WGS analysis of *Listeria monocytogenes* from rural, urban, and farm environments in Norway: Genetic diversity, persistence, and relation to clinical and food isolates

Annette Fagerlund[a]#, Lene Idland[b], Even Heir[a], Trond Møretrø[a], Marina Aspholm[b], Toril Lindbäck[b], and Solveig Langsrud[a]

[a] Nofima, Norwegian Institute of Food, Fisheries and Aquaculture Research, Ås, Norway

[b] Department of Paraclinical Sciences, Faculty of Veterinary Medicine, Norwegian University of Life Sciences, Ås, Norway

Running Head: WGS of environmental *L. monocytogenes* from Norway

Keywords: *Listeria monocytogenes*, whole genome sequencing, WGS, source tracking, persistence, molecular epidemiology, environmental pathogens, dairy farms.

Address correspondence to Annette Fagerlund, annette.fagerlund@nofima.no

Annette Fagerlund and Lene Idland contributed equally to this work.

Abstract

Listeria monocytogenes is a ubiquitous environmental bacterium associated with a wide variety of natural and man-made environments, such as soil, vegetation, livestock, food processing environments, and urban areas. It is also among the deadliest foodborne pathogens, and knowledge about its presence and diversity in potential sources is crucial to effectively track and control it in the food chain. Isolation of *L. monocytogenes* from various rural and urban environments showed higher prevalence in agricultural and urban developments than in forest or mountain areas, and that detection was positively associated with rainfall. Whole genome sequencing (WGS) was performed for the collected isolates and for *L. monocytogenes* from Norwegian dairy farms and slugs, in total 218 isolates. The data was compared with available datasets from clinical and food associated sources in Norway collected within the last decade. Multiple examples of clusters of isolates with 0-8 wgMLST allelic differences were collected over time in the same location, demonstrating persistence of *L. monocytogenes* in natural, urban and farm environments. Furthermore, several clusters with 6-20 wgMLST allelic differences containing isolates collected across different locations, times and habitats were identified, including nine clusters harbouring clinical isolates. The most ubiquitous clones found in soil and other natural and animal ecosystems (CC91, CC11, and CC37) were distinct from clones predominating among both clinical (CC7, CC121, CC1) and food (CC9, CC121, CC7, CC8) isolates. The analyses indicated that ST91 was more prevalent in Norway than other countries and revealed a high proportion of the hypovirulent ST121 among Norwegian clinical cases.

Importance

Listeria monocytogenes is a deadly foodborne pathogen that is widespread in the environment. For effective management, both public health authorities and food producers need reliable tools for source tracking, surveillance, and risk assessment. For this, whole genome sequencing (WGS) is regarded as the present and future gold standard. In the current study, we use WGS to show that *L. monocytogenes* can persist for months and years in natural, urban and dairy farm environments. Notably, clusters of almost identical isolates, with genetic distances within the thresholds often suggested for defining an outbreak cluster, can be collected from apparently unrelated samples. The work highlights the need for a greater knowledge of the genetic relationships between clinical isolates and isolates of *L. monocytogenes* from a wide range of environments, including natural, urban, agricultural, livestock, food production, and food processing environments, in order to correctly interpret and use results from WGS analyses.

Introduction

Listeria monocytogenes is a bacterial pathogen responsible for the life-threatening disease listeriosis. The most common cause of listeriosis is considered to be ingestion of food contaminated by *L. monocytogenes* from unclean food production equipment (1, 2). *L. monocytogenes* is a ubiquitous environmental bacterium that has been associated with a wide variety of environments, such as rivers, soil, vegetation, wild and domesticated animals, food processing environments, and urban areas (3, 4). Consequently, a total absence of *L. monocytogenes* in non-heat-treated foods is difficult, perhaps impossible, to achieve. The literature is, however, not fully consistent about the main habitats of *L. monocytogenes* and the factors affecting its occurrence and spread to humans. It is therefore of importance to increase the understanding of the relationship between *L. monocytogenes* in natural and animal reservoirs, food processing environments, and human clinical disease.

The occurrence of *L. monocytogenes* in soil varies widely, from 0.7% to 45%, depending on the geographic area, season, and humidity (4-6). In comparative investigations, higher frequencies of *L. monocytogenes* have been found after rain, flooding and irrigation events (7, 8). Several studies have reported high incidence of *L. monocytogenes* in water from rivers and lakes, with frequencies from 10-62% of the samples depending on the area and detection method (9-13). A link has been found between the proximity to upstream dairy farms and cropped land and the presence of *L. monocytogenes* in river water (10, 12). An explanation for this could be high frequencies of *L. monocytogenes* in feces from farm animals, e.g., cattle, ducks, and sheep, leaking into surrounding soil and water (9, 14, 15). Dairy farms are, for example, known to hold a *L. monocytogenes* reservoir, and prevalences in environmental samples of 11-24% have been reported (6, 15-17). However, *L. monocytogenes* is not particularly linked to farm animals, and is frequently found in other animals and birds, such as game and urban birds, boars, garden slugs and rodents (9, 18-21). An association between dense populations of humans and occurrence of *L. monocytogenes* in the environment has been reported. A U.S. study showed that 4.4% of samples from urban or residential areas contained *L. monocytogenes*, while the pathogen was less frequently found in samples from forests and mountains (1.3%) (22).

L. monocytogenes comprises four separate deep-branching lineages, which from an evolutionary viewpoint could be considered separate species (23). These are further subdivided by multilocus sequence typing (MLST) into sequence types (STs) and clonal complexes (CCs or clones). The lineage I clones CC1, CC2, CC4, and CC6 are reported to be associated with human disease, while lineage II clones CC9 and CC121 are strongly associated with food and food processing environments (24-28).

79 While many studies have examined the molecular genotypes of *L. monocytogenes* isolates found in
80 food, food processing environments, and clinical disease, much less is known about the diversity
81 present in other environments. In the few published studies, the clonal diversity in environmental
82 samples from soil and water appears to be very high, sometimes dominated by CCs associated with
83 disease (CC1, CC4), although other dominating clones (e.g., CC37) have also been observed (5, 13,
84 29). Several clones are also found in wild animals, e.g., CC7 and CC37 found in moose, boars, slugs,
85 and game birds (18, 21, 30, 31). In environmental samples from dairy/cattle farms in Finland and
86 Latvia, the lineage II clones CC11 (ST451), CC14, CC18, CC20, CC37, and CC91 were most
87 predominant, while lineage I clones were rare (6, 32). Although there are some exceptions, e.g., CC1
88 being predominant in slugs collected in garden and farm environments in Norway (21), the majority
89 of the clones identified in natural and farm environments do not seem to belong to CCs dominant
90 among European food and clinical isolates.

91 Many studies have described persistence over time for *L. monocytogenes* clones in food processing
92 facilities (2, 33, 34) and in individual cattle herds or farm environments (15, 35, 36). Whether *L.*
93 *monocytogenes* can persist over long periods of time also in rural, urban, or agricultural
94 environments has rarely been investigated. Studies of genetic relationships between *L.*
95 *monocytogenes* isolates from natural and animal reservoirs and isolates from food and clinical
96 sources are scarce. High resolution molecular fingerprinting based on whole genome sequencing
97 (WGS) technology has revolutionized the ability to detect outbreaks and the presence of persistent
98 strains (37). Few studies have, however, carried out WGS analyses of *L. monocytogenes* isolates
99 collected from non-food associated locations over the span of months and years. The present study
100 aimed to use WGS to investigate the diversity and genetic relationships between *L. monocytogenes*
101 isolates from rural, agricultural, and urban environments in Norway, and to compare these with
102 available datasets containing genomes of *L. monocytogenes* from human clinical and food-associated
103 sources in Norway collected within the last decade.

104 Results

105 Prevalence of *L. monocytogenes* in rural and urban environments in Norway.

106 A total of 618 distinct environmental sites from rural and urban environments were sampled for *L.*
107 *monocytogenes* between April 2016 and April 2020. The overall sampling scheme was designed to
108 obtain an overview of the presence of *L. monocytogenes* in various habitats, and samples were
109 collected from several geographical regions in Norway (Figure 1A). To study potential persistence of
110 *L. monocytogenes* clones over time, some sites were sampled more than once. At the onset of the
111 study, we hypothesized that the presence of *L. monocytogenes* would be more strongly associated

112 with farm animals, agricultural activity, and urban areas than with natural forests and other wildlands
113 (22). During the first sampling occasion, 10% of sample sites were positive for *L. monocytogenes*
114 (Table 1). In addition, 13 samples of commercial bags of plant soil or compost were negative for *L.*
115 *monocytogenes*. In concordance with our hypothesis, the prevalence of *L. monocytogenes* was
116 significantly higher in urban areas and in areas associated with agriculture and livestock (agricultural
117 fields, grazelands and animal paths) than in forest/mountain areas and on footpaths (Fisher's exact
118 test $p<0.02$). Sampling locations classified as footpaths were generally from non-urban areas such as
119 woods or other areas used for hiking. While 14% of urban areas were positive for *L. monocytogenes*,
120 all samples from footpaths were negative for *L. monocytogenes*, and only 2% of the samples
121 collected in woodland or mountain areas were positive.

122 Detection of *L. monocytogenes* correlated with rain and sample humidity.

123 Previous studies have indicated that *L. monocytogenes* is more frequently isolated after recent
124 rainfall, irrigation, and flooding events (38, 39). In the present study, 271 out of 618 samples were
125 collected on days with rainfall and 347 samples on days with no rain within the previous 24 hours.
126 When collected on days with rain, 20% of samples were positive for *L. monocytogenes*, while on days
127 with no rain within the last 24 hours, only 3% of samples were positive. Thus, our data supports
128 previous studies suggesting that prevalence of *L. monocytogenes* is positively associated with rainfall
129 (Fisher's exact test $p=2\times10^{-12}$).

130 Upon sample collection, the humidity of the sampled material was categorized on a scale from 1
131 (completely dry) to 5 (liquid). Overall, the prevalence of *L. monocytogenes* in samples from the two
132 driest sample categories was 5.6% (4/70) and 5.7% (11/196), while it was 17% (27/164) and 14%
133 (12/86) in the more humid categories 3 and 4. The prevalence was significantly higher in the humid
134 samples (categories 3 and 4) than in the two driest sample categories (Fishers exact test, $p<0.02$).
135 The prevalence in liquid samples (category 5) was 10% (10/102). Among the samples collected in
136 urban environments, the sample humidity was not significantly associated ($p>0.05$) with the
137 prevalence of *L. monocytogenes*, with an overall prevalence of 10% in categories 1 and 2 (10/92), and
138 17% in categories 3 to 5 (14/85).

139 Persistent strains detected in rural and urban environments.

140 To examine whether environmental locations retained their status as *L. monocytogenes* positive or
141 negative over time, and whether the same clones were isolated repeatedly from the same location,
142 70 sites were subjected to one to three additional rounds of sampling the following years. In total,
143 115 *L. monocytogenes* isolates were collected in the current study (S1 Table). All isolates were

144 subjected to WGS, *in silico* MLST, and whole genome MLST (wgMLST) analysis. The distribution of
145 clones (CCs) among the identified isolates is presented in Figure 2.

146 Of the 44 sampling points positive for *L. monocytogenes* in the first round of sampling, 28 sites (64%)
147 were positive on at least one of the subsequent sampling occasions. Of the 26 initially negative sites,
148 five turned out positive during later sampling events (19%), and one of these was positive twice. In
149 total, 29 sampling sites were positive for *L. monocytogenes* more than once, and isolates belonging
150 to the same ST was collected repeatedly from seven sites (Table 2). In six cases, STs repeatedly
151 isolated from the same site were very closely related, with a maximum wgMLST allelic distance of 20.
152 When also adjacent or slightly more distant sampling sites (max 3 km) were included, a total of 14
153 clusters with genetic distances <20 were repeatedly collected from the same location over periods
154 ranging from 4 months to 3 years (S2 Table and S1 Text). When the commonly employed core
155 genome MLST (cgMLST) scheme described by Moura *et al*. (23) was employed, the isolates could not
156 be distinguished, except in one cluster with distances of 0 to 1 cgMLST alleles. Twelve clusters,
157 including two clusters each for CC91, CC11 (ST451) and CC37, represent clusters of highly similar
158 isolates with range 0 to 8 wgMLST allelic differences. Together, the results strongly indicate that *L.*
159 *monocytogenes* clones had persisted in the same environment or were repeatedly reintroduced
160 between sampling events in both rural and urban locations.

161 We also observed a case where a recent common contamination source was obscure: only 9 wgMLST
162 alleles (and no cgMLST alleles) separated a pair of CC6 isolates found 30 km and 3 years apart; one
163 isolate from a grazing pasture in Akershus county in 2020 and the other from soil by the root of a
164 tree in Oslo city centre in 2017.

165 Persistence and cross-contamination on Norwegian dairy farms.

166 In the next step, WGS was performed for a panel of 79 *L. monocytogenes* isolates collected from
167 Norwegian dairy farms (16). A total of 18 dairy herds from four different geographical areas within a
168 100 km radius from downtown Oslo (Figure 1B) had each been sampled four to six times between
169 August 2019 and July 2020. Out of the 556 analyzed samples, *L. monocytogenes* was detected in 12
170 milk filters (13% prevalence), 30 cattle feces samples (30%), 32 samples of cattle feed (silage or silage
171 mixture; 32%), and 5 teat swabs (5%). All bulk tank milk and teat milk samples were negative for *L.*
172 *monocytogenes*, and for one of the farms (farm 16), all 34 collected samples were negative. An
173 overview of the STs of the collected isolates (S1 Table) is presented in Table 3, and a phylogenetic
174 tree showing the genetic relationships between the individual isolates is shown in Figure 3.

175 Twelve clusters, each comprising two to four isolates, with pairwise genetic distances in the range 0
176 to 11 wgMLST alleles, were isolated from the same farm during repeated multiple visits, over periods

177 ranging from 2 to 10 months. These clusters involved 33 of the collected isolates and comprised ten
178 different CCs (S3 Table). These observations strongly support previous studies indicating that the
179 same *L. monocytogenes* clones can persist over time in individual cattle herds or farm environments
180 (15, 35, 36).

181 Out of 12 isolates from milk filters, four belonged to a persistent cluster, and one was closely related
182 to an isolate from a teat swab sample obtained on the same sampling occasion. When the same
183 clone was isolated from several sampling sites at the same farm, the pairwise genetic distances
184 separating milk filter isolates from fecal, feed or teat swab isolates ranged from 0 to 7 wgMLST allelic
185 differences (S3 Table). These links represent likely cross-contamination events where milk filters (and
186 consequently milk) have been contaminated with *L. monocytogenes* clones found in the farm
187 environment.

188 Detection of closely related isolates from different geographic areas.

189 In four cases, closely related isolates belonging to CC11 (ST451), CC226, and CC415 (ST394), were
190 collected from more than one dairy farm. The genetic differences between isolates from different
191 farms were somewhat greater than the diversity between isolates found on the same farm, with
192 between 9 and 20 pairwise wgMLST allelic differences. The number of cgMLST differences within
193 each cluster was 0 or 1 (S3 Table and S1 Text). These data indicate that farms located at different
194 geographical areas may host the same strain of *L. monocytogenes*.

195 Six clusters comprising *L. monocytogenes* from both dairy farms and isolates obtained from rural and
196 urban environments were detected. The genetic distances separating isolates from the two datasets
197 in these clusters ranged from 9 to 27 wgMLST allelic differences, and 0 or 1 cgMLST differences (S4
198 Table and S1 Text). The closest link was observed for a cluster of four CC37 isolates; two from grazing
199 land/pasture in the vicinity of Ås, and two from feed and teat swab samples obtained on two
200 different visits to farm 12, located about 50 km east of Ås. The two pairs of isolates were separated
201 by 9 to 14 wgMLST allelic differences and indistinguishable by cgMLST.

202 To further explore the occurrence of genetic links between Norwegian isolates from natural and
203 animal reservoirs, 24 of the 34 *L. monocytogenes* isolates collected from invading slugs (*Arion*
204 *vulgaris*) from garden and farm environments in Norway by Gismervik *et al*. in 2012 (21) were
205 subjected to WGS analysis (S1 Table). Interestingly, two pairs of slug isolates collected from different
206 geographic locations differed by only 2 (CC14) and 11 (CC1) wgMLST allelic differences. Furthermore,
207 five clusters with 10 to 21 wgMLST allelic differences comprised a slug isolate and one or more
208 isolates from either a rural/urban sampling site or from a dairy farm (S5 Table). The closest genetic
209 relationship concerned two CC1 isolates, in which a slug isolate from the west coast of Norway

210 (collected in 2012) showed only 10 wgMLST allelic differences compared to an isolate collected from
211 a street in a residential area in Oslo in 2017.

212 Thus, counting the previously mentioned pair of CC6 isolates collected in Akershus and Oslo, a total
213 of 17 close genetic links between isolates collected at relatively distant geographic areas in Norway
214 were detected in the set of 218 examined isolates. Presumably, not all clusters represent direct
215 epidemiological links, especially in cases where isolates were collected several years apart. The
216 observed genetic distances within the clusters; ≤21 wgMLST and ≤3 cgMLST allelic differences, are
217 within the thresholds often suggested as an appropriate guide for defining an outbreak cluster, which
218 is about 7-10 cgMLST differences (23, 40, 41) or about 20 single nucleotide polymorphisms (SNPs) in
219 SNP analyses (42, 43), which have a sensitivity comparable to wgMLST (34, 44).

220 Comparison with Norwegian clinical isolates.

221 The identification of close genetic links between isolates from different natural and animal-
222 associated sources without known connections led us to hypothesise that it would be possible to
223 identify clusters containing both environmental and clinical isolates with a similar level of genetic
224 relatedness. A dataset of Norwegian clinical isolates was identified, comprising 130 genomes from
225 2010-2015 (92% of all reported cases in these years) (45) and two genomes from 2018 (ST20 and
226 ST37), made publicly available by the European Centre for Disease Prevention and Control (ECDC)
227 and the Norwegian Institute of Public Health (NIPH), respectively. Sequencing data of sufficient
228 quality for wgMLST analysis was available for 111 of these isolates (S1 Table). An initial comparison
229 between the clinical isolates identified 15 pairs of isolates and nine larger clusters comprising 3 to 12
230 isolates showing genetic distances of ≤10 cgMLST allelic differences (S6 Table). Most clusters
231 contained isolates collected during a time span of several years and could potentially represent
232 listeriosis outbreaks or epidemiologically linked cases.

233 A wgMLST analysis showing the genetic relationships between isolates originating from rural and
234 urban environments, dairy farms, slugs, and clinical cases is shown in Figure 4. Nine clusters
235 contained clinical isolates differentiated from isolates sequenced in the current study by genetic
236 distances in the range 6 to 23 wgMLST allelic differences (0 to 7 cgMLST alleles) (S7 Table and S1
237 Text). The environmental *L. monocytogenes* closely related to clinical isolates comprised isolates
238 from soil samples from both urban and rural locations (belonging to CC4, CC7, CC11/ST451, CC220,
239 CC403, and CC415/ST394), three slug isolates obtained from garden and farm environments (CC7,
240 CC8, and CC9) and a group of CC11/ST451 isolates from dairy farms. The closest genetic link was
241 found between the single CC9 slug isolate (from 2012) and a clinical isolate from 2015, differentiated
242 by only 6 wgMLST alleles (and 0 cgMLST alleles). The analysis shows that *L. monocytogenes* isolates

243 that are genetically very closely related to clinical isolates can be detected in various natural and
244 agricultural environments, even when isolates are collected across timespans ranging several years.

245 Comparison of prevalence and diversity of MLST clones from different sources.

246 Most isolates from natural and agricultural environments belonged to *L. monocytogenes* lineage II,
247 comprising 89%, 94%, and 68% of isolates from rural/urban environments, dairy farms, and slugs,
248 respectively (Figure 5A). The remaining isolates belonged to *L. monocytogenes* lineage I, as lineage III
249 or IV isolates were not detected in the current study. The predominant clones among the rural/urban
250 isolates were CC91 (15%), CC19/ST398 (13%), CC37 (10%), and CC11/ST451 (9%). No specific niches
251 were found for these clones as isolates were spread geographically (3-5 counties), found in 3-5
252 different habitats/areas and in a range of humidity and weather conditions. CC91 appeared most
253 ubiquitous, as it was isolated from five different counties, from different sample types (soil, sand,
254 vegetation, and feces), from five different areas (agricultural fields, urban area, beach, grazeland and
255 forest), during all seasons and from all categories of humidity. CC11/ST451, CC91, and CC37 were
256 also the most frequently isolated clonal groups at the dairy farms (18%, 15%, and 11%, respectively);
257 each detected on seven different farms. Among the slug isolates, the most common clones were CC1
258 (15) and CC91 (12%) (21). A survey of previous studies indicated that CC1, CC7 and CC37 were the
259 clones most commonly detected in various natural and farm environments (S8 Table).

260 Among the examined Norwegian clinical isolates, CC7 was the most prevalent clonal group,
261 accounting for 23% (n=30) of the reported listeriosis cases, followed by CC121 (13%), CC8 (8%), and
262 CC1 (6%) (Figure 5B). In contrast to that observed in many other countries (27, 46), lineage I isolates
263 comprised a minority of the clinical isolates in this dataset (20%). The high prevalence of CC121
264 among the clinical isolates was unexpected, as this clone is commonly regarded as hypovirulent due
265 to the frequent occurrence of premature stop codons (PMSC) in the gene encoding the virulence
266 factor internalin A (*inlA*) (27), a characteristic also shared by the Norwegian CC121 clinical isolates.
267 Interestingly, the single *L. monocytogenes* CC121 isolated in the current study; MF7617 from soil at
268 the edge of a university campus pond in Ås, had an intact and presumably functional copy of *inlA*.
269 This isolate was only distantly related to the clinical CC121 isolates, separated by 195 wgMLST alleles
270 from the nearest clinical isolate. In contrast to CC121, the other three most commonly detected CCs
271 among the Norwegian clinical isolates – CC1, CC7 and CC8 – were relatively common also among the
272 isolates from natural and agricultural environments, with each CC having an average prevalence of
273 between 6% and 7.5% in the rural/urban, dairy farms, and slug isolate datasets (Figure 5A).

274 Since listeriosis is primarily acquired from food, the frequency distribution of CCs for *L.*
275 *monocytogenes* from Norwegian food processing industry (Figure 5C) was estimated from previous

276 work encompassing 680 isolates from five meat and four salmon processing plants, collected during
277 2011-2015 (47). The prevalence of lineage I isolates was <1% among the food processing industry
278 isolates; represented by only two CC2 isolates from the meat industry. In meat processing
279 environments, CC9 was by far the most prevalent clonal group, representing 70% of isolates. It must,
280 however, be noted that most of the collected isolates were from two intensively sampled processing
281 plants (34). One slug isolate and three clinical isolates (2011, 2012, 2015) but none from dairy farms
282 or samples from rural and urban environments belonged to CC9. In salmon processing environments,
283 CC14 was most prevalent (25%), followed by CC121 (22%), CC7 (ST7, ST732 and ST995; 18%), and CC8
284 (ST8 and ST551; 14%). CC14/ST14 was only represented by two clinical and two slug isolates, and not
285 detected among isolates from rural/urban environments or dairy farms. The latter three were among
286 the four most prevalent CCs among the Norwegian clinical isolates.

287 To examine the diversity of the most commonly detected *L. monocytogenes* clonal groups from
288 Norwegian natural environments in an international context, a representative subset of reference
289 genomes belonging to ST37, ST91, and ST451 were selected for comparative analysis using cgMLST
290 (Figure 6). Of the >6000 examined publicly available genomes, 243 belonged to one of the three
291 relevant STs. For ST91, a limited number of international reference sequences were available, and
292 nearly 60% of the analysed isolates were Norwegian. This suggest that this ST could be more
293 prevalent in Norway relative to in many other countries. For ST37, only limited clustering of
294 Norwegian isolates relative to the international isolates was observed. In contrast, for ST91 and
295 ST451, the Norwegian isolates appeared to cluster together with isolates from other countries,
296 indicating that they represent internationally dispersed clones.

297 Discussion

298 It has long been acknowledged that *L. monocytogenes* clones predominating among human clinical
299 isolates differ from those that dominate in food (23, 24, 26, 46, 48, 49), and that persistent clones of
300 *L. monocytogenes* may establish in food processing environments (2, 33, 34). Here, we show that the
301 most ubiquitous clones found in soil and other natural and animal ecosystems (CC91, CC11, and
302 CC37) are distinct from clones predominating among both clinical and food isolates, and that *L.*
303 *monocytogenes* may persist and spread also in urban and rural areas, grazeland, agricultural fields,
304 and farm environments. The correspondence of major CCs was high for the three examined sets of
305 environmental isolates (rural/urban, dairy farms, and slugs). CC37 appeared to be exceptionally
306 widespread in natural environments, and was isolated from nine different counties and a wide
307 variety of habitats. It was also found to persist for years both at a farm and on a bike path in the
308 capital of Norway. The ubiquity of this clone is also reflected by its detection in a large proportion of

309 other studies investigating the identity of *L. monocytogenes* clones from natural and animal
310 reservoirs, including wildlife, forest areas and farms (5, 6, 13, 18, 21, 25, 29-31, 50).

311 The current study identified close genetic relationships between environmental isolates of *L.*
312 *monocytogenes* collected from geographically and temporally unrelated sources, despite a relatively
313 low number of analysed isolates. Although fixed clustering thresholds for defining outbreak clusters
314 are controversial (37, 51, 52), the genetic distances in the observed clusters were well within the
315 thresholds used to guide outbreak analyses (23, 40-43). In the majority of observed clusters, isolates
316 with no known likely association were indistinguishable using cgMLST analysis, which is the method
317 currently employed for surveillance of *L. monocytogenes* by many laboratories including the
318 Norwegian Institute of Public Health (53). This finding underscores the need for careful consideration
319 of additional evidence such as epidemiological data, traceback evidence, and phylogenetic tree
320 topology as part of WGS-based surveillance and outbreak investigations (54). Ideally, evaluation of
321 possible epidemiological links should consider the occurrence of closely related strains in the whole
322 food chain, including external contamination sources in urban and natural environments (52).
323 Currently, a lack of published genomic data on *L. monocytogenes* from various sources is a barrier for
324 effective management of this pathogen, both for public health authorities and for industrial actors.

325 During the last decade (2010-2020), an average of 24 yearly listeriosis cases have been reported in
326 Norway, and most of them (80%) were domestically acquired (http://www.msis.no/). The implicated
327 food is rarely identified. Only two outbreaks have been publicly reported during this period, both
328 associated with traditional fermented fish (rakfisk); one in 2013 (ST802, 4 cases in Norway) (55), and
329 one during the winter of 2018-2019 (ST20, 12 cases in Norway, one in Sweden (56)). A predominance
330 of lineage II was observed among the Norwegian clinical isolates – comprising 80% of isolates during
331 the years 2010-2015; an increase relative to the 56% observed during 1992-2005 (57). A distinct
332 feature among Norwegian clinical isolates was the large proportion of CC121 isolates lacking
333 functional internalin A. The only CC121 isolate collected from a natural environment did not have an
334 *inlA* PMSC, supporting the hypothesis that *inlA* mutations may constitute an adaptation to food
335 industry environments (58). The relatively high proportion of cases caused by clones of a
336 hypovirulent strain in Norway could potentially be linked to national consumption and storage
337 practices leading to sporadic ingestion of large numbers of the pathogen among high risk groups.

338 Worldwide, the hypervirulent clones CC1 and CC4 are significantly more prevalent among clinical
339 isolates than food isolates (27, 59, 60). Together, these two CCs constituted 8% of Norwegian clinical
340 isolates and 11% of the isolates from natural and farm environments. CC1 and CC4 also appear to be
341 prevalent in natural environments in other countries (13, 29). They were, however, not detected in a

342 study of *L. monocytogenes* in nine Norwegian food processing plants (47). Although at least 80% of
343 meat, cheese and fish consumed in Norway is produced domestically (61), imported processed foods
344 remain a potential source of infections. Notably, however, 45% of Norwegian households report that
345 they hunt, fish or collect bivalve molluscs, and about half of the population grow their own
346 vegetables, herbs or fruit, and collect berries in the wild (62). Furthermore, the current study
347 identified clusters comprising closely related isolates from both clinical sources and natural
348 environments, despite comparing temporally non-overlapping sets of isolates. Together, these
349 observations suggest that the relative contribution of industrially processed foods to listeriosis
350 infections could be lower in Norway than in other countries.

351 Materials and Methods

352 Sampling of *L. monocytogenes* from rural and urban environments.

353 Samples were taken to cover what was hypothesized as hot spots and cold spots for *L.*
354 *monocytogenes* in the outer environment. The sampling plan was designed to cover different
355 geographical regions of Norway and areas hypotesized to have respectively high (urban areas,
356 grazeland, animal paths and areas near food processing factories) and low (forests and mountain
357 areas, agricultural fields, beaches and sandbanks) occurrence of *L. monocytogenes*. Samples
358 classified as footpaths were generally from non-urban areas in woods or other areas used for hiking,
359 but separately categorized as we considered footpaths to be associated with human activities to a
360 greater extent than more pristine woodland or mountain areas. A detailed sampling scheme was
361 prepared and convenience sampling was performed by people living in or travelling to different areas
362 to cover Norway geographically and to get detailed results from specific areas (e.g., gardens) and
363 local information. The sampling was performed by trained microbiologists informed about the
364 objective of the study and which types of sites that should be sampled. When possible, several
365 different suspected hot and cold spots were sampled in the same geographical area, e.g., grazeland
366 and a forest nearby where the cattle did not have access. Sampling was performed year around
367 except for winter. For a selection of sampling sites, sampling was repeated once or more over a
368 period of three years.

369 The environmental samples (soil, sand, mud, decaying vegetation, surface water, animal dung, etc.)
370 were collected in sterile 50 mL Nunc tubes. All sampling locations were photographed, and GPS
371 coordinates, sample content, habitat/area, and weather conditions were recorded at the time of
372 sample collection. Specific information about the sample was also noted, such as which animals the
373 area was frequently exposed to (e.g., cattle, deer, sheep, doves) and local information (e.g., popular
374 area for hiking). The humidity of the collected samples was assessed on a scale from 1 to 5, ranging

375 from completely dry (1) to liquid (5). Samples were stored at 4°C for up to a week before processing,
376 and analysed according to ISO 11290-1 (63) with selective enrichment in half-Fraser and Fraser broth
377 (Oxoid) and final plating on RAPID'*L.mono* agar (Bio-Rad).

378 Whole genome sequencing.

379 For each *L. monocytogenes* isolate from rural/urban environments or from *Arion vulgaris* slugs (21), a
380 single colony was picked, inoculated in 5 mL brain heart infusion broth, and grown at 37°C overnight.
381 Culture samples (1 mL) were lysed using Lysing Matrix B and a FastPrep instrument (both MP
382 Biomedicals) and genomic DNA was isolated using the DNeasy Blood and Tissue Kit (Qiagen).
383 Libraries for genome sequencing were prepared using the Nextera XT DNA Sample Preparation Kit
384 (Illumina) and sequenced using 2×300 bp reads on a MiSeq instrument (Illumina). Raw reads were
385 filtered on q15 and trimmed of adaptors before *de novo* genome assembly was performed using
386 SPAdes v3.10.0 or v.3.13.0 (64) with the careful option and six k-mer sizes (21, 33, 55, 77, 99, and
387 127). Contigs with sizes of <500 bp and with coverage of <5 were filtered out.

388 Colonies from the *L. monocytogenes* isolates from dairy farms (16) were inoculated in 20 mL tryptone
389 soy broth and incubated at 37°C for 24 h before 1 mL was pelleted and DNA extracted using the
390 DNeasy Blood & Tissue kit (Qiagen). Libraries for genome sequencing were prepared using the
391 NEBNext Ultra DNA Library Prep Kit (New England Biolabs) with random fragmentation to 350 bp and
392 sequencing of 2×150 bp on a NovaSeq 6000 S4 flow cell (Illumina). *De novo* genome assembly was
393 performed using SPAdes v.3.14.1, incorporated in the software tool Shovill available at
394 https://github.com/tseemann/shovill. Shovill also performed adaptor trimming using Trimmomatic,
395 corrected assembly errors, and removed contigs with sizes <500 bp and coverage <2.

396 Phylogenetic analyses.

397 Classical MLST analysis followed the MLST scheme described by Ragon *et al*. (65) and the database
398 maintained at the Institute Pasteur's *L. monocytogenes* online MLST repository
399 (http://bigsdb.web.pasteur.fr/listeria/listeria.html). *In silico* MLST typing was performed for raw
400 sequencing data using the program available at https://bitbucket.org/genomicepidemiology/mlst
401 (66), and for genome assemblies using the program available at https://github.com/tseemann/mlst.
402 CCs are defined as groups of ST profiles sharing at least six of seven genes with at least one other
403 member of the group, except for CC14, which is divided into CC14 represented by ST14 and ST399 in
404 the current work, and CC91 represented by ST91, as isolates belonging to these two groups do not
405 cluster together in phylogenetic analyses of *L. monocytogenes* populations (27). The wgMLST analysis
406 was performed using a whole-genome schema containing 4797 coding loci from the *L.*
407 *monocytogenes* pan-genome and the assembly-based BLAST approach, implemented in BioNumerics

408 7.6 (http://www.applied-maths.com/news/listeria-monocytogenes-whole-genome-sequence-
409 typing). The cgMLST analysis was performed using the scheme described by Moura *et al*. (23), and
410 results were either extracted from the wgMLST data or obtained by sequence query at the BIGSdb-
411 *Lm* web server (https://bigsdb.pasteur.fr/listeria/listeria.html). Minimum spanning trees were
412 constructed using BioNumerics based on the categorical differences in the allelic cgMLST or wgMLST
413 profiles for each isolate. Loci with no allele calls were not considered in the pairwise comparison
414 between two genomes.

415 Publicly available genomes.

416 Available genomes of clinical isolates from human patients in Norway were identified by searching
417 the NCBI Pathogen Detection database (https://www.ncbi.nlm.nih.gov/pathogens) on the
418 30.08.2021. Available raw sequencing data from NCBI BioProjects submitted by ECDC and NPIH;
419 PRJEB26061 (45) and PRJEB25848, were subjected to *de novo* genome assembly as described for
420 isolates from rural/urban environments. *In silico* MLST genotyping was successful for all genomes
421 except one of the genomes published by the ECDC, and wgMLST analysis was successful for all except
422 21 of the ECDC genomes.

423 Reference genomes included in the cgMLST analysis of ST37, ST91 and ST451 genomes were
424 identified from the following selected sources on the 27.08.2021: (i) cgMLST profiles from the
425 BIGSdb-*Lm* database (http://bigsdb.pasteur.fr/listeria/listeria.html); 15 genomes belonging to
426 relevant STs were identified. (ii) Raw WGS data from the ListAdapt project
427 (https://onehealthejp.eu/jrp-listadapt/) comprising 1552 genomes (BioProject PRJEB38828). *De novo*
428 genome assembly was performed for the 165 genomes of relevant STs. (iii) Genome assemblies from
429 NCBI GenBank; among the 3926 *L. monocytogenes* genomes, 63 genomes belonged to the relvant
430 STs.

431 Data availability.

432 Data from this whole-genome shotgun project has been deposited at the NCBI GenBank database
433 under BioProjects PRJNA689486, PRJNA744724, and PRJNA689487. For GenBank and Sequence Read
434 Archive (SRA) accession numbers, see S1 Table. The assemblies were annotated using the NCBI
435 Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) server
436 (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/).

437 Acknowledgments

438 The authors wish to thank Merete Rusås Jensen, Anette Wold Åsli, Janina Berg, Tove Maugesten at
439 Nofima for excellent technical assistance. We also sincerely thank collaborators in the food and

salmon processing industry, colleagues at Nofima (Runar Gjerp Solstad, Rasmus Karstad, Halvor Nygaard, Kristin Skei Nerdal) and Arild Hugo Solstad for contributing to sample collection. We thank Ann-Katrin Llarena (Faculty of Veterinary Medicine, Norwegian University of Life Sciences) for help during genome assembly of *L. monocytogenes* dairy farm isolates. We thank Ida Skaar (Norwegian Veterinary Institute) for providing the *L. monocytogenes* isolates from slugs (21). We thank the team of curators of the Institut Pasteur MLST system (Paris, France) for importing novel alleles, profiles and/or isolates at http://bigsdb.pasteur.fr/listeria/. This work was funded by the Norwegian Agriculture and Food Industry Research Funds, grant numbers 262306 and 314743.

References

1. Lopez-Valladares G, Danielsson-Tham ML, Tham W. 2018. Implicated food products for listeriosis and changes in serovars of *Listeria monocytogenes* affecting humans in recent decades. Foodborne Pathog Dis 15:387-397.

2. Ferreira V, Wiedmann M, Teixeira P, Stasiewicz MJ. 2014. *Listeria monocytogenes* persistence in food-associated environments: Epidemiology, strain characteristics, and implications for public health. J Food Prot 77:150-170.

3. Liao J, Guo X, Weller DL, Pollak S, Buckley DH, Wiedmann M, Cordero OX. 2021. Nationwide genomic atlas of soil-dwelling *Listeria* reveals effects of selection and population ecology on pangenome evolution. Nat Microbiol 6:1021-1030.

4. Vivant AL, Garmyn D, Piveteau P. 2013. *Listeria monocytogenes*, a down-to-earth pathogen. Front Cell Infect Microbiol 3:87.

5. Linke K, Rückerl I, Brugger K, Karpiskova R, Walland J, Muri-Klinger S, Tichy A, Wagner M, Stessl B. 2014. Reservoirs of *Listeria* species in three environmental ecosystems. Appl Environ Microbiol 80:5583-5592.

6. Terentjeva M, Šteingolde Z, Meistere I, Elferts D, Avsejenko J, Streikiša M, Gradovska S, Alksne L, Kibilds J, Bērziņš A. 2021. Prevalence, genetic diversity and factors associated with distribution of *Listeria monocytogenes* and other *Listeria* spp. in cattle farms in Latvia. Pathogens 10:851.

7. Weller D, Wiedmann M, Strawn LK. 2015. Spatial and temporal factors associated with an increased prevalence of *Listeria monocytogenes* in spinach fields in New York State. Appl Environ Microbiol 81:6059-6069.

469 8. Ivanek R, Grohn YT, Wells MT, Lembo AJ, Sauders BD, Wiedmann M. 2009. Modeling of spatially
470 referenced environmental and meteorological factors influencing the probability of *Listeria*
471 species isolation from natural environments. Appl Environ Microbiol 75:5893-5909.

472 9. Fenlon DR, Wilson J, Donachie W. 1996. The incidence and level of *Listeria monocytogenes*
473 contamination of food sources at primary production and initial processing. J Appl Bacteriol
474 81:641-650.

475 10. Lyautey E, Lapen DR, Wilkes G, McCleary K, Pagotto F, Tyler K, Hartmann A, Piveteau P, Rieu A,
476 Robertson WJ, Medeiros DT, Edge TA, Gannon V, Topp E. 2007. Distribution and characteristics
477 of *Listeria monocytogenes* isolates from surface waters of the South Nation River watershed,
478 Ontario, Canada. Appl Environ Microbiol 73:5401-5410.

479 11. Stea EC, Purdue LM, Jamieson RC, Yost CK, Hansen LT. 2015. Comparison of the prevalences and
480 diversities of *Listeria* species and *Listeria monocytogenes* in an urban and a rural agricultural
481 watershed. Appl Environ Microbiol 81:3812-3822.

482 12. Weller D, Belias A, Green H, Roof S, Wiedmann M. 2020. Landscape, water quality, and weather
483 factors associated with an increased likelihood of foodborne pathogen contamination of New
484 York streams used to source water for produce production. Front Sustain Food Syst 3:124.

485 13. Raschle S, Stephan R, Stevens MJA, Cernela N, Zurfluh K, Muchaamba F, Nuesch-Inderbinen M.
486 2021. Environmental dissemination of pathogenic *Listeria monocytogenes* in flowing surface
487 waters in Switzerland. Sci Rep 11:9066.

488 14. Nightingale KK, Schukken YH, Nightingale CR, Fortes ED, Ho AJ, Her Z, Grohn YT, McDonough PL,
489 Wiedmann M. 2004. Ecology and transmission of *Listeria monocytogenes* infecting ruminants
490 and in the farm environment. Appl Environ Microbiol 70:4458-4467.

491 15. Castro H, Jaakkonen A, Hakkinen M, Korkeala H, Lindström M. 2018. Occurrence, persistence,
492 and contamination routes of *Listeria monocytogenes* genotypes on three Finnish dairy cattle
493 farms: a longitudinal study. Appl Environ Microbiol 84:e02000-17.

494 16. Idland L, Granquist EG, Aspholm M, Lindbäck T. 2021. The occurrence of *Campylobacter* spp.,
495 *Listeria monocytogenes* and Shiga toxin-producing *Escherichia coli* in Norwegian dairy cattle
496 farms; a comparison between free stall and tie stall housing systems. agriRxiv
497 http://doi.org/10.31220/agriRxiv.2021.00091.

498 17. Fox E, O'Mahony T, Clancy M, Dempsey R, O'Brien M, Jordan K. 2009. *Listeria monocytogenes* in
499 the Irish dairy farm environment. J Food Prot 72:1450-1456.

500 18. Sauvala M, Woivalin E, Kivisto R, Laukkanen-Ninios R, Laaksonen S, Stephan R, Fredriksson-
501 Ahomaa M. 2021. Hunted game birds - Carriers of foodborne pathogens. Food Microbiol
502 98:103768.

503 19. Wang Y, Lu L, Lan R, Salazar JK, Liu J, Xu J, Ye C. 2017. Isolation and characterization of *Listeria*
504 species from rodents in natural environments in China. Emerg Microbes Infect 6:e44.

505 20. Hellström S, Kiviniemi K, Autio T, Korkeala H. 2008. *Listeria monocytogenes* is common in wild
506 birds in Helsinki region and genotypes are frequently similar with those found along the food
507 chain. J Appl Microbiol 104:883-888.

508 21. Gismervik K, Aspholm M, Rørvik LM, Bruheim T, Andersen A, Skaar I. 2015. Invading slugs (*Arion*
509 *vulgaris*) can be vectors for *Listeria monocytogenes*. J Appl Microbiol 118:809-816.

510 22. Sauders BD, Overdevest J, Fortes E, Windham K, Schukken Y, Lembo A, Wiedmann M. 2012.
511 Diversity of *Listeria* species in urban and natural environments. Appl Environ Microbiol 78:4420-
512 4433.

513 23. Moura A, Criscuolo A, Pouseele H, Maury MM, Leclercq A, Tarr C, Björkman JT, Dallman T,
514 Reimer A, Enouf V, Larsonneur E, Carleton H, Bracq-Dieye H, Katz LS, Jones L, Touchon M,
515 Tourdjman M, Walker M, Stroika S, Cantinelli T, Chenal-Francisque V, Kucerova Z, Rocha EPC,
516 Nadon C, Grant K, Nielsen EM, Pot B, Gerner-Smidt P, Lecuit M, Brisse S. 2016. Whole genome-
517 based population biology and epidemiological surveillance of *Listeria monocytogenes*. Nat
518 Microbiol 2:16185.

519 24. Painset A, Björkman JT, Kiil K, Guillier L, Mariet JF, Félix B, Amar C, Rotariu O, Roussel S, Perez-
520 Reche F, Brisse S, Moura A, Lecuit M, Forbes K, Strachan N, Grant K, Møller-Nielsen E, Dallman
521 TJ. 2019. LiSEQ - whole-genome sequencing of a cross-sectional survey of *Listeria*
522 *monocytogenes* in ready-to-eat foods and human clinical cases in Europe. Microb Genom
523 5:000257.

524 25. Félix B, Feurer C, Maillet A, Guillier L, Boscher E, Kerouanton A, Denis M, Roussel S. 2018.
525 Population genetic structure of *Listeria monocytogenes* strains isolated from the pig and pork
526 production chain in France. Front Microbiol 9:684.

527 26. Maury MM, Bracq-Dieye H, Huang L, Vales G, Lavina M, Thouvenot P, Disson O, Leclercq A,
528 Brisse S, Lecuit M. 2019. Hypervirulent *Listeria monocytogenes* clones' adaptation to mammalian
529 gut accounts for their association with dairy products. Nat Commun 10:2488.

530 27. Maury MM, Tsai YH, Charlier C, Touchon M, Chenal-Francisque V, Leclercq A, Criscuolo A,
531 Gaultier C, Roussel S, Brisabois A, Disson O, Rocha EPC, Brisse S, Lecuit M. 2016. Uncovering
532 *Listeria monocytogenes* hypervirulence by harnessing its biodiversity. Nat Genet 48:308-313.

533 28. Cabal A, Pietzka A, Huhulescu S, Allerberger F, Ruppitsch W, Schmid D. 2019. Isolate-based
534 surveillance of *Listeria monocytogenes* by whole genome sequencing in Austria. Front Microbiol
535 10:2282.

536 29. Papić B, Pate M, Félix B, Kušar D. 2019. Genetic diversity of *Listeria monocytogenes* strains in
537 ruminant abortion and rhombencephalitis cases in comparison with the natural environment.
538 BMC Microbiol 19:299.

539 30. Fredriksson-Ahomaa M, London L, Skrzypczak T, Kantala T, Laamanen I, Biström M, Maunula L,
540 Gadd T. 2020. Foodborne zoonoses common in hunted wild boars. Ecohealth 17:512-522.

541 31. Sauvala M, Laaksonen S, Laukkanen-Ninios R, Jalava K, Stephan R, Fredriksson-Ahomaa M. 2019.
542 Microbial contamination of moose (*Alces alces*) and white-tailed deer (*Odocoileus virginianus*)
543 carcasses harvested by hunters. Food Microbiol 78:82-88.

544 32. Castro H, Douillard FP, Korkeala H, Lindström M. 2021. Mobile elements harboring heavy metal
545 and bacitracin resistance genes are common among *Listeria monocytogenes* strains persisting on
546 dairy farms. mSphere 6:e0038321.

547 33. Carpentier B, Cerf O. 2011. Review - Persistence of *Listeria monocytogenes* in food industry
548 equipment and premises. Int J Food Microbiol 145:1-8.

549 34. Fagerlund A, Langsrud S, Møretrø T. 2020. In-depth longitudinal study of *Listeria monocytogenes*
550 ST9 isolates from the meat processing industry: Resolving diversity and transmission patterns
551 using whole-genome sequencing. Appl Environ Microbiol 86:e00579-20.

552 35. Haley BJ, Sonnier J, Schukken YH, Karns JS, Van Kessel JA. 2015. Diversity of *Listeria*
553 *monocytogenes* within a U.S. dairy herd, 2004-2010. Foodborne Pathog Dis 12:844-850.

554 36. Latorre AA, Van Kessel JA, Karns JS, Zurakowski MJ, Pradhan AK, Zadoks RN, Boor KJ, Schukken
555 YH. 2009. Molecular ecology of *Listeria monocytogenes*: evidence for a reservoir in milking
556 equipment on a dairy farm. Appl Environ Microbiol 75:1315-1323.

557 37. Pightling AW, Pettengill JB, Luo Y, Baugher JD, Rand H, Strain E. 2018. Interpreting whole-
558 genome sequence analyses of foodborne bacteria for regulatory applications and outbreak
559 investigations. Front Microbiol 9:1482.

560 38. Weller D, Wiedmann M, Strawn LK. 2015. Spatial and temporal factors associated with an
561 increased prevalence of *Listeria monocytogenes* in spinach fields in New York State. Appl Environ
562 Microbiol 81:6059-6069.

563 39. Ivanek R, Grohn YT, Wells MT, Lembo AJ, Jr., Sauders BD, Wiedmann M. 2009. Modeling of
564 spatially referenced environmental and meteorological factors influencing the probability of
565 *Listeria* species isolation from natural environments. Appl Environ Microbiol 75:5893-5909.

566 40. Zamudio R, Haigh RD, Ralph JD, De Ste Croix M, Tasara T, Zurfluh K, Kwun MJ, Millard AD,
567 Bentley SD, Croucher NJ, Stephan R, Oggioni MR. 2020. Lineage-specific evolution and gene flow
568 in *Listeria monocytogenes* are independent of bacteriophages. Environ Microbiol 22:5058-5072.

569 41. Ruppitsch W, Pietzka A, Prior K, Bletz S, Fernandez HL, Allerberger F, Harmsen D, Mellmann A.
570 2015. Defining and evaluating a core genome multilocus sequence typing scheme for whole-
571 genome sequence-based typing of *Listeria monocytogenes*. J Clin Microbiol 53:2869-2876.

572 42. Wang Y, Pettengill JB, Pightling A, Timme R, Allard M, Strain E, Rand H. 2018. Genetic diversity of
573 *Salmonella* and *Listeria* isolates from food facilities. J Food Prot 81:2082-2089.

574 43. Allard MW, Strain E, Rand H, Melka D, Correll WA, Hintz L, Stevens E, Timme R, Lomonaco S,
575 Chen Y, Musser SM, Brown EW. 2019. Whole genome sequencing uses for foodborne
576 contamination and compliance: Discovery of an emerging contamination event in an ice cream
577 facility using whole genome sequencing. Infect Genet Evol 73:214-220.

578 44. Jagadeesan B, Baert L, Wiedmann M, Orsi RH. 2019. Comparative analysis of tools and
579 approaches for source tracking *Listeria monocytogenes* in a food facility using whole-genome
580 sequence data. Front Microbiol 10:947.

581 45. Van Walle I, Bjorkman JT, Cormican M, Dallman T, Mossong J, Moura A, Pietzka A, Ruppitsch W,
582 Takkinen J, European Listeria Wgs Typing G. 2018. Retrospective validation of whole genome
583 sequencing-enhanced surveillance of listeriosis in Europe, 2010 to 2015. Euro Surveill
584 23:1700798.

585 46. Chenal-Francisque V, Lopez J, Cantinelli T, Caro V, Tran C, Leclercq A, Lecuit M, Brisse S. 2011.
586 Worldwide distribution of major clones of *Listeria monocytogenes*. Emerg Infect Dis 17:1110-
587 1112.

588 47. Møretrø T, Schirmer BCT, Heir E, Fagerlund A, Hjemli P, Langsrud S. 2017. Tolerance to
589 quaternary ammonium compound disinfectants may enhance growth of *Listeria monocytogenes*
590 in the food industry. Int J Food Microbiol 241:215-224.

591 48. Henri C, Leekitcharoenphon P, Carleton HA, Radomski N, Kaas RS, Mariet JF, Felten A, Aarestrup
592 FM, Gerner Smidt P, Roussel S, Guillier L, Mistou MY, Hendriksen RS. 2017. An assessment of
593 different genomic approaches for inferring phylogeny of *Listeria monocytogenes*. Front
594 Microbiol 8:2351.

595 49. Gray MJ, Zadoks RN, Fortes ED, Dogan B, Cai S, Chen Y, Scott VN, Gombas DE, Boor KJ,
596 Wiedmann M. 2004. *Listeria monocytogenes* isolates from foods and humans form distinct but
597 overlapping populations. Appl Environ Microbiol 70:5833-5841.

598 50. Kim SW, Haendiges J, Keller EN, Myers R, Kim A, Lombard JE, Karns JS, Van Kessel JAS, Haley BJ.
599 2018. Genetic diversity and virulence profiles of *Listeria monocytogenes* recovered from bulk
600 tank milk, milk filters, and milking equipment from dairies in the United States (2002 to 2014).
601 PLoS One 13:e0197053.

602 51. Schürch AC, Arredondo-Alonso S, Willems RJL, Goering RV. 2018. Whole genome sequencing
603 options for bacterial strain typing and epidemiologic analysis based on single nucleotide
604 polymorphism versus gene-by-gene-based approaches. Clin Microbiol Infect 24:350-354.

605 52. Gerner-Smidt P, Besser J, Concepción-Acevedo J, Folster JP, Huffman J, Joseph LA, Kucerova Z,
606 Nichols MC, Schwensohn CA, Tolar B. 2019. Whole genome sequencing: Bridging one-health
607 surveillance of foodborne diseases. Front Public Health 7:172.

608 53. Løvdal T, Brandal LT, Sundaram AYM, Naseer U, Roth B, Lunestad BT. 2021. Small-scale
609 comparative genomic analysis of *Listeria monocytogenes* isolated from environments of salmon
610 processing plants and human cases in Norway. Hygiene 1:43–55.

611 54. Brown E, Dessai U, McGarry S, Gerner-Smidt P. 2019. Use of whole-genome sequencing for food
612 safety and public health in the United States. Foodborne Pathog Dis 16:441-450.

613 55. Nielsen EM, Björkman JT, Kiil K, Grant K, Dallman T, Painset A, Amar C, Roussel S, Guillier L, Félix
614 B, Rotariu O, Perez-Reche F, Forbes K, Strachan N. 2017. Closing gaps for performing a risk
615 assessment on *Listeria monocytogenes* in ready-to-eat (RTE) foods: activity 3, the comparison of
616 isolates from different compartments along the food chain, and from humans using whole
617 genome sequencing (WGS) analysis. EFSA supporting publication EN-1151:170.

618 56. Public Health Agency of Sweden (Folkhälsomyndigheten). 2020. Listerios 2019. Sammanfattning
619 och bedömning. https://www.folkhalsomyndigheten.se/folkhalsorapportering-statistik/statistik-
620 a-o/sjukdomsstatistik/listeriainfektion/kommentarer-och-specialstatistik/2019/. Accessed
621 07.09.2021.

622 57. Rosef O, Klæboe H, Paulauskas A, Ambrasiene D. 2012. Diversity of *Listeria monocytogenes*
623 isolated from humans, food, and environmental sources in Norway. Vet Med Zoot 59:71-79.

624 58. Harrand AS, Jagadeesan B, Baert L, Wiedmann M, Orsi RH. 2020. Evolution of *Listeria*
625 *monocytogenes* in a food processing plant involves limited single-nucleotide substitutions but
626 considerable diversification by gain and loss of prophages. Appl Environ Microbiol 86:e02493-19.

627 59. Lee S, Chen Y, Gorski L, Ward TJ, Osborne J, Kathariou S. 2018. *Listeria monocytogenes* source
628 distribution analysis indicates regional heterogeneity and ecological niche preference among
629 serotype 4b clones. mBio 9:e00396-18.

630 60. Kwong JC, Mercoulia K, Tomita T, Easton M, Li HY, Bulach DM, Stinear TP, Seemann T, Howden
631 BP. 2016. Prospective whole-genome sequencing enhances national surveillance of *Listeria*
632 *monocytogenes*. J Clin Microbiol 54:333-342.

633 61. Helsedirektoratet. 2021. Utviklingen i norsk kosthold 2020. Matforsyningsstatistikk [*online*].
634 https://www.helsedirektoratet.no/rapporter/utviklingen-i-norsk-kosthold. Accessed 21.10.2121.

635 62. Bugge AB. 2015. Moralities of food and meals – how to eat properly [*online*]. Published by:
636 Statens Institutt for Forbruksforskning (SIFO), SIFO-report no. 3-2015.
637 https://hdl.handle.net/20.500.12199/927.

638 63. International Organization of Standardization. 2017. Microbiology of the food chain - Horizontal
639 method for the detection and enumeration of *Listeria monocytogenes* and of *Listeria* spp. —
640 Part 1: Detection method (ISO 11290-1). https://www.iso.org/standard/60313.html.

641 64. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Son
642 P, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012.
643 SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J
644 Comput Biol 19:455-477.

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

647 66. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Ponten
648 T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total-genome-
649 sequenced bacteria. J Clin Microbiol 50:1355-1361.

650 # Tables

651 *Table 1: Prevalence of L. monocytogenes in rural and urban environments*

Habitat or sampling area	No. of collected isolates	Prevalence of *L. monocytogenes* (%)
Grazeland or animal path	85	21 %
Urban or residential area	177	14 %
Agricultural field	70	11 %
Near food processing plant	106	10 %
Beach or sandbank	24	4 %
Forest or mountain area	121	2 %
Footpaths	35	0 %

652

653 *Table 2: STs identified at sampling points positive for L. monocytogenes on repeated occasions*

Site no.	Sampling point description	2016, Oct	2017, Jun, Oct, Nov	2018, Jun	2018, Sept, Oct	2019, Sept	2020, Jan
Urban or residential area							
33	Brook in residential area	ST451				ST4	
48	Garden compost heap	ST451 [a]				ST451 [a]	ST425
49	Garden compost heap	ST451				ST425 [b]	ST425 [b]
66	Puddle next to road	ST20				ST4	
120	Flowerbed in town centre		ST91		negative	ST1813	ST398
121	In front of park bench by flowerbed		ST399		ST451	ST398 [c]	ST398 [c]
123	Grass lawn in town centre		ST398 [d]		ST398 [d]	negative	negative
129	Roadside close to brook		ST37			ST91	negative
251	Decaying leaves/vegetation on bike path		ST18			ST37 [e]	ST37 [e]
252	Soil near horse paddock		ST204			ST398	ST7
253	Along sidewalk curb		ST1			negative	ST425
259	Flowerbed with pigeons, city centre		ST204 [f]			ST204 [f]	ST398
262	At foot of tree, city centre		ST6			ST204	ST120
268	Puddle on gravel path, city park		ST6			ST451	negative
Grazeland or animal path							
53	Decaying vegetation by feeding station	ST4				ST399	ST37
54	Soil close to cattle feeding station	ST91				ST399	ST451
55	Mud close to cattle enclosure	ST37				ST399	ST91
56	Puddle of mud close to cattle enclosure	ST91				ST398	ST6
98	Sheep grazing pasture		ST398		ST451	ST2343	ST91
99	Sheep manure		ST398		negative	negative	ST91
100	Animal tracks by feeding station		ST398		negative	ST37	negative
101	Animal tracks by feeding station		ST398		negative	ST91	negative
130	Soil at edge of pond		ST91		negative	ST121	negative
133	Decaying vegetation at edge of pond		ST398		negative	ST20 [g]	negative
134	Decaying vegetation at edge of pond		ST29		negative	ST4	ST20 [g]
Near food processing plant							
279	Grass next to cold storage entrance		ST732	ST732	ST732		
287	Storm drain outside plant		ST1	negative	ST647		negative
363	Gravel from quay outside factory			negative	ST732		ST647
406	Gravel from quay outside factory				ST1		ST647

654 Allelic distance between isolates with same superscript letter: [a] 3 wgMLST differences, [b] 2 wgMLST differences, [c]
655 2 wgMLST differences, [d] 34 wgMLST differences and 0 cgMLST differences, [e] 7 wgMLST differences, [f] 3 wgMLST
656 differences, [g] 0 wgMLST differences (sites 133 and 134 are located 5 m apart), [h] 0-4 wgMLST differences.

657 *Table 3: L. monocytogenes STs identified on dairy farms*

Visit	Sample	Farm no.																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	17	18
Aug/ Sept 2019	MF			*ns*														
	Feces	*ns*				*ns*			ST91		ST11	ST4				ST7	ST21	
	Feed					ST226	ST37						ST37				ST6	
	TS															*ns*		
Nov/Des 2019	MF	*ns*		*ns*	ST451			ST2760	*ns*			ST451	ST451					
	Feces	ST18			ST16	ST394		ST451			ST204			ST8				
	Feed	ST18			ST226	ST226	ST451	ST37					ST91					
	TS															ST37		
Jan 2020	MF	ST18							*ns*				ST37					
	Feces	ST18		ST2761									ST451				ST18	
	Feed	ST425	ST394	ST2761		ST7				ST224		ST4	ST451		ST91	ST91	ST18	
	TS			ST20														
Feb/Mar 2020	MF	*ns*						*ns*			*ns*	*ns*	ST177			*ns*		
	Feces	ST20		ST2761			ST451				ST21	ST451	ST177		ST91			
	Feed	ST37	ST394				ST451				ST412		ST177		ST91			
	TS	ST20																
May 2020	MF	*ns*	*ns*	*ns*									ST91	ST91			*ns*	
	Feces	*ns*	*ns*	*ns*									ST91	ST37			*ns*	
	Feed	*ns*	*ns*	*ns*									ST511	ST8			*ns*	ST91
	TS	*ns*	*ns*	*ns*									ST37	ST91			*ns*	
Jun 2020	MF	*ns*		*ns*			ST91		ST451	ST451								
	Feces	*ns*		*ns*		ST37				ST21	ST412	ST4	ST451	ST124			*ns*	
	Feed	*ns*		*ns*						ST451	ST394		ST177				*ns*	
	TS	*ns*	*ns*	*ns*					*ns*								*ns*	

658 MF; milk filter, TS; teat swab, *ns*; not sampled, empty cells; negative for *L. monocytogenes*

Figures

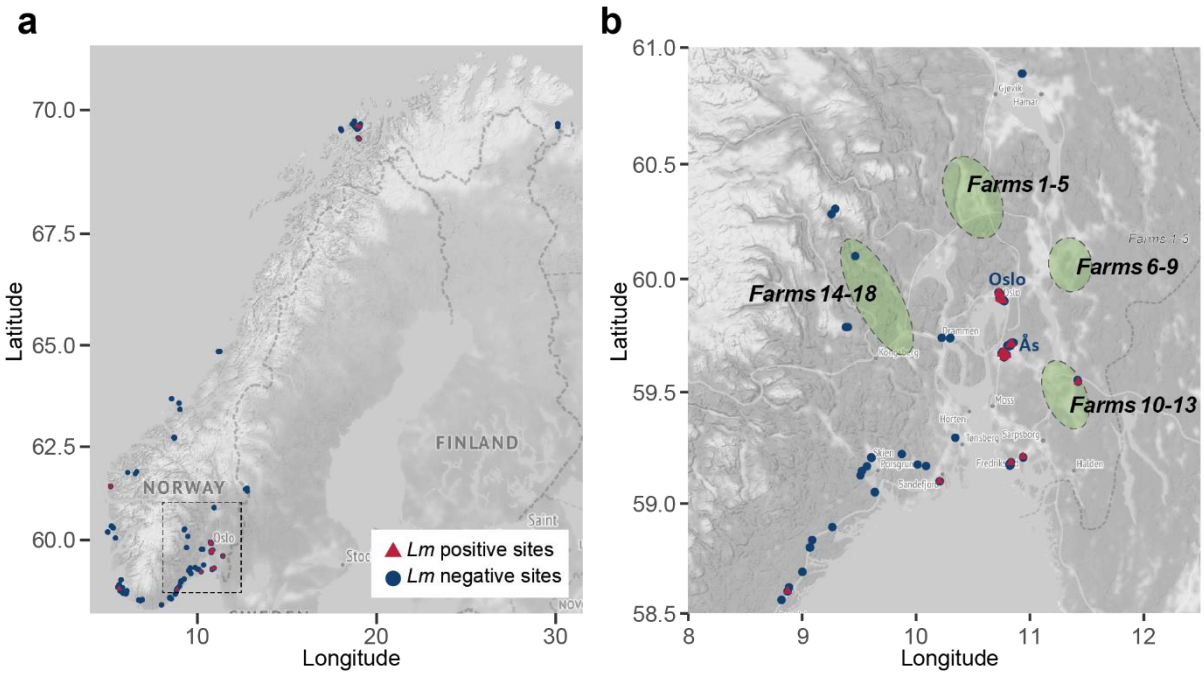


Figure 1: Maps showing the geographic location of sampling sites. A) The location of sampling sites in rural and urban environments in Norway, with red triangles representing *L. monocytogenes* positive samples and blue circles negative sampling points. The area outlined by the dashed square in **a)** is the area shown in **b)**. The green shaded areas in **b)** show the geographical origins of the dairy cattle farms sampled for *L. monocytogenes* in Idland *et al*. (16).

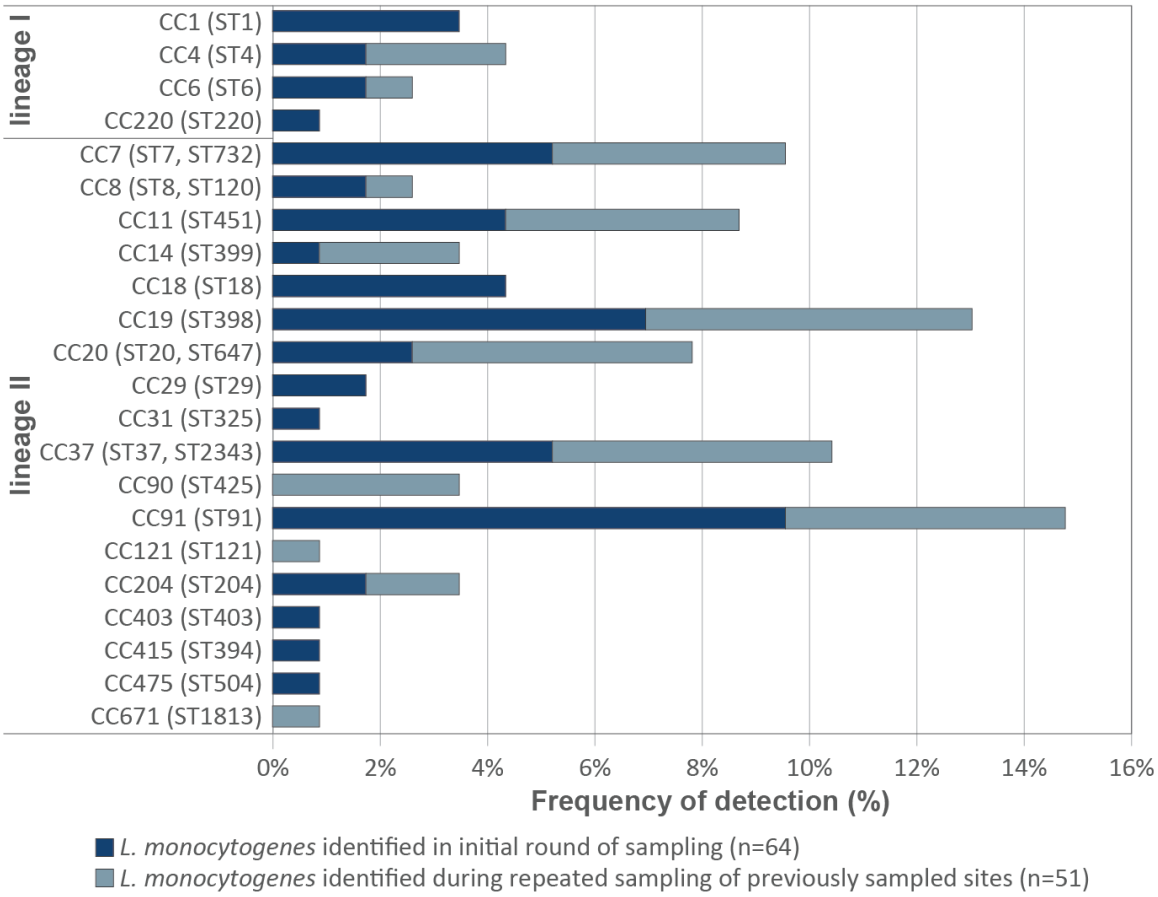
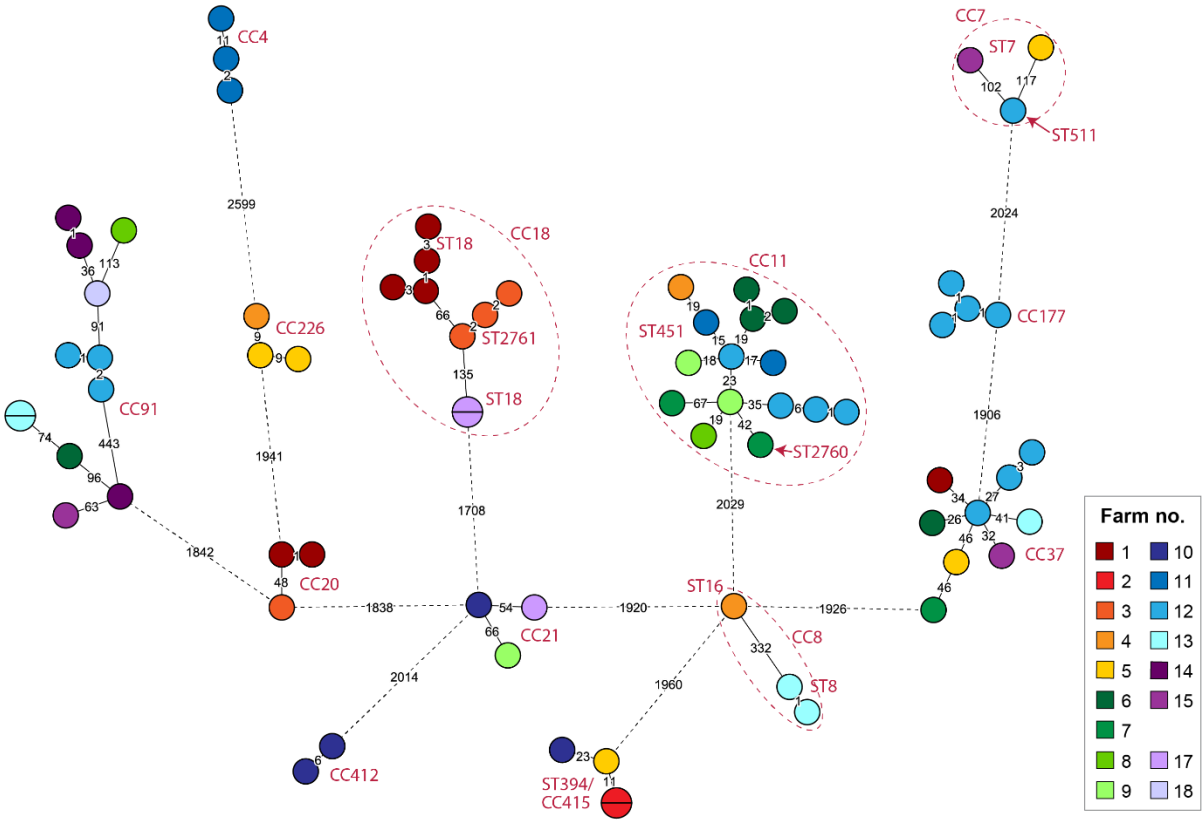


Figure 2: Distribution of CCs among identified isolates from rural and urban environments. The data is reported as percentage of the grand total number of isolates (n=115). STs represented within each clonal complex (CC) are given in parentheses.



670

671 **Figure 3: Phylogeny for the *L. monocytogenes* isolates from dairy farms.** A minimum spanning tree
672 based on wgMLST analysis. The area of each circle is proportional to the number of isolates
673 represented, and the number of allelic differences between isolates is indicated on the edges
674 connecting two nodes. The CCs and STs are indicated next to each node (the CC number is the same
675 as the ST number unless indicated). Isolates separated from the nearest other isolate by >1700
676 wgMLST alleles (D011L, D058L, D080L, D084L, D144L, and D190L) were excluded from the figure for
677 clarity.

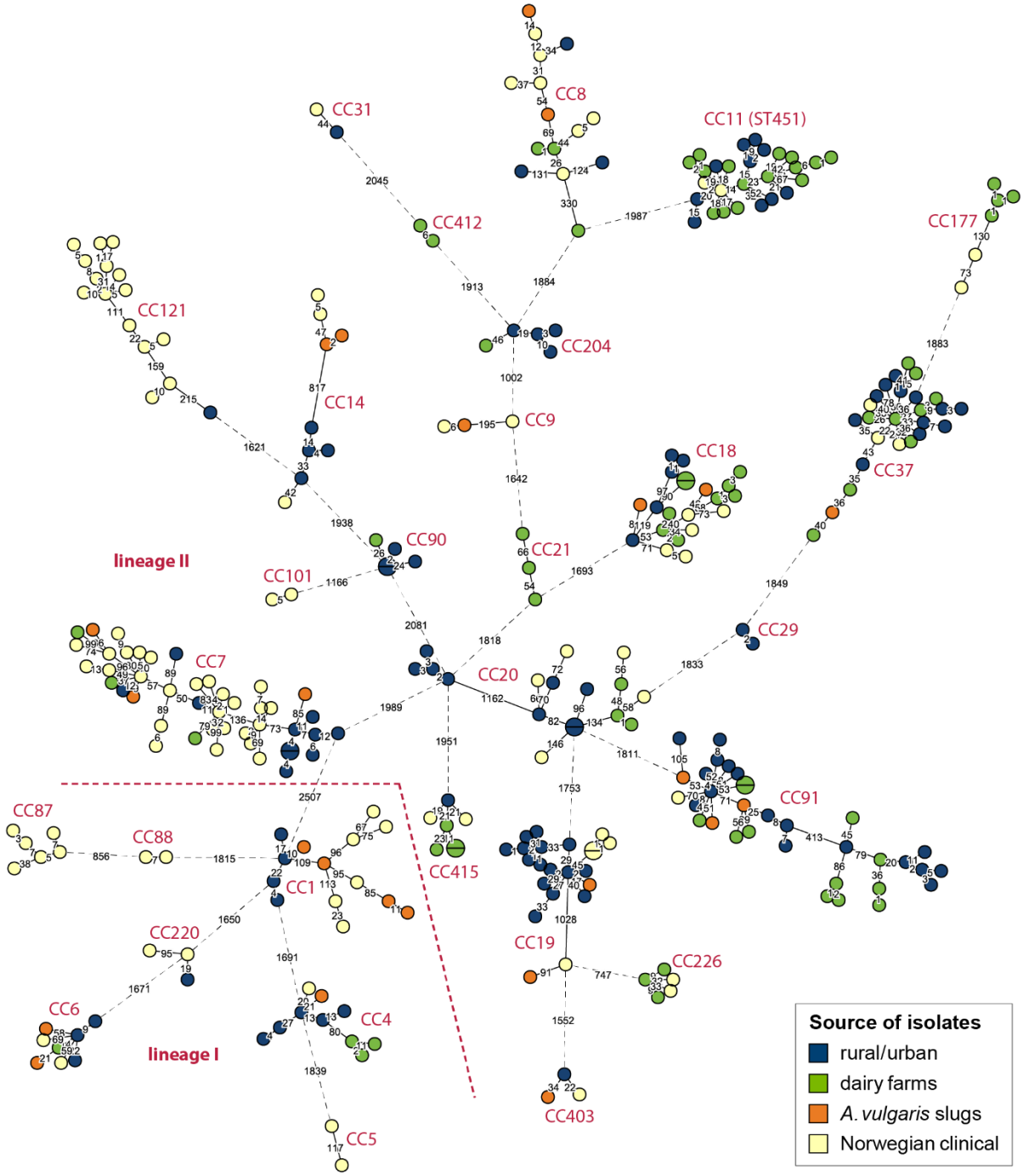
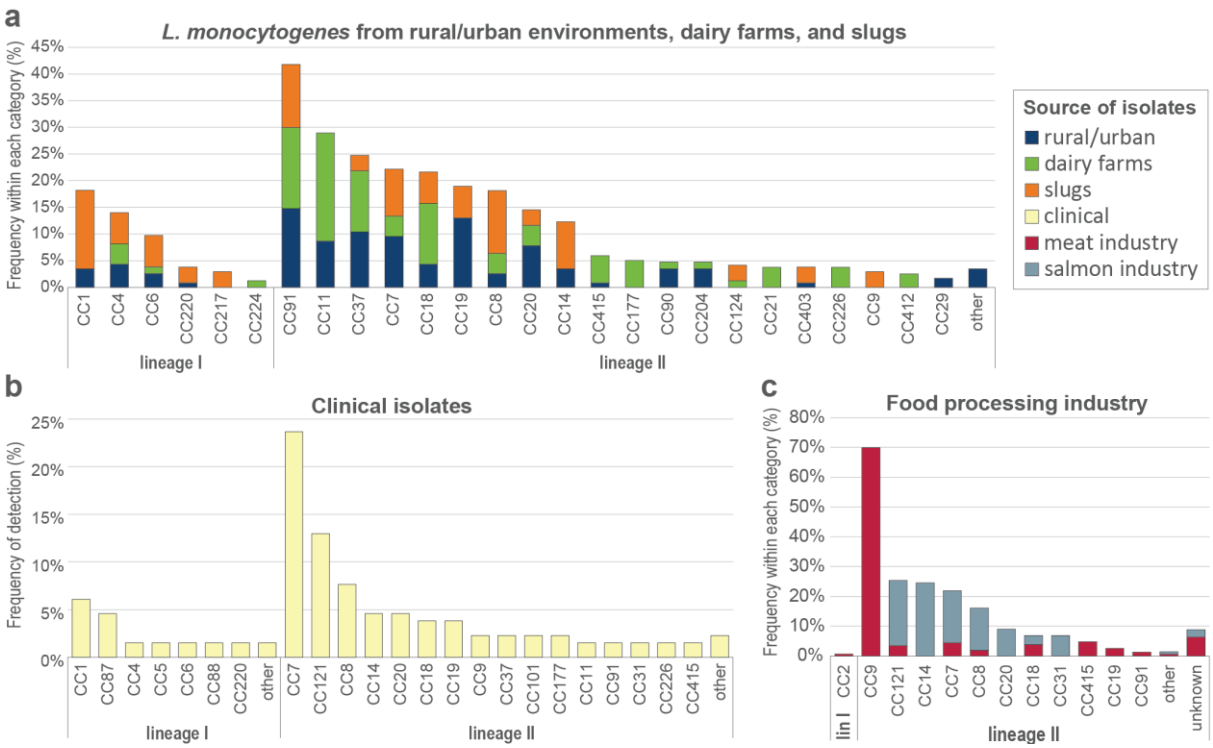


Figure 4: wgMLST phylogeny for *L. monocytogenes* isolates from Norway. Shown is a minimum spanning tree based on wgMLST analysis. The area of each circle is proportional to the number of isolates represented, and the number of allelic differences between isolates is indicated on the edges connecting two nodes. Edges shown as dashed lines separate clusters belonging to different clonal complexes, which are indicated next to each node. Isolates separated from the nearest other isolate by >900 wgMLST alleles (MF7614, MF6841, D084L, D144L, D190L, ERR2522251, ERR2522285, ERR2522267, ERR2522291, and ERR2522298) were excluded from the figure for clarity.

686



687 **Figure 5: Prevalence and distribution of *L. monocytogenes* MLST clonal complexes (CCs) from**
688 **different sample types in Norway.** The data are reported as percentages of isolates within a given CC
689 in each source category. **a)** Prevalence in rural and urban environments (isolated during 2016-2020;
690 n=115 isolates), dairy farms (2019-2020; n=87), and slugs (2012; n=34). The "other" category
691 comprises one isolate each for CC31, CC121, CC475, and CC671. **b)** Prevalence in publicly available
692 genomes from human cases of listeriosis in Norway (2010-2015; n=129 and 2018; n=2). The lineage I
693 "other" category comprises a CC3 and a CC59 isolate, and the lineage II "other" category includes
694 one isolate each for CC11, CC101, and CC177. **c)** Prevalence within food processing facilities in
695 Norway. The CCs were inferred for isolates from five meat and four salmon processing facilities
696 (meat: 2012-2015; n=293, and salmon: 2011-2014; n=358). The data used to predict the CC for each
697 isolate were multiple-locus variable-number tandem repeat analysis (MLVA) obtained for all isolates
698 and MLST data obtained for representative isolates from each obtained MLVA profile (47). The
699 "unknown" category represents isolates with MLVA profiles identified only once and not subjected to
700 MLST.

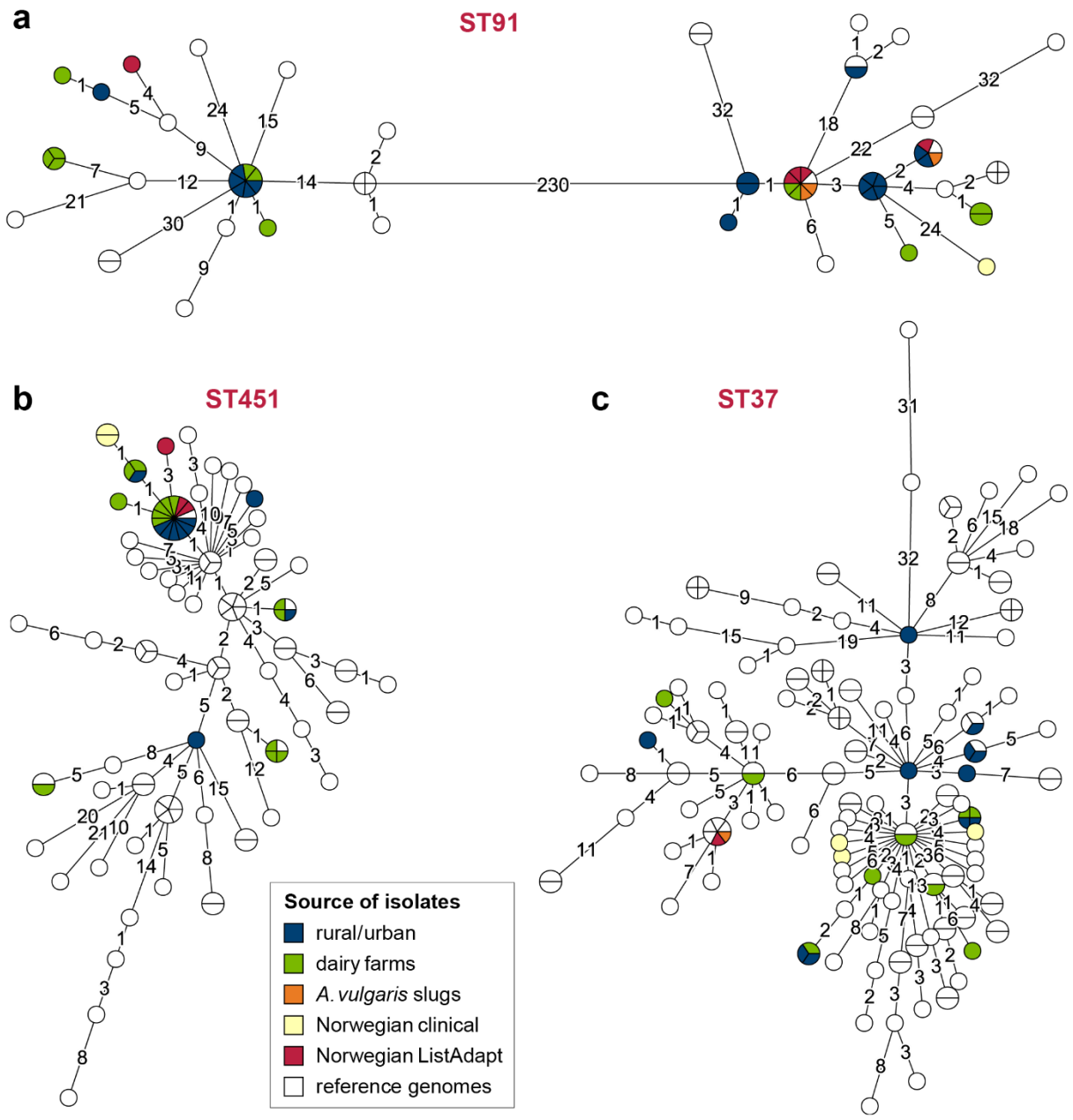


Figure 6: cgMLST phylogeny for the most common STs identified in the current study. Minimum spanning trees based on cgMLST allelic profiles for: **a)** ST91, **b)** ST451, and **c)** ST37, showing the relationship between the Norwegian isolates from natural environments, Norwegian clinical isolates, and reference genomes obtained from public databases. Reference genomes were obtained from the BIGSdb-*Lm* database hosted at the Pasteur Institute, WGS data from the EU project ListAdapt (also including genomes from Norwegian sources, labelled in red), and genome assemblies from NCBI GenBank. The area of each circle is proportional to the number of isolates represented, and the number of allelic differences between isolates is indicated on the edges connecting two nodes.

711 # List of supplemental information

712 **S1 Text** (*Supplemental_S1_Text.docx*): Additional information regarding clusters of closely related *L.*
713 *monocytogenes* isolates.

714 Supplemental Tables (*Supplemental_Tables.xlsx*):

715 **S1 Table**: Isolates included in the current study

716 **S2 Table**: Clusters of isolates from rural or urban environments

717 **S3 Table**: Clusters of dairy farms isolates

718 **S4 Table**: Clusters containing isolates from both rural/urban environments and dairy farms

719 **S5 Table**: Clusters containing isolates from slugs or both slugs and rural/urban/farm environments

720 **S6 Table**: Clusters of clinical isolates

721 **S7 Table**: Clusters containing both clinical isolates and isolates from rural/urban/farm
722 environments or slugs

723 **S8 Table**: Survey of STs and CCs found in previous studies of various natural and farm
724 environments