y. enterocolitica* biotype 1A strains and *Y. enterocolitica* reference strains were above 95% and the ANI values between our 5 *ail*-positive *Y. kristensenii* and the *Y. kristensenii* reference strain were above 98% (Table S1). Our 11 *Y. enterocolitica* biotype 1A strains formed three groups (G1, G2 and G3) (Figure S1), which could also be confirmed with the ANI values (Table S1).

In total, we found four different *ail* sequences (A6–A9) among the *Y. enterocolitica* strains and two sequences (A6 and A10) among the *Y. kristensenii* strains (Table 3). Four sequence patterns (A7–A10) have not been reported before. We also report, for the first time, two *ail* genes in the *Y. enterocolitica* and *Y. kristensenii* strains. We named the *ail* gene found in all the strains *ail*1 and the second *ail* gene found in one *Y. enterocolitica* biotype 1A strain and in all five *Y. kristensenii* strains we named *ail*2 (Table 3). The CDSs of *ail*1 in all strains and CDSs of *ail*2 in the *Y. kristensenii* strains were 537 bp long, while the CDS of *ail*2 in *Y. enterocolitica* 1A was 543 bp long. Our *Y. enterocolitica* biotype 1A strains with *ail* sequence patterns A6, A7 and A8 formed groups G3, G2 and G1, respectively, in the phylogenetic analysis based on aligned core sequences (Figure S1). *Y. enterocolitica* strain (PS23) in the group G3 carried also *ail*2 gene with pattern A9. All *Y. kristensenii* strains carrying *ail*1 of pattern A6 and *ail*2 of pattern A10 belonged to group G4.

Table 3

Different *ail* patterns found in *Y. enterocolitica* (YE) and *Y. kristensenii* (YK) strains.

Sequence pattern		Species/ biotype	Strain ID	Sequence size (bp)		Reference
*ail*1	*ail*2			*ail*1	*ail*2	
A1	ND	YE/4	$Y11^T$	537		Huang et al. 2010
A1	ND	YE/1A	SDWL-003	537		Liang et al. 2014
A2	ND	YE/1B	8081	537		Huang et al. 2010
A3	ND	YE/2	NX1997	537		Huang et al. 2010
A4	ND	YE/1A	2006RAT	537		Liang et al. 2014
A5	ND	YE/1A	256-P	338		Platt-Samoraj et al. 2017
A6	ND	YE/1A	Y30/09	537		Kraushaar et al. 2011
A6	A9	YE/1A	PS23	537	543	This study
A6	A10	YK	HR100, M47, M70, M73, M75	537	537	This study
A7	ND	YE/1A	SO16, LAS383, HR88, PR4, PR18, PR20	537		This study
A8	ND	YE/1A	IHI111299, F528D1, M29A27, M34	537		This study

ND=not detected

The *ail*1 sequence patterns A6–A8 showed a similarity between 99.4% and 99.8% with only one to three base mutations (Table 4). All mutations were missense mutations. Sequence patterns A6–A8 were highly similar with the *ail*1 sequence pattern A2 formed by high-pathogenic *Y. enterocolitica* biotype 1B strains reported by Huang et al. (2010). The *ail*2 sequence patterns A9 and A10 found in *Y. enterocolitica* biotype 1A and *Y. kristensenii*, respectively, had a low sequence similarity (79.2%) and also differed clearly from *ail*1 sequence patterns A1–A8 (Table 4). Most of the point mutations in the sequence patterns A9 (65/106) and A10 (80/125) were missense mutations.

Table 4

Similarity (%) of sequence patterns (A1–A10) of whole CDSs of the *ail* genes in *Y. enterocolitica* (YE) and *Y. kristensenii* (YK) strains.

Sequence pattern	Yersinia species	Biotype	A1	A2	A3	A4	A6	A7	A8	A9	A10
A1	YE	2-4	100.0	96.1	99.4	92.9	96.3	96.1	96.3	80.1	76.0
A2	YE	1B	96.1	100.0	95.5	94.2	99.8	99.6	99.4	80.5	75.6
A3	YE	2	99.4	95.5	100.0	92.4	95,7	95.5	95.7	79.7	75.6
A4	YE	1A	92.9	94.2	92.4	100.0	94.4	94.2	94.0	80.7	78.9
A6	YE YK	1A	96.3	99.8	95.7	94.4	100.0	99.8	99.6	80.7	77.0
A7	YE	1A	96.1	99.6	95.5	94.2	99.8	100.0	99.4	80.5	76.8
A8	YE	1A	96.3	99.4	95.7	94.0	99.6	99.4	100.0	80.7	77.0
A9	YE	1A	80.1	80.5	79.7	80.7	80.7	80.5	80.7	100.0	79.2
A10	YK		76.0	75.6	75.6	78.9	77.0	76.8	77.0	79.2	100.0

Similarity (%):

>99, >95, >90, >80, >75

The *ail*1 sequences patterns A6–A8 clustered together with the *ail* sequence pattern A2 of high-pathogenic *Y. enterocolitica* biotype 1B (Figure 1). The *ail*2 patterns A9 and A10 formed their own branches. The *ail*1 sequence pattern A6 was found in all of our *Y. kristensenii* strains and in one *Y. enterocolitica* strain. The same pattern was also found in a German *Y. enterocolitica* 1A strain (Y30/09) earlier described by Kraushaar et al. (2011). The *ail*1 sequence patterns A6–A8 were clearly different from sequence patterns A1 and A3 found in low-pathogenic *Y. enterocolitica* strains (Huang et al. 2010) and pattern A4 found in a *Y. enterocolitica* biotype 1A strain (Liang et al. 2014).

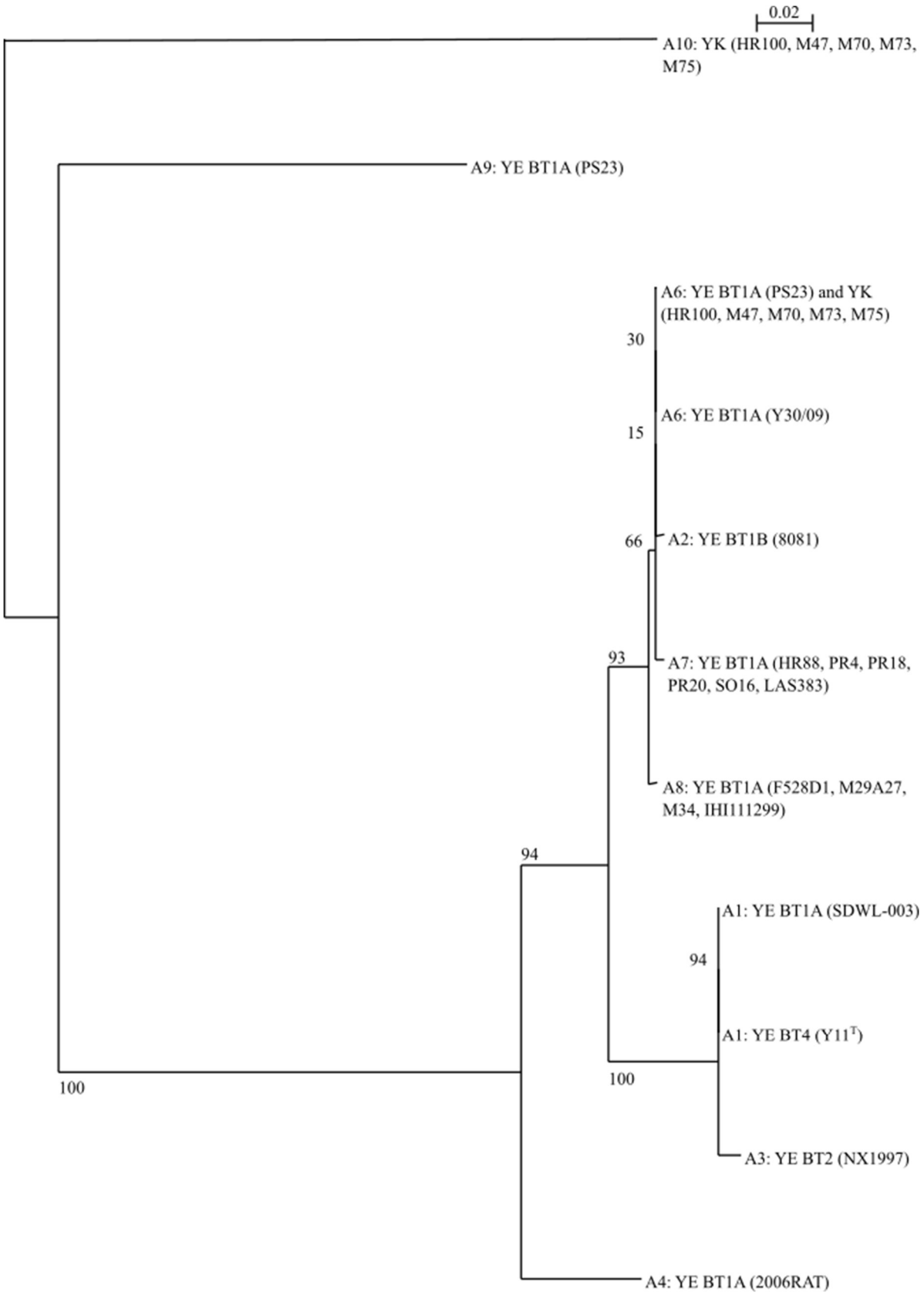

Fig.1. Maximum likelihood phylogenetic tree based on the whole CDSs of the *ail* genes of *Y. enterocolitica* (YE) and *Y. kristensenii* (YK) strains generated using PhyML. Numbers at the nodes indicate the % likelihood of that branch assignment. The scale represents a distance of 0.02 residue substitutions per site for the branch length.

The *ail*1 sequence pattern A5 was found in a *Y. enterocolitica* biotype 1A strain (256-P) from Poland (Platt-Samoraj et al. 2017) (Figure S2). Only a partial CDS (339 bp) was reported in this study. In the multiple sequence analysis of partial (339 bp) *ail* sequences, all sequence patterns A6–A8 found in our study were clustered together (Figure S2). This group also included two partial CDSs (394 bp) of the *ail* reported in *Y. enterocolitica* biotype 1A strains from Finland (Sihvonen et al. 2011) and 21 partial coding sequences (339 bp) of the *ail* reported in *Y. enterocolitica* biotype 1A strains from Poland (Platt-Samoraj et al. 2017).

The amino acid sequences AA6–AA8 of the Ail protein of our 16 *ail*-positive *Yersinia* strains were highly similar with amino acid sequence AA2 of the high-pathogenic *Y. enterocolitica* biotype 1B (Table 5). One to three amino acid replacements were found. They were located in loops 3 and 4 of the Ail protein (Figure 2A and Figure S3a-d). The amino acid sequences AA9 and AA10 encoded by the *ail*2 gene were not conserved. All loops of the Ail2 protein contained several amino acid replacements (Figure S3e-f).

Table 5
Similarity (%) of the amino acid sequences of the Ail protein of *Y. enterocolitica* (YE) and *Y. kristensenii* (YK) strains.

Amino acid sequences	Yersinia species	Biotype	Amino acid sequences							
			AA1	AA2	AA3	AA6	AA7	AA8	AA9	AA10
AA1	YE	2-4	100.0	94.9	99.4	95.5	94.9	95.5	75.6	70.2
AA2	YE	1B	94.9	100.0	94.4	99.4	98.9	98.3	76.1	71.9
AA3	YE	2	99.4	94.4	100.0	94.9	94.4	94.9	75,0	69.7
AA6	YE YK	1A	95.5	99.4	94.9	100.0	99.4	98.9	76.1	71.9
AA7	YE	1A	94.9	98.9	94.4	99.4	100.0	98.3	75.6	71.3
AA8	YE	1A	95.5	98.3	94.9	98.9	98.3	100.0	76.1	71.9
AA9	YE	1A	75.6	76.1	75.0	76.1	75.6	76.1	100.0	73.3
AA10	YK		70.2	71.9	69.7	71.9	71.3	71.9	73.3	100.0

Similarity (%):
>98, >94

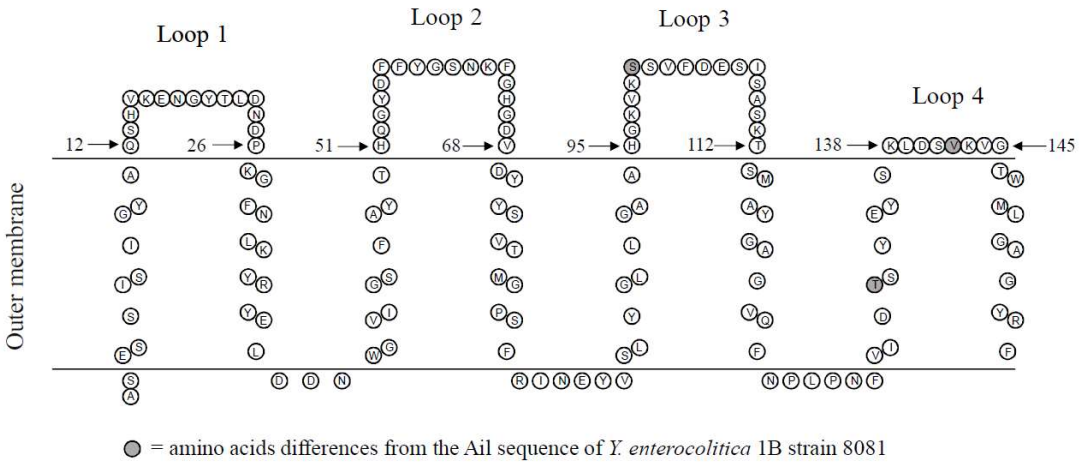

Fig. 2. Position of Ail residues in *Y. enterocolitica* strains encoded by *ail*1.

4. Discussion

All human pathogenic *Yersinia* spp. (*Y. enterocolitica*, *Y. pseudotuberculosis* and *Y. pestis*) carry the *ail* gene in their chromosome. This gene has only sporadically been found in *Y. enterocolitica* biotype 1A strains, which are regarded as non-pathogenic strains because they typically miss the most important virulence genes (Hunter et al. 2019). In this study, we characterised 11 *ail*-positive *Y. enterocolitica* biotype 1A and 5 *ail*-positive *Y. kristensenii* strains using several methods. We could confirm with the whole-genome alignment based on conserved sequences and ANI values that all our *Y. enterocolitica* and *Y. kristensenii* strains had been correctly identified using the PCR based on 16S rRNA (Neubauer et al. 2000). The sucrose-negative *Y. enterocolitica* strains could not be correctly identified with the API20E system. Correct identification of *Yersinia* spp. with biochemical tests may sometimes be impossible (Fredriksson-Ahomaa et al. 2018). All our strains were negative for the *vir*F and *yad*A genes located on the virulence plasmid and for the chromosomal *yst*A and *myf*A genes by PCR. These genes are all important virulence genes found in pathogenic strains (Batzilla et al. 2011). All our *Y. enterocolitica* strain were *inv*A- and *yst*B-positive by PCR. These virulence-associated chromosomal genes have been detected in *Y. enterocolitica* biotype 1A strains (Batzilla et al. 2011; Hunter et al. 2019). All our *Y. kristensenii* strains were *inv*A-negative but they were all *yst*B-positive by PCR. However, the *yst*B sequence found in our *Y. kristensenii* strains was different from the *yst*B found in our *Y. enterocolitica* strains. The function of YstB enterotoxin in the pathogenesis of yersiniosis remains unclear.

We detected a highly conserved *ail* gene (*ail*1) in all *Y. enterocolitica* biotype 1A and *Y. kristensenii* strains, which was very similar and clustered together with the *ail* gene in the high-pathogenic *Y. enterocolitica* biotype 1B. Only one to three nucleotide changes were

observed; however, they were all missense mutations, which may have an effect on the function and virulence of the gene. An identical *ail*1 sequence was found in one *Y. enterocolitica* biotype 1A strain isolated from minced pork in Germany (Kraushaar et al. 2011), from human and lettuce samples in Finland (Sihvonen et al. 2011) and wild boars in Poland (Platt-Samoraj et al. 2017), suggesting that *ail*1 is very conserved in non-pathogenic *Y. enterocolitica* biotype 1A in Europe. However, we detected three phylogenetically slightly different *ail*1 sequences in our *Y. enterocolitica* biotype 1A strains, indicating that strains with different *ail* sequence patterns may have originated from different sources.

Unexpectedly, we detected a second *ail* (*ail*2) gene in all five *Y. kristensenii* strains and in one *Y. enterocolitica* strain. This gene was very non-conserved and highly different from *ail*1. Several missense mutations occurred in the *ail*2 gene, which probably affect the function of this gene. The *ail*2 sequence was identical in all *Y. kristensenii* strains but was highly different from the *ail*2 of the *Y. enterocolitica* strain, indicating that the *ail*2 gene has most likely been gained from a different source. The *ail*2 was possibly located on a prophage of our *Y. kristensenii* strains, while it was located on a plasmid in our *Y. enterocolitica* strain. More studies are needed concerning the presence and function of various *ail* genes in non-pathogenic *Yersinia* strains.

The few point mutations in the *ail*1 sequences of *Y. enterocolitica* biotype 1A and *Y. kristensenii* strains were missense mutations changing the amino acids of the Ail protein. One to three amino acid replacements occurred and they were all located in loops 3 and 4 of the Ail. The single amino acid change at A100 (A100S) in loop 3 may decrease the serum resistance but not the invasion activity according to Miller et al. (2001). This indicates that non-pathogenic *ail*-positive *Yersinia* strains have an ability to colonize the animal host, which

could explain why *ail*-positive *Y. enterocolitica* 1A and *Y. kristensenii* were frequently isolated from the intestine of voles and shrew in Finland (Joutsen et al. 2016). However, further studies are needed to explore whether these amino acid changes alter the function of the Ail protein and the virulence of the strains. Numerous missense point mutations occurred in the *ail*2 gene, which strongly affected the amino acid composition of the Ail protein. Several amino acid replacements were located in all loops. Mutations in loops 2 and 3 of the Ail protein have shown to significantly decrease and even eliminate the attachment and invasion capacity and the serum resistance of *Y. enterocolitica* strains (Miller et al. 2001). The meaning of the Ail protein encoded by *ail*2 should be explored more.

The pYV virulence plasmid is essential for the pathogenesis of yersiniosis but it may be lost during subculturing, leading to false-negative results, and therefore chromosomal virulence genes are preferred as PCR targets (Petsios et al. 2016). The chromosomal *ail* gene is one of the most frequently used targets for detection and identification of pathogenic *Yersinia*: however, the *ail* gene has frequently been reported in *Y. enterocolitica* biotype 1A strains from wildlife (Joutsen et al. 2017; Platt-Samoraj et al. 2017). The *ail* primers used in the European accredited methods (ISO 2015, 2017) detected the *ail*1 gene of our *Y. enterocolitica* biotype 1A and *Y. kristensenii* strains and therefore other targets are also needed. Parallel with *ail*, we suggest the use of a PCR target located on the pYV for detection and identification of pathogenic *Yersinia* isolates, especially when the detection of pathogenic *Yersinia* is performed directly from clinical or food samples and no isolates are available for further characterisation. When the isolate is available, analysis of whole-genome sequencing data would provide information on both the bioserotype and potential pathogenicity of the isolate. More research is needed to assess the potential virulence of *Yersinia* strains harbouring the chromosomal *ail*1 and *yst*B genes but missing the pYV.

5. Conclusions

Our results demonstrated that the *ail*1 gene is conserved among *Y. enterocolitica* biotype 1A and *Y. kristensenii* strains and is highly similar with the *ail* gene found in high-pathogenic *Y. enterocolitica* biotype 1B strains. The functionality and virulence of the *ail*1 gene found in our study needs to be clarified. A second *ail* gene (*ail*2), which was not conserved, was found in all *Y. kristensenii* strains and in one *Y. enterocolitica* strain. The Ail protein encoded by *ail*2 had several amino acid replacements in loops 2 and 3, which probably eliminate the attachment and invasion capacity and cause loss of serum resistance. The prevalence and meaning of the *ail*2 gene in *Yersinia* strains need more studies. The validated standard methods used to detect pathogenic *Y. enterocolitica* detect the *ail*1 gene found in non-pathogenic *Yersinia* strains, thus giving a false-positive result.

Acknowledgements

We thank Mirko Rossi (EFSA) for scientific discussions and Maria Stark (University of Helsinki) for technical assistance.

References

Adeolu, M., Alnajar, S., Naushad, S., Gupta, R.S., 2016. Genome-based phylogeny and taxonomy of the '*Enterobacteriales*': proposal for *Enterobacterales* ord. nov. divided into the families *Enterobacteriaceae, Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov. Int. J. Syst. Evol. Microbiol. 66, 5575-5599.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403-410.

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.

356 Batzilla, J., Heesemann, J., Rakin, A., 2011. The pathogenic potential of *Yersinia*
357 *enterocolitica* 1A. Int. J. Med. Microbiol. 301, 556-561.

358 Bohn, E., Sonnabend, M., Klein, K., Autenrieth, I.B., 2019. Bacterial adhesion and host cell
359 factors leading to effector protein injection by type III secretion system. Int. J. Med.
360 Microbiol. 309, 344-350.

361 EFSA and ECDC, 2019. The European Union one health 2018 zoonoses report. EFSA J. 17,
362 5926.

363 Fredriksson-Ahomaa, M., Fredriksson-Ahomaa, M., Cernela, N., Cernela, N., Hächler, H.,
364 Hächler, H., Stephan, R., Stephan, R., 2012. Yersinia enterocolitica strains associated
365 with human infections in Switzerland 2001–2010. Eur. J. Clin. Microbiol. Infect. Dis.
366 31, 1543-1550.

367 Fredriksson-Ahomaa, M., Joutsen, S., Laukkanen-Ninios, R., 2018. Identification of *Yersinia*
368 at the species and subspecies levels is challenging. Curr. Clin. Microbiol. Rep. 5, 135-
369 142.

370 Huang, Y., Wang, X., Cui, Z., Yang, Y., Xiao, Y., Tang, L., Kan, B., Xu, J., Jing, H., 2010.
371 Possible use of *ail* and *foxA* polymorphisms for detecting pathogenic *Yersinia*
372 *enterocolitica*. BMC Microbiol. 10, 211.

373 Hunter, E., Greig, D.R., Schaefer, U., Wright, M.J., Dallman, T.J., McNally, A., Jenkins, C.,
374 2019. Identification and typing of *Yersinia enterocolitica* and *Yersinia*
375 *pseudotuberculosis* isolated from human clinical specimens in England between 2004
376 and 2018. J. Med. Microbiol. 68, 538-548.

377 ISO, 2017. Microbiology of the food chain - Horizontal method for the detection of
378 pathogenic *Yersinia enterocolitica* (ISO 10273). International Organization for
379 Standardization, 1-47.

380 ISO, 2015. Microbiology of the food chain – Polymerase chain reaction (PRC) for the
381 detection of food-borne pathogens – Detection of pathogenic *Yersinia enterocolitica* and
382 *Yersinia pseudotuberculosis* (ISO 18867). International Organization for
383 Standardization, 1-30.

384 Joutsen, S., Eklund, K., Laukkanen-Ninios, R., Stephan, R., Fredriksson-Ahomaa, M., 2016.
385 Sheep carrying pathogenic *Yersinia enterocolitica* bioserotypes 2/O:9 and 5/O:3 in the
386 feces at slaughter. Vet. Microbiol. 197, 78-82.

387 Joutsen, S., Laukkanen-Ninios, R., Henttonen, H., Niemimaa, J., Voutilainen, L., Kallio,
388 E.R., Helle, H., Korkeala, H., Fredriksson-Ahomaa, M., 2017. *Yersinia* spp. in wild
389 rodents and shrews in Finland. Vector-Borne Zoonotic Dis. 17, 303-311.

390 Katoh, K., Rozewicki, J., Yamada, K.D., 2019. MAFFT online service: multiple sequence
391 alignment, interactive sequence choice and visualization. Briefings in bioinformatics 20,
392 1160-1166.

393 Kraushaar, B., Dieckmann, R., Wittwer, M., Knabner, D., Konietzny, A., Mäde, D., Strauch,
394 E., 2011. Characterization of a *Yersinia enterocolitica* biotype 1A strain harbouring an
395 *ail* gene. J. Appl. Microbiol. 111, 997-1005.

396 Lefort, V., Longueville, J., Gascuel, O., 2017. SMS: smart model selection in PhyML. Mol.
397 Biol. Evol. 34, 2422-2424.

398 Liang, J., Bi, Z., Shi, G., Xiao, Y., Qiu, H., Kou, Z., Hu, B., Jing, H., Wang, X., 2014. Two
399 novel *ail*-positive biotype 1A strains of *Yersinia enterocolitica* isolated in China with
400 unequal adhesion and invasion properties. Inf. Gen. Evol. 27, 83-88.

401 Mäde, D., Reiting, R., Strauch, E., Ketteritzsch, K., Wicke, A., 2008. A real-time PCR for
402 detection of pathogenic *Yersinia enterocolitica* in food combined with an universal
403 internal amplification control system. J. Verbrauch. Lebensm. 3, 141-151.

404 Miller, V.L., Beer, K.B., Heusipp, G., Young, B.M., Wachtel, M.R., 2001. Identification of
405 regions of Ail required for the invasion and serum resistance phenotypes. Mol.
406 Microbiol. 41, 1053-1062.

407 Neubauer, H., Aleksic, S., Hensel, A., Finke, E.J., Meyer, H., 2000. *Yersinia enterocolitica*
408 16S rRNA gene types belong to the same genospecies but form three homology groups.
409 Int. J. Med. Microbiol. 290, 61-64.

410 Nguyen, S.V., Muthappa, D.M., Hurley, D., Donoghue, O., McCabe, E., Anes, J., Schaffer,
411 K., Murphy, B.P., Buckley, J.F., Fanning, S., 2019. *Yersinia hibernica* sp. nov., isolated
412 from pig-production environments. Int. J. Syst. Evol. Microbiol.

413 Nousiainen, L., Joutsen, S., Lunden, J., Hänninen, M., Fredriksson-Ahomaa, M., 2016.
414 Bacterial quality and safety of packaged fresh leafy vegetables at the retail level in
415 Finland. Int. J. Food Microbiol. 232, 73-79.

416 Petsios, S., Fredriksson-Ahomaa, M., Sakkas, H., Papadopoulou, C., 2016. Conventional and
417 molecular methods used in the detection and subtyping of *Yersinia enterocolitica* in
418 food. Int. J. Food Microbiol. 237, 55-72.

419 Platt-Samoraj, A., Syczylo, K., Szczerba-Turek, A., Bancerz-Kisiel, A., Jablonski, A., Labuc,
420 S., Pajdak, J., Oshakbaeva, N., Szweda, W., 2017. Presence of *ail* and *yst*B genes in
421 *Yersinia enterocolitica* biotype 1A isolates from game animals in Poland. Vet. J. 221,
422 11-13.

423 Reuter, S., Corander, J., de Been, M., Harris, S., Cheng, L., Hall, M., Thomson, N.R.,
424 McNally, A., 2015. Directional gene flow and ecological separation in *Yersinia*
425 *enterocolitica*. Microb. Genom. 1, 1-9.

426 Richter, M., Rossello-Mora, R., 2009. Shifting the genomic gold standard for the prokaryotic
427 species definition. PNAS 106, 19126-19131.

428 Sauvala, M., Laaksonen, S., Laukkanen-Ninios, R., Jalava, K., Stephan, R., Fredriksson-
429 Ahomaa, M., 2019. Microbial contamination of moose (*Alces alces*) and white-tailed

deer (*Odocoileus virginianus*) carcasses harvested by hunters. Food Microbiol. 78, 82-88.

Sihvonen, L.M., Hallanvuo, S., Haukka, K., Skurnik, M., Siitonen, A., 2011. The *ail* gene is present in some *Yersinia enterocolitica* biotype 1A strains. Foodborne Pathog. Dis. 8, 455-457.

Thisted Lambertz, S., Nilsson, C., Hallanvuo, S., Lindblad, M., 2008. Real-time PCR method for detection of pathogenic *Yersinia enterocolitica* in food. Appl. Environ. Microbiol. 74, 6060-6067.

Treangen, T.J., Ondov, B.D., Koren, S., Phillippy, A.M., 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol. 15, 524.

Vielva, L., de Toro, M., Lanza, V.F., de la Cruz, F., 2017. PLACNETw: a web-based tool for plasmid reconstruction from bacterial genomes. Bioinf. 33, 3796-3798.