



GenEpi-BioTrain – Genomic Epidemiology and Public Health Bioinformatics

Genomic epidemiology applied to *Listeria* and cluster detection

May 2024

Intended Learning Objectives



Specific objectives of this session:

Background information on the **biology of *Listeria monocytogenes*** (genomic and molecular aspects, virulence, global distribution, clinical aspects, diagnostics methods and treatments)

Know how to **integrate genomic and epidemiological** data using tools (BIGSdb, GrapeTree, ReporTree, MicroReact, itol, PastML) to analyze **transmission patterns** and **identify sources** of infection during disease outbreaks

Learn techniques for *Listeria* **cluster detection, characterization, surveillance,** and **risk assessment** in relation to *Listeria* infections and foodborne disease outbreaks

Listeriosis – the deadliest foodborne disease



FIGURES

Listeria monocytogenes

- **Foodborne pathogen (G+, Firmicutes)**

- gastroenteritis
- bacteremia
- materno-fetal infections
- meningoencephalitis

wide incubation period (1-67 days)

- **High case-fatality rate** (20-30%) among at-risk patients:

- immunocompromised
- elderly
- pregnant women

- **Asymptomatic carriage:**

- 5-10% of healthy people

- **Genomic surveillance** (FR ≥ 2015; ~450 cases/year)

- improved outbreak detection
- reduced nr of cases per cluster
- still ~80% of human clusters w/unknown source of contamination

- Reservoirs: **soil** and **asymptomatic hosts** (mostly humans, bovines, pigs and chicken)

FIGURES

Radoshevich & Cossart, Nat Microb Rev 2018
Lecuit, Microbes Infect 2007



Important economic and public health impact

- Able to **survive at 4°C** and **persist for long periods** in food-production plants
 - food recalls/withdraws
 - outbreaks

South Africa declares end to largest ever Listeria outbreak

L1-SL6-ST6-CT4148

By Joe Whitworth on September 4, 2018

The largest ever detected Listeria outbreak in known history is over, according to officials in South Africa.

Between Jan. 1, 2017, and July 17, 2018, South Africa recorded **1,060 laboratory-confirmed cases of listeriosis, including 216 deaths.**

The outbreak was linked to a ready-to-eat (RTE) processed meat plant owned by Tiger Brands, the Enterprise Foods production facility in Polokwane. However, investigators for the government and in the private sector have not been able to determine how the outbreak strain was introduced in the affected factory.

A total of 12 million South African rand (\$810,000) was spent to deal with the outbreak.

<https://www.foodsafetynews.com/>

FIGURES



RECALL

Lm outbreaks in Europe

Seven face charges in deadly Listeria outbreak in Spain

By Joe Whitworth on July 1, 2022

Seven people are set to stand trial as part of Spain's largest ever Listeria outbreak, which occurred in 2019.

A judge in a court in Seville this week decided to continue proceedings against seven defendants for offences including alleged crimes against public health and injury to a fetus resulting in abortion. At the conclusion of an investigation into the outbreak, Pilar Ordóñez also considered that Seville Council could be held civilly liable, this means it would have to pay compensation if convicted.

The outbreak from "La Mecha" brand chilled roasted pork produced by Magrudis affected almost 250 people. During the health alert between mid-August and mid-October 2019, four people died and there were six abortions.

<https://www.foodsafetynews.com/>

FIGURES

EFSA Journal 2023

Sources of *Listeria* Infection

The usual suspects....

FIGURES



But virtually anything (raw/non-pasteurized or processed)...

***LISTERIA* AND BLUE BELL ICE CREAM**
Contaminated production facilities and illnesses linked to Blue Bell Creameries

FIGURES

L. monocytogenes cell invasion



Maudet *et al.*, Science 2022

FIGURES

Lecuit *et al.*, Infect Immun 1997

Gessain *et al.*, J Exp Med 2015

FIGURES

Vasanthakrishnan *et al.*, Env Microb 2015

Pizarro-Cerdá *et al.*, Cold Spring Harb Perspect Med 2012
Portnoy, Auerbuch & Glomski, J Cell Biol 2009

FIGURES

Listeria Pathogenic Island 1 (LIPI-1)

Lm population structure

MLST (7 genes)

Ragon *et al.*, PLoS Pathog 2008



- Clonal population structure
 4 phylogenetic lineages
 >170 clonal complexes (MLST)
- Different clonal complexes are associated to
 different food matrixes and ecological niches.

FIGURES

FIGURES

FIGURES

Biodiversity of *Lm*

Maury, Tsai et al., Nature Genetics 2016

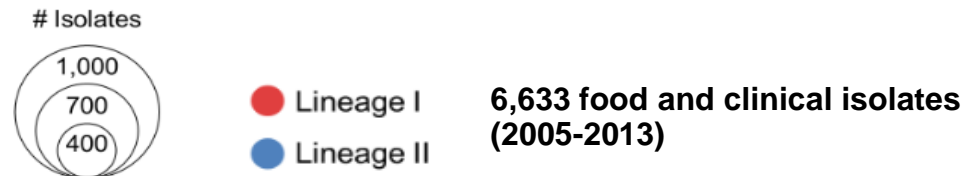
Infection-associated clones

Food-associated clones

Bacterial load in the brain



FIGURES



Hypervirulent clones are more often associated with the infection of people with less comorbidities
(MONALISA study, Charlier-Woerther et al., Lancet ID 2017)

Interplay of *Lm* virulence and host susceptibility



Disson, Moura & Lecuit, Trends Microb 2021

FIGURES

Maury*, Tsai* *et al*, Nat Microb 2016
Charlier et al., Lancet Inf Dis 2016

Why *Lm* genomic surveillance?

- **Maximum resolution** on strain discrimination
- **Increased specificity** of cluster investigation
 - > Optimized **investigation resources**
- **Increased sensitivity** of outbreak detection
 - > **Earlier detection and control** of outbreaks
- **Enhanced detection** of persistent and geographically widespread strains.
- **Prediction of virulence and resistance** traits

L. monocytogenes
2.9 Mb

Genome-wide approaches for surveillance

Multiple Genome Alignment:

- Difficult to interpret with draft genomes
- Computational demanding and limited to few genomes
- Not expandable

Single Nucleotide Polymorphism (SNP) calling:

- Good discrimination; unlimited no. of genomes
- Computational demanding and not expandable
- Problematic with rearrangements and recombination events

Genome-wide allele typing (genome-wide MLST):

- Good discrimination; unlimited no. of genomes
- Minimizes biases of recombination and rearrangements
- Expandable

Listeria monocytogenes core genome Multilocus Sequence Typing (cgMLST)



■ cgMLST1748 loci
■ GC content

EGD-e
NC_0032
10
2867 genes
2.94 Mb

Institut Pasteur scheme

Nat Microbiol 2016

1748 loci (~1.57 Mb; 53% EGDe genome)

Validated in >1000 *Lm* genomes from different countries and sources
Discarded loci consistently absent and/or w/conflicting detection (paralogues)

High locus detection (>96%)

Low error rates (< 0.02%)

High congruence with wgSNP (>99%)

Validated across:

- independent cultures and sequencing runs
- different sequencing coverages (**recommended: >40X in Illumina**)
- assembly-free and assembly-based pipelines

Implemented in:

- BIGSdb-*Listeria* (open-source web; <https://bigsdb.pasteur.fr/listeria/>)
- BioNumerics (commercial software; Applied Maths)
- Chewbbaca/chewie-NS (open-source; Mamede *et al.*, Nucl Acids Res 2020)

The cgMLST principle

Sequence variants (alleles) are defined for each locus

Allele

1
1
2
3
3

FIGURES

1748 loci

cgMLST type

CT304
CT257
CT274
CT274
CT275
CT314
CT308

Lm genomes

Loci detection

NGS data

cgMLST (1748 loci) vs MLST (7 loci) nomenclature



sublineages ~ clonal complexes

MLST (7 loci):

- sequence types (STs)
- clonal complexes (CCs)

FIGURES

cgMLST (1748 loci):

- cgMLST types (CTs)
- sublineages (SLs)

cgMLST types (full string): **lineage - sublineage - ST(MLST) - cgMLSTtype**

e.g. L1-SL1-ST1-CT304

Discrimination levels

L. monocytogenes

phylogenetic lineage (I-IV or L1-L4)

serogroup (PCR-serotyping: IVb, IIb, ...)

[serotype (agglutination: 4b, 1/2b, ...)]

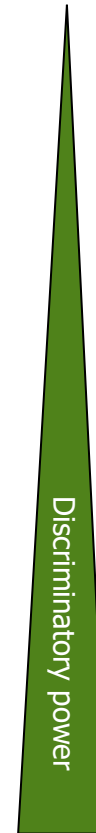
clonal complex (MLST: CC1, CC2, ...)

sublineage (cgMLST: SL1, SL2, ..)

sequence type (MLST: ST1, ST2, ...)

cgMLST type (cgMLST: CT1, CT2, ...)

wgSNPs



FIGURES

Single Nucleotide Polymorphisms Analysis vs cgMLST

wgSNP

cgMLST

FIGURES

cgMLST profile distances strongly correlate with whole genome SNP distances.

Sequenced based phylogeny vs cgMLST clustering

1748 concatenated sequences (1.57 Mb)

IV
III
I

cgMLST profiles

IV
III
I

FIGURES

II

II

Single Linkage

Recombination purged-phylogenies strongly correlate with cgMLST profile distances.

BIGSdb-*Listeria*

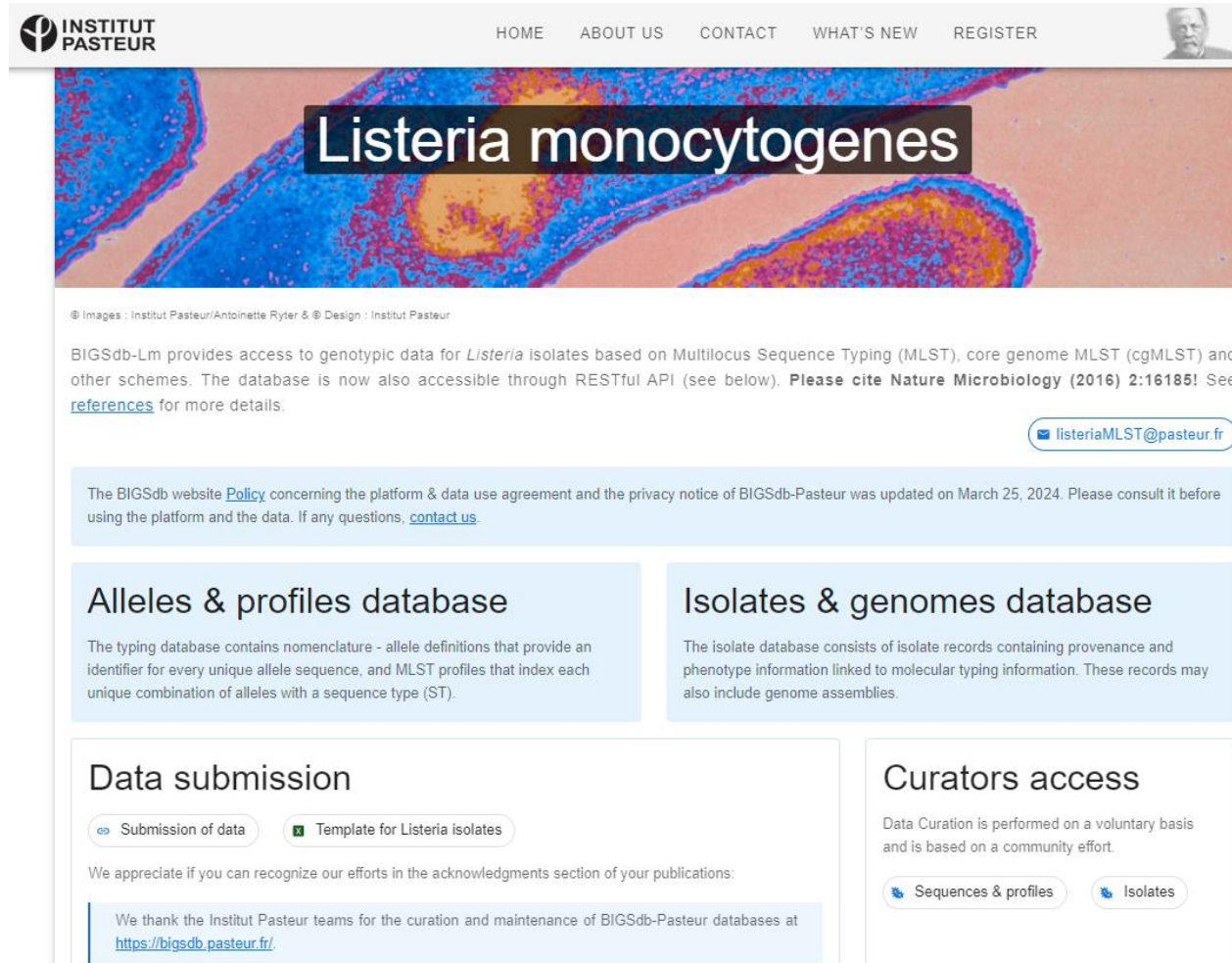
cgMLST

MLST

Serogrouping

Virulence
Genes

Resistance
Genes
(AMR,
QACs,SSIs,...)



The screenshot shows the homepage of the BIGSdb-Listeria website. At the top is the Institut Pasteur logo and a navigation bar with links: HOME, ABOUT US, CONTACT, WHAT'S NEW, and REGISTER. A small portrait of Louis Pasteur is in the top right. The main header features a colorful heatmap image of a *Listeria monocytogenes* cell with the text "Listeria monocytogenes" overlaid. Below this, a paragraph states: "BIGSdb-Lm provides access to genotypic data for *Listeria* isolates based on Multilocus Sequence Typing (MLST), core genome MLST (cgMLST) and other schemes. The database is now also accessible through RESTful API (see below). Please cite *Nature Microbiology* (2016) 2:16185! See [references](#) for more details." A contact button "listeriaMLST@pasteur.fr" is on the right. A light blue box contains a policy update notice dated March 25, 2024. Below are two main sections: "Alleles & profiles database" (describing nomenclature and MLST profiles) and "Isolates & genomes database" (describing isolate records and genome assemblies). At the bottom, there are sections for "Data submission" (with buttons for "Submission of data" and "Template for Listeria isolates") and "Curators access" (with buttons for "Sequences & profiles" and "Isolates"). A footer note thanks the Institut Pasteur teams and provides the URL <https://bigsdbs.pasteur.fr/>.

- Curated global genomic db
- Access to public isolates, profiles and allele sequences (through browser and RESTful API)
- **User accounts for private projects**

May 2024:

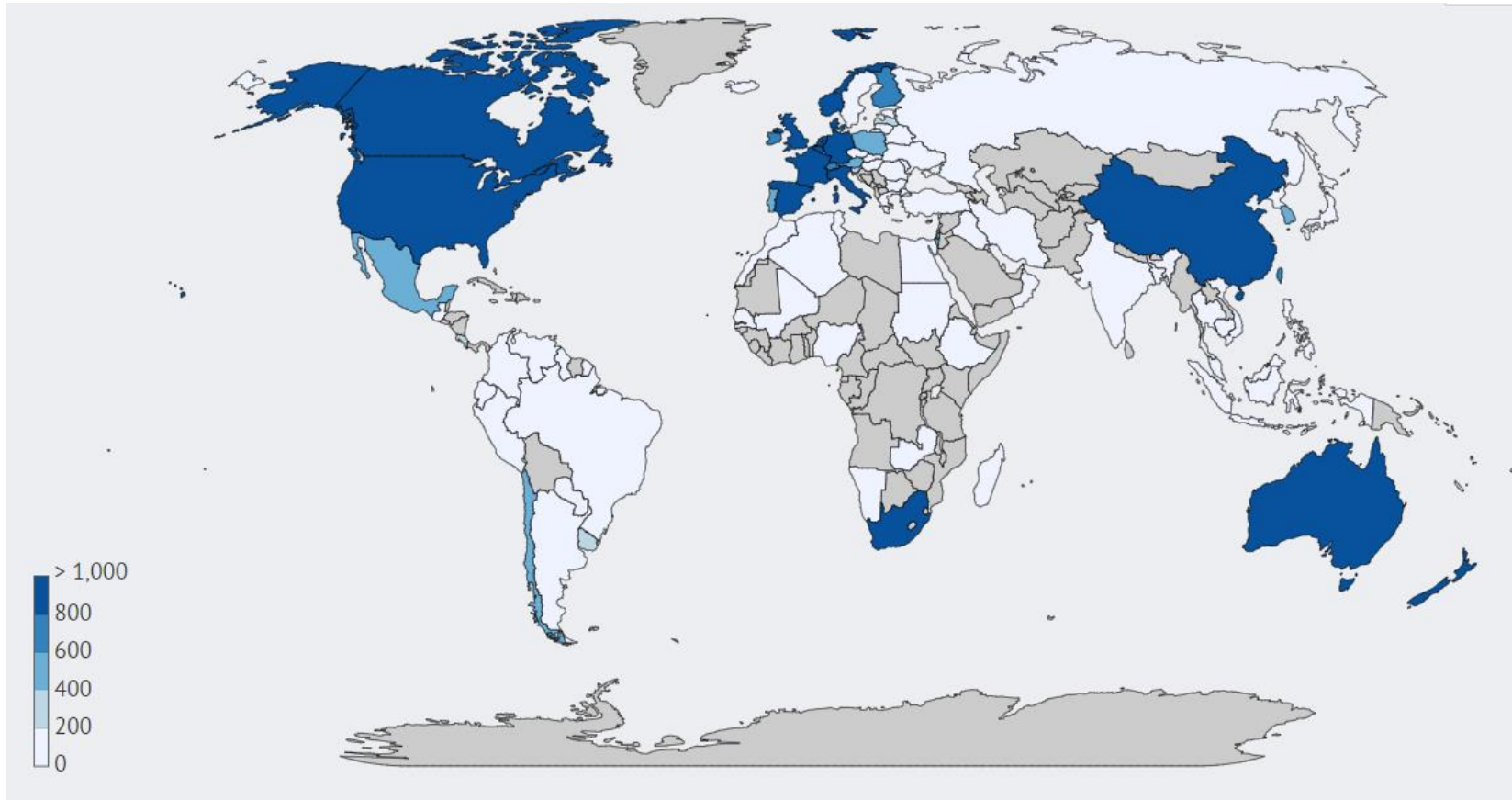
> 90,000 genomes

~ 350 submitters
(research & public health labs)

<http://bigsdbs.pasteur.fr/listeria>

BIGSdb-*Listeria*

90,477 genomes (>90 countries)



- Curated global genomic db
- Access to public isolates, profiles and allele sequences (through browser and RESTful API)
- **User accounts for private projects**

Apr 2024:

> 90,000 genomes

~ 350 submitters
(research & public health labs)

<http://bigsdb.pasteur.fr/listeria>

Cut-offs in *Lm* cgMLST scheme (IPasteur)

Pairwise allelic differences (AD) among isolates

(cgMLST types)

7AD

FIGURES

(sublineages) 150AD

(N=1,696 isolates)

(n=426, 49 outbreaks)

Cut-offs established for a unified nomenclature for global communication

Cut-offs in *Lm* cgMLST scheme (IPasteur)

	nr. allelic differences	min. cgMLST similarity
closely related isolated	0-7	99.600%
likely unrelated	>8	<99.600%

→ cgMLST cluster alert

FIGURES

Other *Listeria monocytogenes* cgMLST schemes

Different starting datasets (diversity)
Different schemes
Different validations steps
Different cut-offs



Moura *et al.*, Nat Microbiol 2016
[~1700 genomes]
1748 loci



FIGURES

Ruppitsch *et al.*, J Clin 2015
[~40 genomes]
1701 loci

Pightling *et al.*, BMC Microb 2015
[~200 genomes]
1013 loci

Enhanced understading of *Lm* biology

- **LIPI-1** is highly conserved
- **LIPI-3** almost exclusively in lineage 1
- **LIPI-4** mostly in SL4 (CC4), but also in other rare clinical SLs

Traits associated to increased fitness within hosts

FIGURES

- ***inlA* truncations** strongly associated to lineage II and food isolates ($p < 0.0001$)
- **Benzalkonium chloride resistance** more frequent in lineage II and SL121 (CC121) ($p < 0.0001$)

Traits associated to increased fitness in food-production environments

Moura *et al.*, Nat Microb 2016

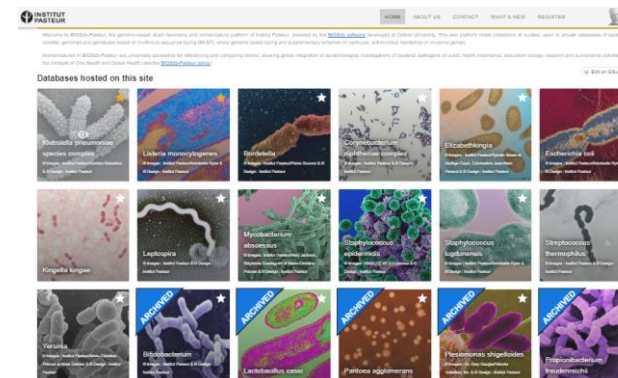
Integration of genomic and epidemiological data to understand transmission

Useful user-friendly tools :

Genome and cgMLST database, allele caller and typing info:

<https://bigsdbs.pasteur.fr/>

Listeria monocytogenes



Core genome sequence **alignment** (from assemblies...; useful for species without cgMLST):

<https://github.com/marbl/parsnp> **Parsnp**

wgSNP **alignment** (from assemblies or raw data; requires reference genome; useful to confirm consistance cgMLST & wgSNP):

<https://bigsdbs.pasteur.fr/> **Snippy**

(Automatic) **Clusters breakdown & summary** (from profiles, seq alns, etc):

++ <https://github.com/insapathogenomics/ReporTree>



Clusters visualisation (MSTrees):

++ <https://github.com/achtman-lab/GrapeTree>
++ <https://www.phyloviz.net/>



Tree visualisation (+multiple metadata):

++ <https://itol.embl.de/>
++ <https://microreact.org/>
<https://auspice.us/>
<https://treeviewer.org/>




Ancestral reconstruction (eg. source, country, etc):

<https://pastml.pasteur.fr/>



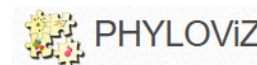
++ (also) in BIGSdb!

<https://bigsdbs.pasteur.fr/listeria>

id	isolate	aliases	species	isolation year	country	region	outbreak info	source	infection	food type	Isolate fields 			source lab	phylogenetic lineage	clonal complex	sublineage	cgmlst type	comments	accession number	abcZ	bgIA	cat	dapE	dat	MLST					CC	Lineage	PCR-serogroup
											source description	sample type	other source info													ldh	thkA	ST	serogroup				
1	EGD-e		L. monocytogenes	1960	Unknown			Unknown					Institut Pasteur, Paris, France	II	CC9	SL9	CT637		NC_003210	6	5	6	20	1	4	1	35	CC9	II	IIc			
2	F2365	FSLJ1-110	L. monocytogenes	1985	USA		Jalisco cheese outbreak, California 1985	Food			Cheese			I	CC1	SL1	CT304		NC_002973; SRR2229351	3	1	1	1	3	1	3	1	CC1	I	IVb			
3	SLCC2482		L. monocytogenes	1966	Unknown			Unknown					Institut Pasteur, Paris, France	I	CC3	SL3	CT83		NC_018591; NC_018888	4	4	4	3	2	1	5	3	CC3	I	IIb			
4	LM80459		L. monocytogenes	1999	France		Outbreak 1999	Human			BACT		Institut Pasteur, Paris, France	I	CC4	SL4	CT188		NC_012488	1	2	12	3	2	5	3	4	CC4	I	IVb			
5	CLIP 2005/00704		L. monocytogenes	2005	France	dpt92		Human	MN		MN		Institut Pasteur, Paris, France	I	CC5	SL5	CT125		ERR1100912	2	1	11	3	3	1	7	5	CC5	I	IIb			
6	CLIP 2009/00372		L. monocytogenes	2009	France	dpt56		Human	CNS				Institut Pasteur, Paris, France	I	CC6	SL6	CT467		ERR1100913	3	9	9	3	3	1	5	6	CC6	I	IVb			
7	SLCC5850		L. monocytogenes		UK			Animal			rabbit		Institut Pasteur, Paris, France	II	CC7	SL7	CT720	PrfA mutation Tyr76Stop	NC_018592	5	8	5	7	6	22	1	12	CC7	II	IIa			
8	08-5578		L. monocytogenes	2008	Canada		Outbreak 2008	Human			BACT		Public Health Agency of Canada, Winnipeg, Canada	II	CC8	SL8	CT786		NC_013766; NC_013767	57	6	2	29	5	3	1	292	CC8	II	IIa			
9	FSL R2-561		L. monocytogenes		Unknown			Unknown			Unknown			II	CC9	SL9	CT603		NC_017546	6	5	6	4	1	4	1	9	CC9	II	IIc			
10	CLIP 2009/00277		L. monocytogenes	2007	Tunisia			Food			Unknown		Institut Pasteur, Paris, France	II	CC101	SL101	CT974		ERR1100932	7	15	15	8	6	14	9	101	CC101	II	IIa			

Analysis tools

Breakdown: [Fields](#) [Two Field](#) [Combinations](#) [Polymorphic sites](#) [Publications](#) [Sequence bin](#)
 Analysis: [BURST](#) [Codons](#) [Gene Presence](#) [Genome Comparator](#) [BLAST](#) [rMLST species id](#) [PCR](#)
 Export: [Dataset](#) [Contigs](#) [Sequences](#)
 Third party: [GrapeTree](#) [iTOL](#) [Microreact](#) [PhyloViz](#) [ReporTree](#)



Integration of genomic and epidata to understand transmission



eg. *Lm*-CC1 (most widespread and pathogenic *Lm* clonal complex)

- **3 sublineages** (SL1, SL150, SL404), with SL1 comprising 99% of isolates
- **8 genetic clusters within SL1** with uneven prevalence

FIGURES

International transmission of *Lm*



FIGURES

● L1-SL1-ST1-CT288 (DK-UK, 2002-2013, $n=4$)
● L1-SL1-ST1-CT320 (US-UK, 2012, $n=2$)
● L1-SL6-ST6-CT443 (DK-UK-FR, 2003-2013, $n=8$)

● L2-SL7-ST7-CT732 (US-DK, 2002-2013, $n=3$)
● L2-SL9-ST9-CT628 (UK-FR, 2008-2013, $n=2$)
● L2-SL9-ST9-CT617 (US-UK, 2013, $n=4$)
● L2-SL9-ST9-CT604 (US-CA, 2010-2013, $n=6$)
● L2-SL14-ST14-CT956 (UK-FR, 2014-2015, $n=2$)
● L2-SL121-ST121-CT917 (UK-FR, 2007-2009, $n=3$)

- Evidence for multiple cross-country transmissions occurring in the past.
- Identical cgMLST types in different countries also evidences recent international transmission of *Lm*.

CC1 is ancient but sublineages diversification is recent

FIGURES

CC1
197 AD [860 BC - 1045 AD]

SL1 emerged 160 years ago and quickly spread worldwide



FIGURES



Ancestral geographic origin

PastML

FIGURES

Recent SL1 transmission chains are mostly local



Noémie Lefrancq & Henrik Salje
(Simon Cauchemez's Unit, IPasteur)

Recent transmissions chains are mostly **locally acquired** with **limited inter-country spread**:

FIGURES

- Pairs of isolates from the same 2-year period and the same country are 18.7 times (95% CI: 4.7-190.7) more likely to have their MRCA within the past 5 years than pairs of isolates coming from other countries, regardless their distance.
- **Strong local spatial structure persisting for decades-long**

within country
(France)

Moura et al. Sci Adv 2021

Most clusters are country and region-specific ->
importance to **enhance national surveillance programs** through all economic sectors

Detection of clusters

eg. *Lm* strain and cluster diversity - Taiwan



FIGURES

MSTree of cgMLST profiles

(clearly show allelic distances between isolates,
highlight isolates within the same cluster, ...)

Taiwan - Strain & traits diversity



FIGURES

Single linkage of cgMLST profiles
iTol tree visualization w/selected traits

(show the phylogenetic placement of selected traits,
the distribution of different sources or regions
across *Lm* diversity, ...)

Taiwan – Mandatory declaration since 2018



TABLE 2 cgMLST types detected in this study comprising more than two isolates ($n = 33$ out of 117)

FIGURES

Long timespan of clusters,
some with national spread

Real-time surveillance

Software | [Open access](#) | Published: 15 June 2023

ReporTree: a surveillance-oriented tool to strengthen the linkage between pathogen genetic clusters and epidemiological data

[Verónica Mixão](#), [Miguel Pinto](#), [Daniel Sobral](#), [Adriano Di Pasquale](#), [João Paulo Gomes](#) & [Vítor Borges](#) 

[Genome Medicine](#) **15**, Article number: 43 (2023) | [Cite this article](#)

FIGURES

Summary of clusters & epidata:

[partitions_summary.tsv](#)

Take home messages:



Lm is a growing **opportunistic pathogen**, likely prone by

- industrialization, refrigeration, and globally animal and food trades
- increasing ageing and immuno-compromised population; microbiota dysbiosis

Majority clusters and transmission chains are local

(importance of comprehensive surveillance systems, including in animal farms)

Evolution towards **loss of virulence** and acquisition of **resistances in food environments**

Genomic surveillance is crucial to uncover **clusters, carriage, reservoirs, diversity, and dynamics** of *Listeria*, with **impacts on control strategies and public health**.

Take home messages:

FIGURES

Majority of investigations rely on known at-risk products

**WGS coupled with
exhaustive sampling in
food and animal environments
may disclosure unknown sources of infection**

Integration of WGS and epidemiological data contribute to
better understand *Lm* transmission and define strategies to reduce the risk of infections.

Acknowledgements

The creation of this training material was commissioned by ECDC to Institut Pasteur with the direct involvement of Alexandra Moura (WHO-Collaborating Centre and National Reference *Centre Listeria*, Institut Pasteur, Paris).