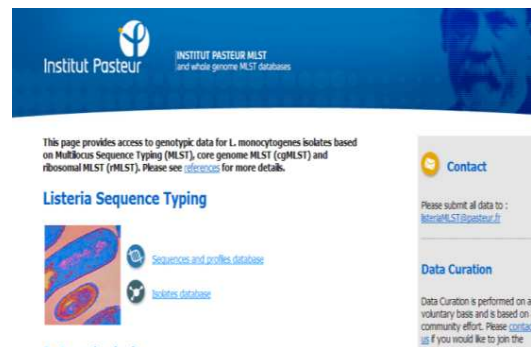
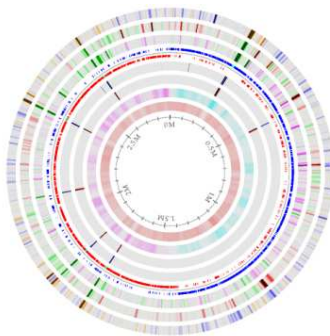
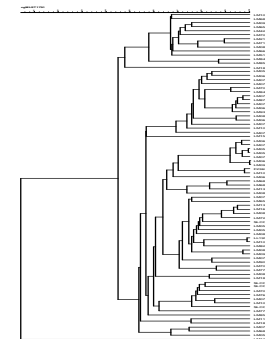


Development of a core genome MLST scheme for global epidemiology and population biology of *Listeria monocytogenes*

Alexandra Moura
amoura@pasteur.fr

Biology of Infection Unit & National Reference Center for Listeria

A screenshot of the Institut Pasteur MLST website. The header features the Institut Pasteur logo and the text "INSTITUT PASTEUR MLST and whole genome MLST databases". Below the header, a paragraph states: "This page provides access to genotypic data for L. monocytogenes isolates based on Multilocus Sequence Typing (MLST), core genome MLST (cgMLST) and ribosomal MLST (rMLST). Please see references for more details." The main content area is titled "Listeria Sequence Typing" and includes links to "Sequences and profiles database" and "Isolates database". On the right side, there is a "Contact" section with the email "Intern@MLST@pasteur.fr" and a "Data Curation" section with a "contact" link. The footer contains a small disclaimer about data curation.

Acknowledgments

Biology of Infection Unit & National Reference Center for *Listeria*, Institut Pasteur

Viviane CHENAL-FRANCISQUE, H  l  ne DIEYE, Morgane LAVINA, Pierre THOUVENOT, Alexandre LECLERCQ, Marc LECUIT

Microbial Evolutionary Genomics, Institut Pasteur

Mylène MAURY, Marie TOUCHON, Alexis CRISCUOLO, Sylvain BRISSE

Center for information Technology, Institut Pasteur

Louis JONES



Institut Pasteur



Applied Maths, Belgium

Hannes POUSEELE, Bruno POT

Public Health
Agency of Canada

PHAC, Canada

Aleisha REIMER, Matthew WALKER,
Celine NADON



CDC, USA

Peter GERNER-SMIDT, Cheryl TARR
Heather CARLETON, Lee KATZ,
Zuzana KUCEROVA, Steven STROIKA



Oxford University, UK

Keith JOLLEY



SSI, Denmark

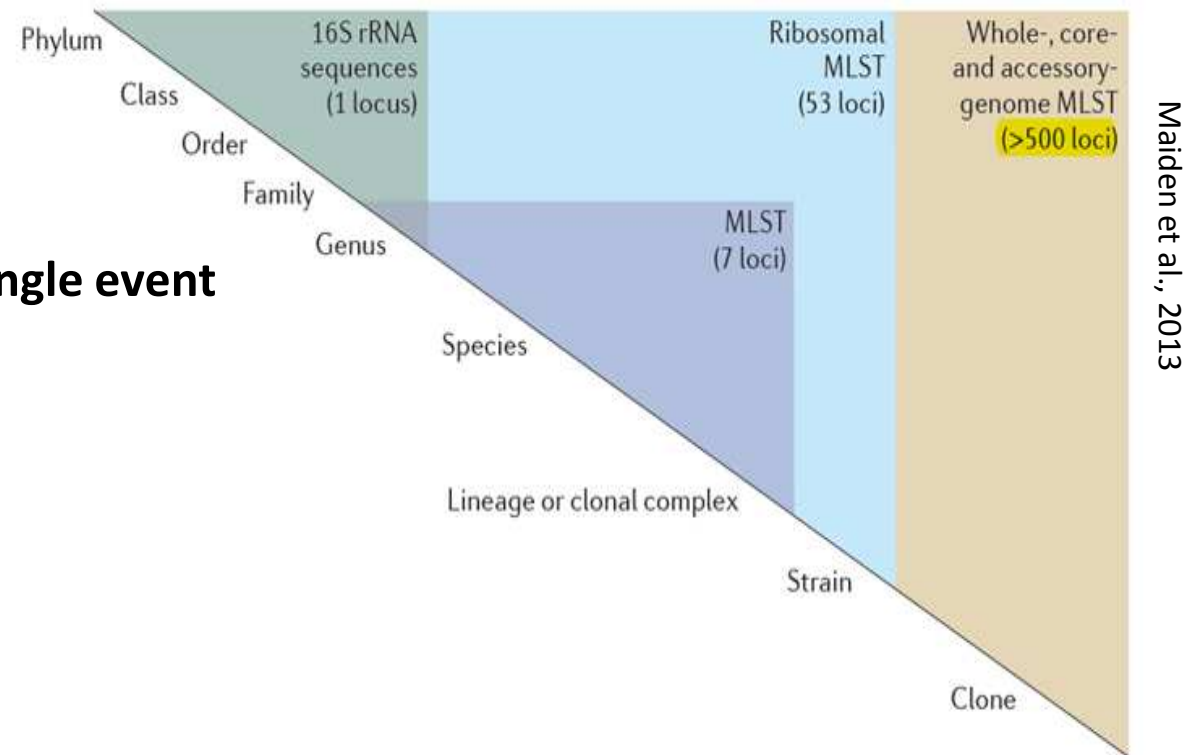
Jonas LARSSON, Eva NIELSEN

Funding:

FONDS NATIONAL SUISSE
DE LA RECHERCHE SCIENTIFIQUE

Whole genome Multilocus Sequence Typing

- **High discriminatory power**
- **Highly reproducible**
- **Recombination & point mutations in a single event**
- **Expandable**
- **Fast and easy to perform**
- **Allows standardised nomenclature**
- **Portable/comparable across different labs**



Whole genome MLST in epidemiological surveillance

Published genome-based MLST schemes

Campylobacter jejunii (2012)
Neisseria meningitidis (2012)
Mycobacterium tuberculosis (2014)
Staphylococcus aureus (2014)
Klebsiella pneumoniae (2014)

Genes 2012, 3, 261-277; doi:10.3390/genes3020261

Communication

A Gene-By-Gene Approach to Bacterial Population Genomics: Whole Genome MLST of *Campylobacter*

Samuel K. Sheppard ^{1,2,†}, Keith A. Jolley ^{1,2} and Martin C. J. Maiden ^{1,2,4}



Whole-Genome-Based *Mycobacterium tuberculosis* Surveillance: a Standardized, Portable, and Expandable Approach

Thomas A. Kohl,^a Roland Diel,^b Dag Harmsen,^c Jörg Rothgänger,^d Karen Meywald Walter,^e Matthias Merker,^a Thomas Weniger,^d Stefan Niemann^{a,f}

Molecular Mycobacteriology, Forschungszentrum Borstel, Borstel, Germany^a; Institute for Epidemiology, Schleswig-Holstein University Hospital, Kiel, Germany^b; Department of Periodontology, University Hospital Münster, Münster, Germany^c; Ridom GmbH, Münster, Germany^d; Public Health Department Hamburg-Central, Hamburg, Germany^e; German Center for Infection Research, Borstel Site, Borstel, Germany^f

Emerging Infectious Diseases • www.cdc.gov/eid • Vol. 20, No. 11, November 2014

Genomic Definition of Hypervirulent and Multidrug-Resistant *Klebsiella pneumoniae* Clonal Groups

Suzanne Bialek-Davenet,¹ Alexis Criscuolo,¹ Florent Ailloud, Virginie Passet, Louis Jones, Anne-Sophie Delannoy-Vieillard, Benoit Garin, Simon Le Hello, Guillaume Arlet, Marie-Hélène Nicolas-Chanoine, Dominique Decré, and Sylvain Brisse



Resolution of a Meningococcal Disease Outbreak from Whole-Genome Sequence Data with Rapid Web-Based Analysis Methods

Keith A. Jolley,^a Dorothea M. C. Hill,^a Holly B. Bratcher,^a Odile B. Harrison,^a Ian M. Feavers,^b Julian Parkhill,^c and Martin C. J. Maiden^a
Department of Zoology, University of Oxford, Oxford, United Kingdom^a; Division of Bacteriology, Blanche Lane, South Mimms, United Kingdom^b; and Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom^c



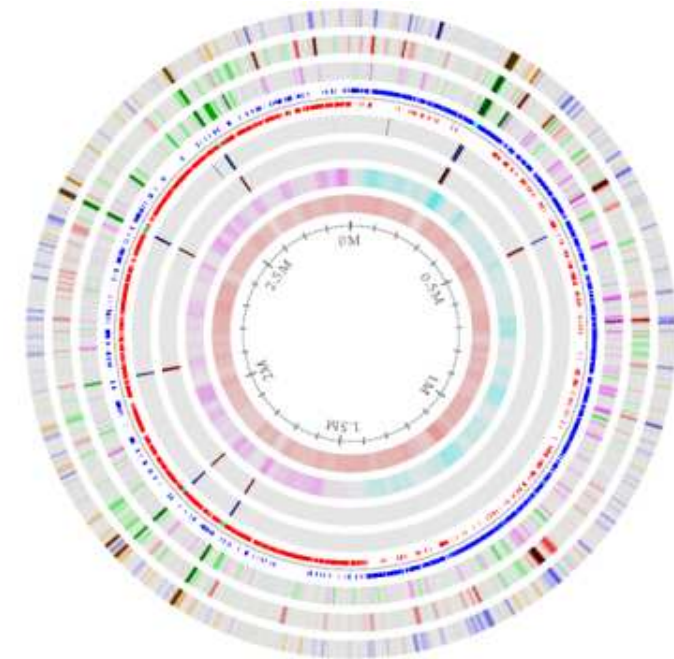
Bacterial Whole-Genome Sequencing Revisited: Portable, Scalable, and Standardized Analysis for Typing and Detection of Virulence and Antibiotic Resistance Genes

Shana R. Leopold,^{a,*} Richard V. Goering,^b Anika Witten,^c Dag Harmsen,^d Alexander Mellmann^a

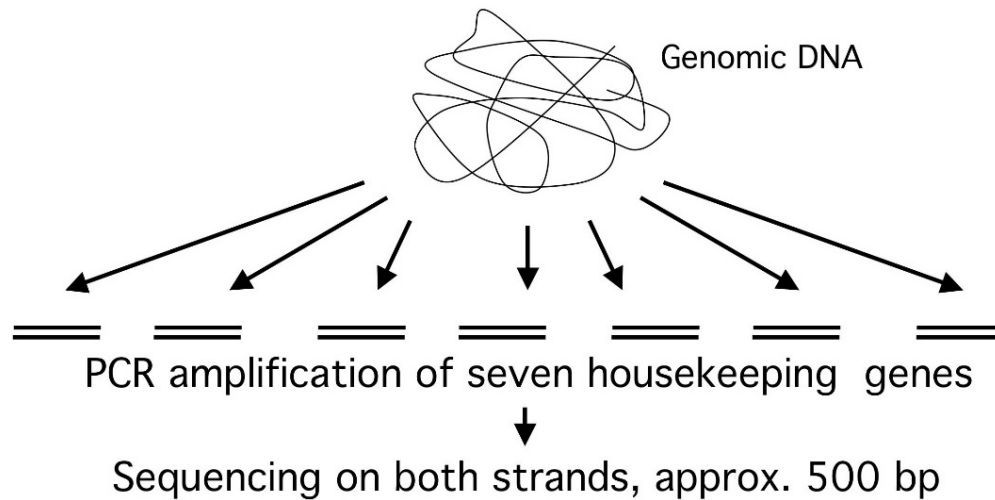
Institute of Hygiene, University of Münster, Münster, Germany^a; Department of Medical Microbiology and Immunology, Creighton University School of Medicine, Omaha, Nebraska, USA^b; Leibniz Institute for Arteriosclerosis, University of Münster, Münster, Germany^c; Department of Periodontology, University of Münster, Münster, Germany^d

Aims

- To establish a **genome-based typing system** for *Listeria monocytogenes* (*Lm*) to be shared on a **global scale**
- To address the **reproducibility** of the proposed system
- To advance on the understanding of *Lm* **global population structure**
- To test its applicability for **outbreak investigation**
- To propose a genome-based **nomenclature system**
- To implement an online **reference *Lm* genome database**

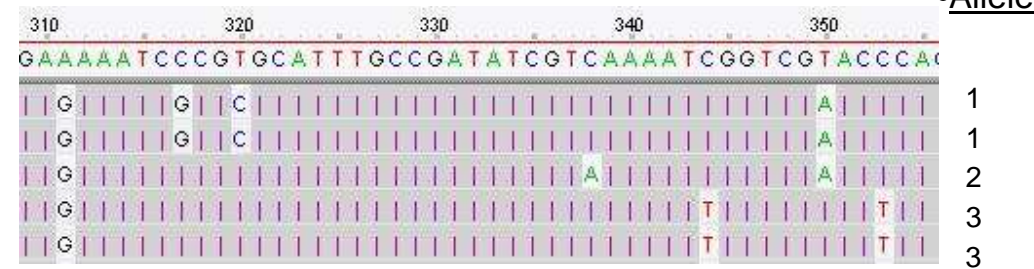


Multilocus Sequence Typing (MLST)



Maiden et al *et al.* 1998

e.g. *abcZ* locus



Loci:

Alleles:

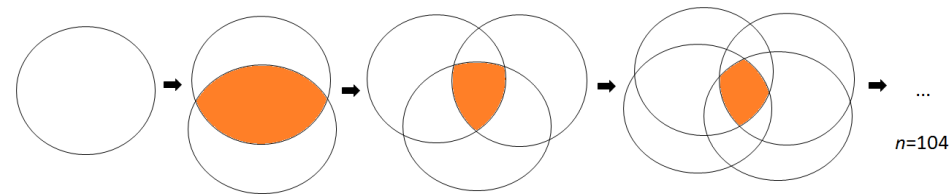
<i>abcZ</i>	<i>bglA</i>	<i>cat</i>	<i>dapE</i>	<i>dat</i>	<i>ldh</i>	<i>lhkA</i>
3	1	1	1	3	1	3



Sequence type (ST)	Alleles						
1	3	1	1	1	3	1	3
2	1	1	11	11	2	1	5
3	4	4	4	3	2	1	5

Lm Core Genome MLST (cgMLST)

Identification of 1791 genes in highly conserved genes in 104 genomes



M. Maury, M. Touchon, S. Brisse

core genome



Define allelic variation

(EGD-e as reference strain)

1791 loci

104 genomes

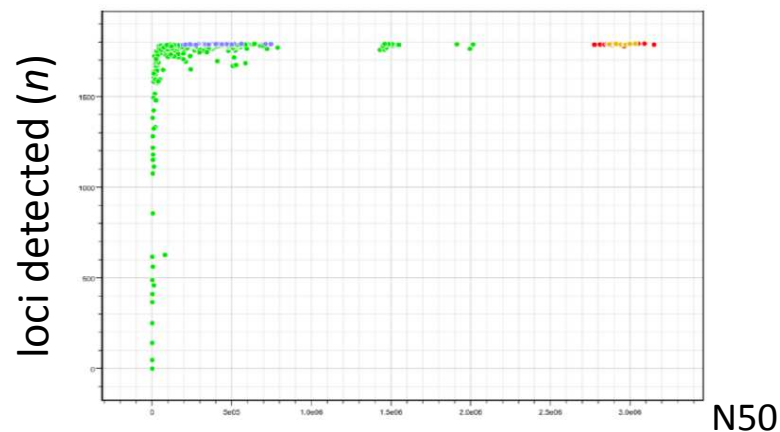
isolate	lmo0002	lmo0003	lmo0005	lmo0006	lmo0007	lmo0009	lmo0010	lmo0011	lmo0012	lmo0015	lmo0016	lmo0018	lmo0019	lmo0020	lmo0021
F2385	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
LL195	2	2	2	4	18	2	2	2	2	2	2	2	2	2	2
LM05-00008	2	2	2	4	2	2	2	2	15	2	2	2	2	2	2
LM05-00172	2	2	2	4	2	2	2	2	15	2	2	2	2	2	2
LM07-00596	2	2	2	4	2	2	2	2	15	2	2	2	2	2	2
LM08-00154	2	2	2	4	2	2	2	2	2	2	2	2	2	2	2
SLCC2378	2	2	2	4	2	2	2	2	2	2	2	2	2	2	2
LM09-00141	2	2	2	4	2	2	2	2	15	2	2	2	2	2	2
LM09-00155	2	2	2	4	2	2	2	2	2	2	2	2	2	2	2
LM07-00648	2	2	2	4	2	2	2	2	15	2	2	2	2	2	2
LM06-01196	2	2	2	4	2	2	2	2	15	2	2	2	2	2	2
LM07-00666	2	2	2	4	18	2	2	2	2	2	2	2	2	2	2
LM07-00523	2	2	2	4	2	2	2	2	2	2	2	2	30	2	2
LM07-00792	2	2	2	4	2	2	2	2	2	2	2	2	30	2	2
LM08-00333	2	25	2	4	2	2	2	2	2	2	2	2	30	2	2
LM08-01429	2	2	2	4	18	2	2	2	2	2	2	2	2	2	2
LM06-01254	2	2	2	4	2	2	2	2	2	2	2	2	30	2	2
LM07-01306	2	2	2	31	2	2	2	2	2	2	2	2	30	2	2
LM07-01317	2	2	2	4	2	2	2	2	2	2	2	2	30	2	2

cgMLST loci validation

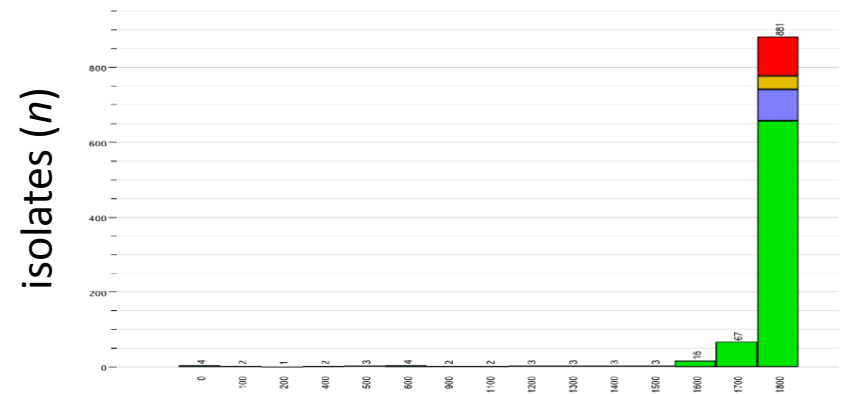
996 *Lm* isolates from 4 surveillance centers



cgMLST loci validation



Number of core loci detected in de novo assembled genomes in function of N50.

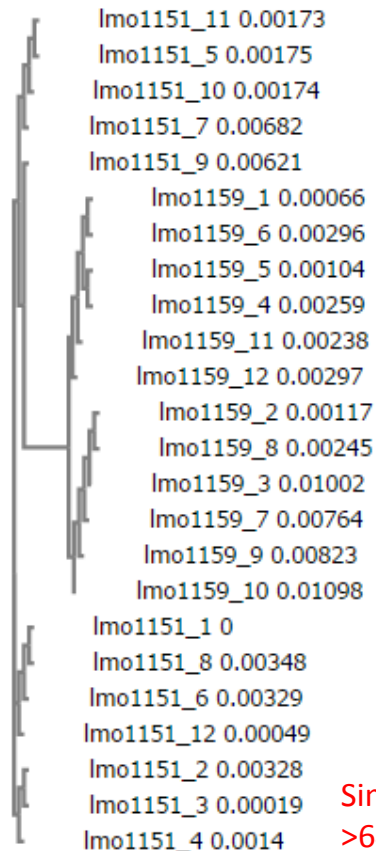


Distribution of the number of loci detected among isolates. Colours indicate the source of the isolates.

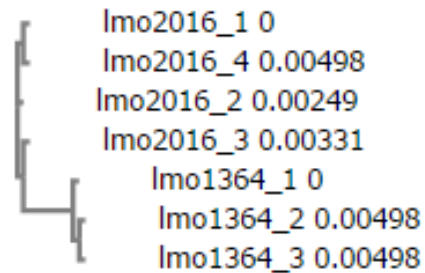
- $n=38$ genomes ($N50 < 2E10^4$) discarded from further analysis
- **No discrepancies in allele callings between raw reads and *de novo* assembled genomes.**

cgMLST loci validation

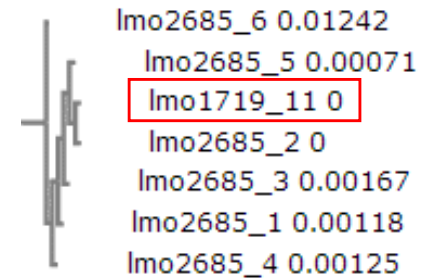
- High inter-locus similarity**



Similarity between locus
>64%



Similarity between locus
>78%



Similarity between locus
>78%

LocusId	LocusId	AvgLen	MinLen	MaxLen	AvgSIBetween
lmo1151	lmo1159	288	288	288	0.638953
lmo1159	lmo1151	276	276	276	0.684963
lmo1364	lmo2016	201	201	201	0.79385
lmo1719	lmo2685	303	303	303	0.788107
lmo2016	lmo1364	201	201	201	0.79385
lmo2685	lmo1719	303	303	303	0.788107

Consequence: one allele of a given locus may fall within the diversity of another locus.

Excluded loci and alleles

excluded loci	<i>n</i>
systematically absent (missing >5.0% in G958)	33
redundant with MLST	4
conflicting detection	6

TOTAL **43**
(2.5% *N*=1791)

excluded alleles	<i>n</i>
non-GATC characters	39
no start codon	15
no stop codon	36
internal/earlier stop	124
frameshifts	284

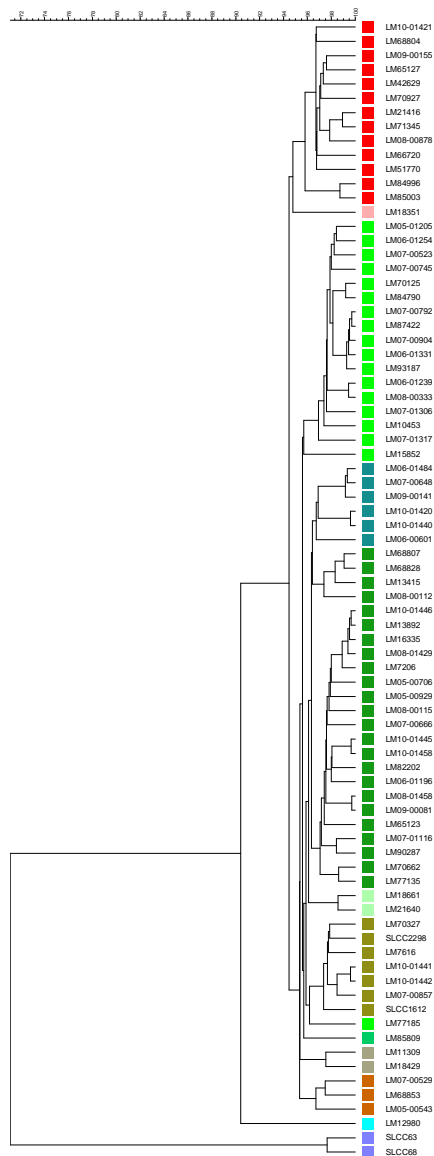
TOTAL **498**
(0.27%, *N*=1791*104)

Final cgMLST scheme with **1748 loci (61.8% of EGD-e genome)**

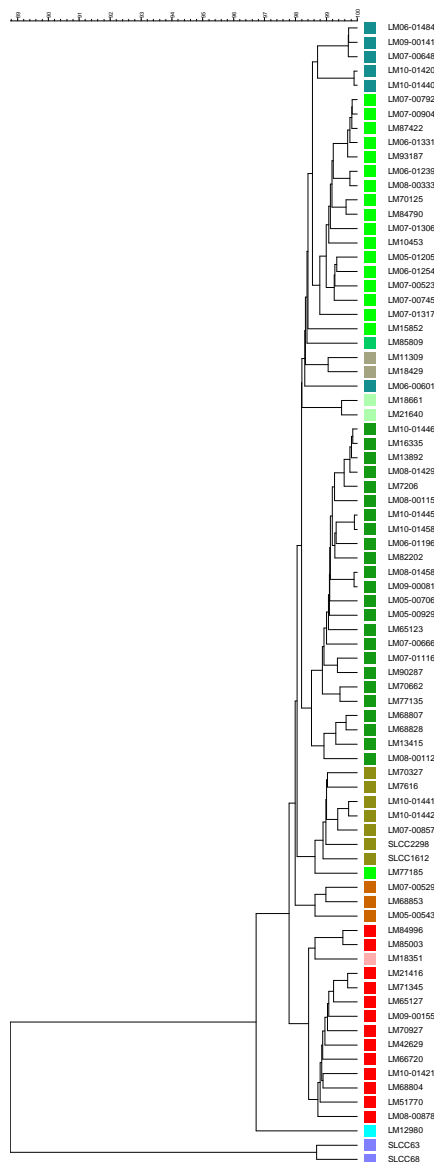
Comparison among methods



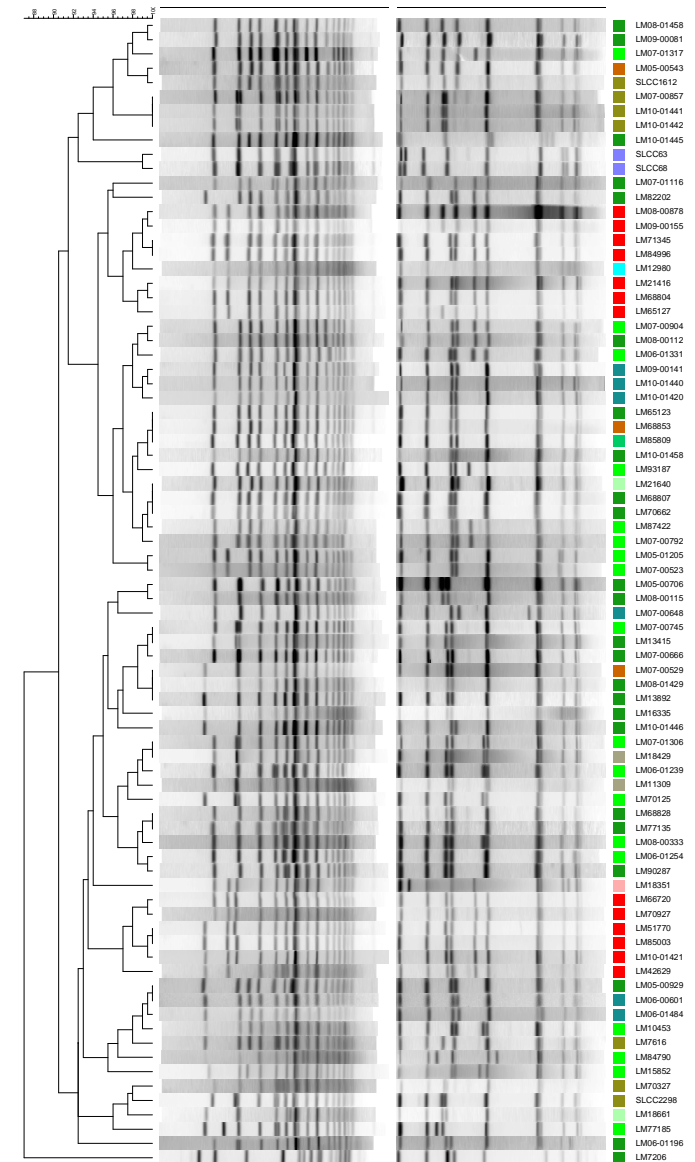
cgMLST



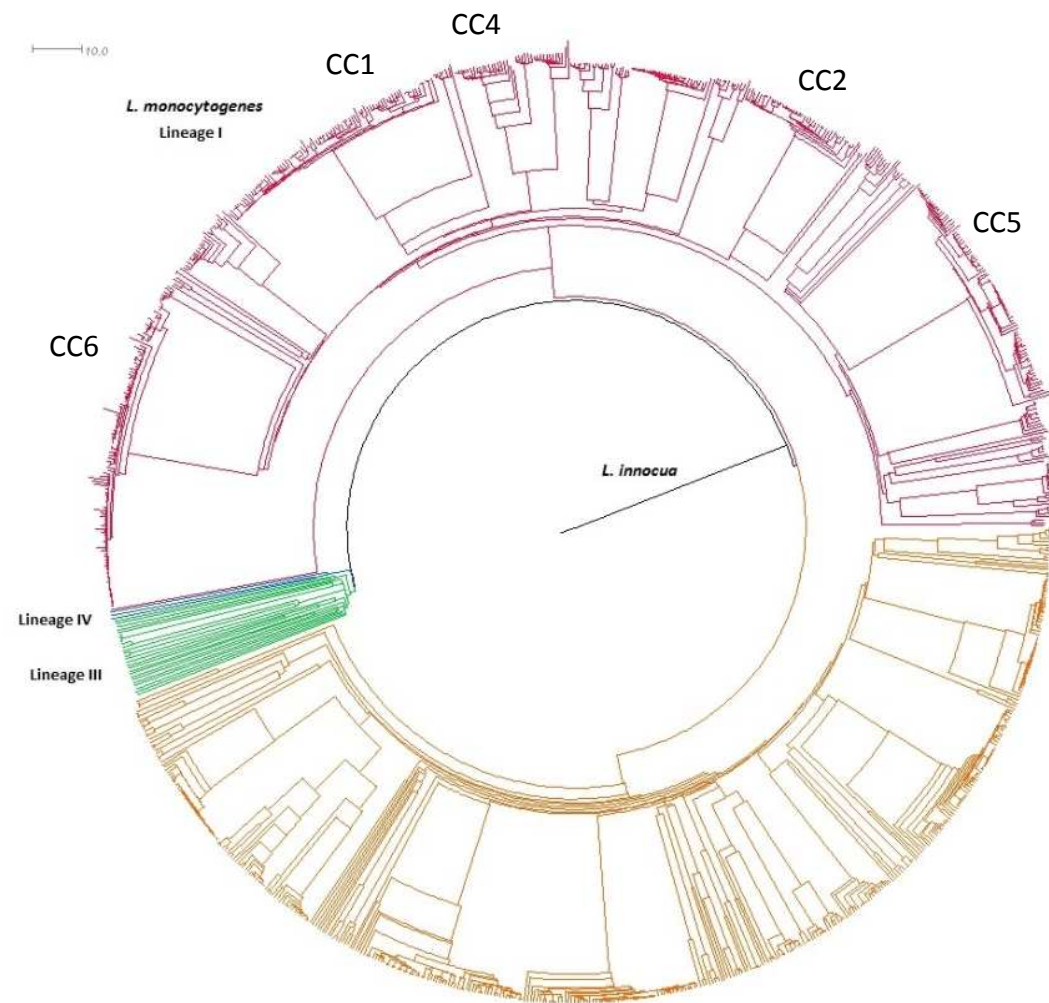
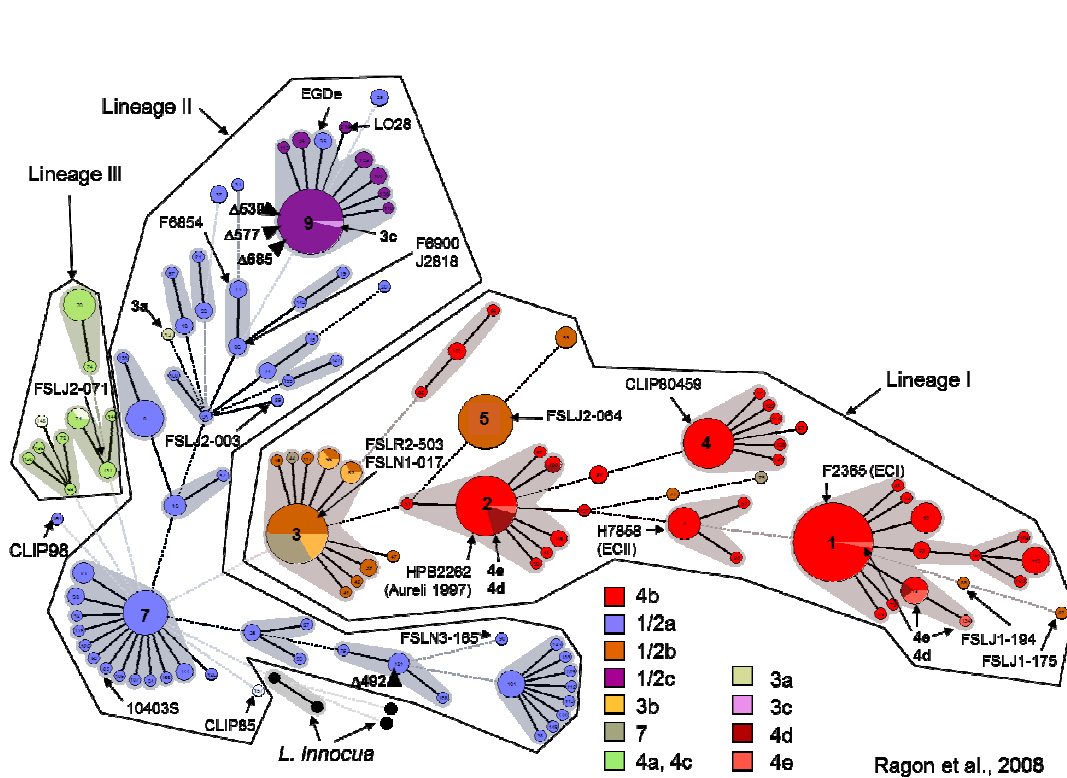
SNPs



PFGE-ApaI-AscI

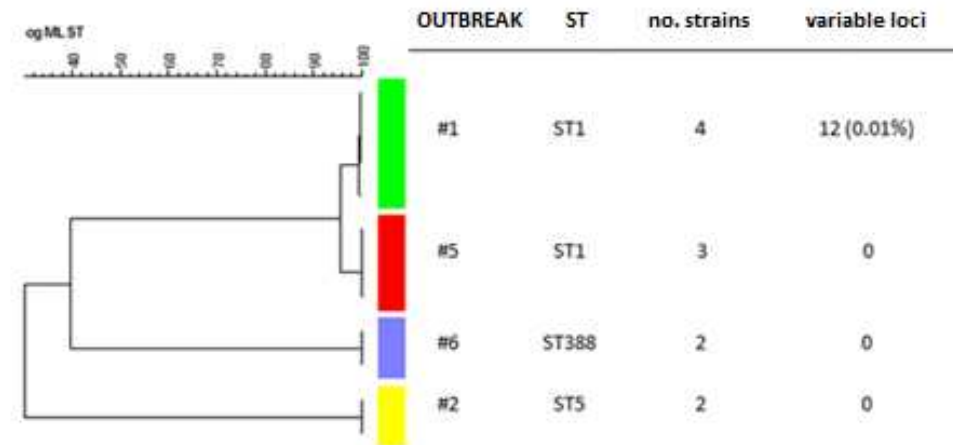


Applications: *Lm* population structure studies



- well delineated clusters corresponding to previously defined CCs

Applications: *Lm* outbreak analysis



Lineage I



Lineage II

cgMLST profiles are **highly conserved within single outbreaks**

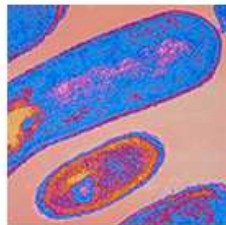
Lm BIGSdb



INSTITUT PASTEUR MLST
and whole genome MLST databases

This page provides access to genotypic data for *L. monocytogenes* isolates based on Multilocus Sequence Typing (MLST), core genome MLST (cgMLST) and ribosomal MLST (rMLST). Please see [references](#) for more details.

Listeria Sequence Typing



[Sequences and profiles database](#)



[Isolates database](#)

Data submission

- [Primers used for amplification and sequencing \(MLST\)](#)
- [Submission of MLST data \(Sanger sequencing\)](#)



Contact

Please submit all data to :
listeriaMLST@pasteur.fr

Data Curation

Data Curation is performed on a voluntary basis and is based on a community effort. Please [contact us](#) if you would like to join the curators team.



[Sequences and Profiles](#)



[Isolates](#)

<http://bigsdB.web.pasteur.fr/listeria>

Lm BIGSdb

Genome sequence query for loci profiling

Sequence query - *Listeria* locus/sequence definitions

Please paste in your sequence to query against the database. Query sequences will be checked first for an exact match against the chosen (or all) loci - they do not need to be trimmed. The nearest partial matches will be identified if an exact match is not found. You can query using either DNA or peptide sequences. [i](#)

Please select locus/scheme — Order results by —

locus

Action

isi

Dropdown menu contents:

- All loci
- MLST
- cgMLST1766
- Virulence
- abcZ
- accA (lmo1572)
- ackA (lmo1581)
- acpD (lmo0611)
- acpP (lmo1806)
- acpS (lmo0885)
- actA (lmo0204)
- addB (lmo2268)
- adeC (lmo1742)
- adk (lmo2611)
- agrA (lmo0051)
- agrC (lmo0050)
- alaS (lmo1504)
- alsS (lmo2006)
- ami (lmo2558)
- ansB (lmo1663)

Lm – BioNumerics v.7.5

The screenshot displays the BioNumerics v.7.5 software interface. The main window shows a list of database entries for Listeria sequences. The selected entry, LMWGLST0000171, is highlighted in yellow. A detailed view of this entry is shown on the right, including a sequence editor, a sequence viewer, a restriction analysis, and a genome map.

Database entries table:

Key	Level	Parent key	Modified date	FastqId	Denovold	location
LMWGLST0000190			2014-11-21 10:34:59	07PF0776	limo.024.c01.14	
LMWGLST0000171			2014-11-21 10:34:59	08-5578	limo.005.c01.14	
LMWGLST0000172			2014-11-21 10:34:59	08-5923	limo.006.c01.14	
LMWGLST0000173			2014-11-21 10:34:59	10403S	limo.007.c01.14	
LMWGLST0000189			2014-11-21 10:34:59	ATCC19117	limo.023.c01.15	
LMWGLST0000170			2014-11-21 10:34:59	CLIP80459 (LM80459)	limo.004.c01.14	
WGMLST_LISTERIA0000001			2014-11-04 11:01:40	EGDe-CNR	limo.001.c01.14	
LMWGLST0000168			2014-11-21 10:34:59	F2365	limo.002.c01.14	
LMWGLST0000191			2014-11-21 10:34:59	F6854	limo.025.c01.08	
LMWGLST0000192			2014-11-21 10:34:59	F6900	limo.026.c01.08	
LMWGLST0000193			2014-11-21 10:34:59	FSLJ1-194	limo.030.c01.08	
LMWGLST0000194			2014-11-21 10:34:59	FSLJ1-208	limo.031.c01.08	
LMWGLST0000195			2014-11-21 10:34:59	FSLN1-017	limo.035.c01.08	
LMWGLST0000196			2014-11-21 10:34:59	FSLN3-165	limo.036.c01.08	
LMWGLST0000197			2014-11-21 10:34:59	FSLR2-503	limo.037.c01.08	
LMWGLST0000175			2014-11-21 10:34:59	FSLR2-561	limo.009.c01.14	
LMWGLST0000169			2014-11-21 10:34:59	HCC23	limo.003.c01.14	
LMWGLST0000198			2014-11-21 10:34:59	HPB2262	limo.039.c01.08	
LMWGLST0000174			2014-11-21 10:34:59	J0161	limo.008.c01.14	
LMWGLST0000199			2014-11-21 10:34:59	J2818	limo.042.c01.08	
LMWGLST0000179			2014-11-21 10:34:59	L312	limo.013.c01.15	
LMWGLST0000178			2014-11-21 10:34:59	L99	limo.012.c01.14	
LMWGLST0000200			2014-11-21 10:34:59	LL195	limo.043.c01.10	
WGMLST_LISTERIA0000018			2014-11-21 10:57:39	LM05-00008	limo.044.c01.09	
WGMLST_LISTERIA0000019			2014-11-04 16:04:07	LM05-00065	limo.045.c01.09	
WGMLST_LISTERIA0000020			2014-11-04 16:04:06	LM05-00172	limo.046.c01.09	
WGMLST_LISTERIA0000021			2014-11-21 10:57:28	LM05-00190	limo.047.c01.09	
WGMLST_LISTERIA0000022			2014-11-21 10:57:16	LM05-00401	limo.048.c01.09	
WGMLST_LISTERIA0000069			2014-11-21 10:57:51	LM05-00704	limo.049.c01.09	
WGMLST_LISTERIA0000002			2014-11-04 11:01:24	LM06-00444	limo.050.c01.09	
WGMLST_LISTERIA0000037			2014-11-04 16:04:43	LM06-00581	limo.051.c01.09	
WGMLST_LISTERIA0000074			2014-11-04 16:04:27	LM06-00727	limo.052.c01.09	
WGMLST_LISTERIA0000023			2014-11-04 16:04:33	LM06-00918	limo.053.c01.09	
WGMLST_LISTERIA0000038			2014-11-04 11:00:26	LM06-00983	limo.054.c01.09	
WGMLST_LISTERIA0000024			2014-11-04 16:04:38	LM06-01023	limo.055.c01.09	
WGMLST_LISTERIA0000009			2014-11-04 11:01:41	LM06-01244	limo.056.c01.09	
WGMLST_LISTERIA0000003			2014-11-04 11:01:23	LM06-01598	limo.057.c01.09	
WGMLST_LISTERIA0000010			2014-11-04 11:01:37	LM06-01814	limo.058.c01.09	
WGMLST_LISTERIA0000008			2014-11-03 16:21:58	LM07-00514	limo.059.c01.09	
WGMLST_LISTERIA0000025			2014-11-04 16:04:45	LM07-00596	limo.060.c01.09	
WGMLST_LISTERIA0000017			2014-11-04 11:01:18	LM07-00876	limo.061.c01.09	
WGMLST_LISTERIA0000067			2014-11-04 16:04:31	LM07-00948	limo.062.c01.09	
WGMLST_LISTERIA0000014			2014-11-04 16:04:37	LM07-01014	limo.063.c01.09	
WGMLST_LISTERIA0000064			2014-11-04 16:04:28	LM07-01042	limo.064.c01.09	
WGMLST_LISTERIA0000026			2014-11-04 16:04:34	LM07-01067	limo.065.c01.09	

Sequence Editor (LMWGLST0000171):

```

ccacaagcc attgtgtgta attaaccact aattgtgtat aagtttaaac 50
taattgaaa ggttatccac aataaaaagg cggtattcag gagttatcca 100
cactttctag gaaaggattt cattgcgcca atgtgttaaa ctatttacog 150
aatacgaaaa aaagacaaat aaatgaggtt gtgaaaaatg atattttcac 200
ggcttttgaa aactagatat acagagtttt atcgagttat acaaaaacag 250
aatattgacg acgtatttgg atacttatta attcacgata aacgggaacc 300
agcagaaatt gacgatttta aggtatttgc aaaaagtaat ataaataaag 350
  
```

Sequence Viewer: A linear view of the sequence with a scale from 0 to 25,000 bp. The sequence is shown as a blue line with various markers and annotations.

Restriction Analysis: A table showing the results of a restriction analysis. The table has columns for Name, Count, Map, Cut1, Fragment, and Cut2. The analysis is performed on the sequence LMWGLST0000171 using the experiment denovo.

Genome: A circular genome map showing the distribution of the sequence across the genome. The map is a circle with a scale from 0 to 3,000,000 bp. The sequence is shown as a blue line with various markers and annotations.

Database entries: A table showing the database entries for the selected sequence. The table has columns for Key, Level, Parent key, Modified date, FastqId, Denovold, and location. The entries are listed in a table with a scrollable list of keys.

Database design: A table showing the database design for the selected sequence. The table has columns for Key, Level, Parent key, Modified date, FastqId, Denovold, and location. The entries are listed in a table with a scrollable list of keys.

Status bar: Database: Listeria-wgMLST_DefaultUser_ Entries: Loaded=102, View=102, Selected=0 4 experiments C:\Users\Public\Documents\BioNumerics7.5\Data\Listeria-wgMLST This is a time limited package valid until 2015-01-31

Lm cgMLST - concluding remarks

- ✓ **Robust and highly discriminatory method** for genome-based typing.
- ✓ Loci were validated on nearly 1,000 genomes of representative *Lm* strains.
- ✓ **Consistent** using both raw reads and *de novo* assembled genomes.
- ✓ Potential highly valuable tool for **outbreak investigation** (validation in progress).
- ✓ Available online through ***Lm* BIGSdb database** <http://bigsdb.web.pasteur.fr/>

The *Lm* cgMLST standard nomenclature of allele numbers and clusters facilitates the international collaboration on research and epidemiology of *Lm* using high-throughput genome sequencing.

Acknowledgments

Biology of Infection Unit & National Reference Center for *Listeria*, Institut Pasteur

Viviane CHENAL-FRANCISQUE, Hélène DIEYE, Morgane LAVINA, Pierre THOUVENOT, Alexandre LECLERCQ, Marc LECUIT

Microbial Evolutionary Genomics, Institut Pasteur

Mylène MAURY, Marie TOUCHON, Alexis CRISCUOLO, Sylvain BRISSE

Center for information Technology, Institut Pasteur

Louis JONES



Applied Maths, Belgium

Hannes POUSEELE, Bruno POT



CDC, USA

Peter GERNER-SMIDT, Cheryl TARR
Heather CARLETON, Lee KATZ,
Zuzana KUCEROVA, Steven STROIKA



SSI, Denmark

Jonas LARSSON, Eva NIELSEN



PHAC, Canada

Aleisha REIMER, Matthew WALKER,
Celine NADON



Oxford University, UK

Keith JOLLEY

Funding:



FONDS NATIONAL SUISSE
DE LA RECHERCHE SCIENTIFIQUE