Science Advances
AAAS

# Supplementary Materials for

## Emergence and global spread of *Listeria monocytogenes* main clinical clonal complex

Alexandra Moura*, Noémie Lefrancq, Thierry Wirth, Alexandre Leclercq, Vítor Borges, Brent Gilpin, Timothy J. Dallman, Joachim Frey, Eelco Franz, Eva M. Nielsen, Juno Thomas, Arthur Pightling, Benjamin P. Howden, Cheryl L. Tarr, Peter Gerner-Smidt, Simon Cauchemez, Henrik Salje, Sylvain Brisse, Marc Lecuit*, *Listeria* CC1 Study Group

*Corresponding author. Email: amoura@pasteur.fr (A.M.); marc.lecuit@pasteur.fr (M.L.)

Published 1 December 2021, *Sci. Adv.* **7**, eabj9805 (2021)
DOI: 10.1126/sciadv.abj9805

**The PDF file includes:**



**Other Supplementary Material for this manuscript includes the following:**

## 1. The *Listeria* CC1 Study Group


Caroline Charlier, Institut Pasteur, French National Reference Center and WHO Collaborating Center Listeria, Paris, France; Institut Pasteur, Biology of Infection Unit, Paris-France; Inserm U1117, Paris, France; Université de Paris, Necker-Enfants Malades University Hospital, Department of Infectious Diseases and Tropical Medicine, AP-HP, Paris, France

Guillaume Vales, Institut Pasteur, French National Reference Center and WHO Collaborating Center *Listeria*, Paris, France

Hélène Bracq-Dieye, Institut Pasteur, French National Reference Center and WHO Collaborating Center Listeria, Paris, France

Nathalie Tessaud-Rita, Institut Pasteur, French National Reference Center and WHO Collaborating Center *Listeria*, Paris, France

Pierre Thouvenot, Institut Pasteur, French National Reference Center and WHO Collaborating Center *Listeria*, Paris, France

Viviane Chenal-Francisque, Institut Pasteur, French National Reference Center and WHO Collaborating Center *Listeria*, Paris, France

Zuzana Kucerova, Centers for Disease Control and Prevention, Atlanta, Georgia, United States

Heather Carleton, Centers for Disease Control and Prevention, Atlanta, Georgia, United States

Steven Stroika, Centers for Disease Control and Prevention, Atlanta, Georgia, United States

Anders Gonçalves da Silva, Microbiological Diagnostic Unit Public Health Laboratory, Department of Microbiology and Immunology, The Doherty Institute for Infection and Immunity, University of Melbourne, Victoria, Australia

Karolina Mercoulia, Microbiological Diagnostic Unit Public Health Laboratory, Department of Microbiology and Immunology, The Doherty Institute for Infection and Immunity, University of Melbourne, Victoria, Australia

Anthony Marius Smith, National Institute for Communicable Diseases, Division of the National Health Laboratory Service, Johannesburg, South Africa.

Jonas T. Björkman, Statens Serum Institut, Copenhagen, Denmark

Anna Oevermann, Division of Neurological Diseases, DCR-VPH, Vetsuisse Faculty, University of Bern, Bern, Switzerland

Lisandra Aguillar-Bultet, Vetsuisse Faculty, University of Bern, Bern, Switzerland

Thijs Bosch, National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Control, Bilthoven, Netherland

Sjoerd Kuiling, National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Control, Bilthoven, The Netherlands

Maaike van den Beld, National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Control, Bilthoven, The Netherlands

Anaïs Painset, Public Health England, London, United Kingdom

Kathie Grant, Public Health England, London, United Kingdom

Leonor Silveira, National Institute of Health Dr. Ricardo Jorge, Department of Infectious Diseases, Lisbon, Portugal

Ângela Pista, National Institute of Health Dr. Ricardo Jorge, Department of Infectious Diseases, Lisbon, Portugal

Mónica Oleastro, National Institute of Health Dr. Ricardo Jorge, Department of Infectious Diseases, Lisbon, Portugal

Sven Halbedel, Consultant Laboratory for *Listeria monocytogenes*, FG11, Robert Koch Institute, Wernigerode, Germany

## 2. Supplementary tables

### Table S1. Genetic characteristics of CC1 sublineages and SL1 genetic clades and statistics by world region

| | | | cgMLST allelic distances | cg1748 SNP distances (1.29 Mb, 11,976 ATGC sites) | | | | | | wgF2365 SNP distances (2.28 M, 29,108 ATGC sites) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *n* | mean ± stdev | mean ± stdev | Tajima's D | Fu & Li's D | Fu & Li's F | nucleotide diversity | haplotype diversity | mean ± stdev | Tajima's D | Fu & Li's D | Fu & Li's F | nucleotide diversity | haplotype diversity |
| Clade | SL1 | 2002 | 68 ± 23.27 | 38 ± 14.8 | -2.69 | -22.40 | -11.24 | 4.33 | 0.966 | 80 ± 27.4 | -2.77 | -26.84 | -12.04 | 63.31 | 0.999 |
| | SL404 | 2 | 57 | 50 | nd | nd | nd | 50.00 | 1.000 | 107 | nd | nd | nd | 107.00 | 1.000 |
| | SL150 | 17 | 42 ± 16.90 | 38 ± 16.8 | -1.46 | -1.41 | -1.50 | 33.41 | 1.000 | 86 ± 30.9 | -1.70 | -1.86 | -1.93 | 79.62 | 1.000 |
| | GC1 | 710 | 39 ± 11.16 | 32 ± 9.5 | -2.78 | -16.87 | -9.79 | 10.13 | 0.992 | 67 ± 15.9 | -2.82 | -18.41 | -10.29 | 49.05 | 0.997 |
| | GC2 | 684 | 51 ± 18.03 | 45 ± 16.4 | -2.70 | -17.18 | -9.92 | 14.92 | 0.991 | 92 ± 30.6 | -2.80 | -17.98 | -10.14 | 48.27 | 0.997 |
| | GC3 | 282 | 47 ± 17.41 | 38 ± 14.8 | -2.63 | -11.92 | -8.32 | 12.79 | 0.991 | 85 ± 25.8 | -2.77 | -12.43 | -8.51 | 65.25 | 0.999 |
| | GC4 | 142 | 46 ± 19.97 | 39 ± 17.3 | -2.59 | -7.73 | -6.29 | 24.68 | 0.990 | 100 ± 29.9 | -2.73 | -8.80 | -7.00 | 86.73 | 0.992 |
| | GC5 | 63 | 34 ± 13.23 | 28 ± 11.3 | -2.38 | -5.66 | -5.13 | 22.38 | 0.995 | 73 ± 23.5 | -2.51 | -5.95 | -5.38 | 68.35 | 0.999 |
| | GC6 | 47 | 45 ± 16.00 | 36 ± 13.6 | -2.22 | -4.33 | -4.13 | 32.66 | 0.988 | 79 ± 25.6 | -2.29 | -4.63 | -4.37 | 73.67 | 0.990 |
| | GC7 | 46 | 28 ± 24.28 | 24 ± 20.8 | -1.60 | -2.77 | -2.71 | 19.97 | 0.966 | 63 ± 51.7 | -2.03 | -4.09 | -3.86 | 58.94 | 0.998 |
| | GC8 | 28 | 29 ± 7.94 | 25 ± 6.7 | -2.39 | -3.32 | -3.41 | 23.18 | 0.997 | 50 ± 10.5 | -2.47 | -3.85 | -3.85 | 45.31 | 1.000 |
| World Region | Africa | 54 | 66 ± 91.55 | 64 ± 148.7 | -2.72 | -7.02 | -6.28 | 41.06 | 0.962 | 146 ± 220.0 | -2.62 | -6.59 | -5.92 | 119.17 | 0.997 |
| | Asia | 25 | 85 ± 119.86 | 100 ± 204.3 | -2.55 | -4.41 | -4.29 | 67.85 | 0.990 | 187 ± 332.3 | -2.49 | -4.38 | -4.24 | 180.76 | 1.000 |
| | Europe | 1236 | 61 ± 24.50 | 53 ± 28.9 | -2.74 | -20.39 | -10.50 | 13.42 | 0.995 | 110 ± 44.9 | -2.78 | -22.24 | -11.05 | 59.04 | 0.999 |
| | North America | 513 | 92 ± 92.77 | 96 ± 160.3 | -2.58 | -11.41 | -7.34 | 28.06 | 0.992 | 194 ± 232.0 | -2.68 | -14.20 | -8.64 | 164.93 | 0.998 |
| | South America | 22 | 63 ± 31.91 | 52 ± 27.6 | -1.39 | -1.66 | -1.72 | 47.01 | 0.987 | 117 ± 55.0 | -1.57 | -2.19 | -2.19 | 114.57 | 0.996 |
| | Oceania | 22 | 63 ± 31.91 | 52 ± 27.6 | -1.39 | -1.66 | -1.72 | 47.01 | 0.987 | 117 ± 55.0 | -1.57 | -2.19 | -2.19 | 114.57 | 0.996 |

## Table S2. Sublineage-specific genes present in at least 50% of isolates

Gene families were determined using a 95% amino acid identity cut-off. Gray shades highlight genes within the same genomic context.

| Remarks | Roary family | Reference locus | Ortholog | Annotation | Length (nt) | No. isolates | % Isolates (in SL) | #Order in contig |
|---|---|---|---|---|---|---|---|---|
| exclusively in SL1 | group_2369 | ID32421_02477 | lmo0671 | hypothetical protein | 293 | 1852 | 93% | |
| exclusively in SL404 | group_7853 | ID31663_01719 | lmo0804 | hypothetical protein | 176 | 2 | 100% | 1067 |
| | group_7852 | ID31663_01117 | LMOf2365_0494 | hypothetical protein | 2150 | 1 | 50% | 6972 |
| exclusively in SL150 | group_897 | ID32037_02878 | not found | hypothetical protein | 260 | 16 | 94% | 1226 |
| | group_6423 | ID32037_02875 | LMOSA_10 | hypothetical protein | 467 | 16 | 94% | 1229 |
| | group_5149 | ID32037_02874 | LMOSA_20 | replication-associated protein | 305 | 14 | 82% | 1230 |
| | group_6471 | ID32037_02873 | LMOf2365_0352 | hypothetical protein | 284 | 12 | 71% | 1231 |
| | group_6421 | ID32037_02872 | HG964404_5465 | hypothetical protein | 314 | 16 | 94% | 1232 |
| | group_5148 | ID32037_02871 | LMOSA_12420 | hypothetical protein | 116 | 12 | 71% | 1233 |
| | group_6420 | ID32037_02870 | lmo0339 | inorganic pyrophosphatase | 371 | 16 | 94% | 1234 |
| | group_6409 | ID32037_00117 | LMOSA_3400 | hypothetical protein | 290 | 14 | 82% | 3282 |
| | group_2611 | ID32037_00115 | lmo2044 | peptide ABC transporter substrate-binding protein | 1664 | 13 | 76% | 3284 |
| | group_6408 | ID32037_00121 | lmo2749 | glutamine amidotransferase | 572 | 16 | 94% | 3302 |
| | group_6406 | ID32037_00123 | lmo2375 | hypothetical protein | 395 | 16 | 94% | 3304 |
| | group_6417 | ID32037_01337 | not found | hypothetical protein | 110 | 16 | 94% | 4171 |
| | group_6418 | ID32037_01335 | lmo2688 | cell division protein FtsW | 1130 | 16 | 94% | 4175 |
| exclusively in SL150 & SL404 (absent in SL1) | recD | ID32037_02916 | LMOSA_12110 | DNA helicase; RecBCD enzyme subunit RecD | 1355-3608 | 18 | 95% | 2391 |
| | group_5134 | ID32037_02915 | not found | hypothetical protein | 1061-1340 | 18 | 95% | 2392 |
| | group_6411 | ID32037_02914 | lmo0303 | putative secreted, lysin rich protein | 551 | 18 | 95% | 2393 |
| | group_6412 | ID32037_02912 | lmo0305 | L-allo-threonine aldolase | 1082 | 18 | 95% | 2395 |
| | group_6413 | ID32037_02911 | lmo0306 | hypothetical protein | 467 | 18 | 95% | 2396 |
| | group_6415 | ID32037_02907 | lmo0310 | hypothetical protein | 1076 | 17 | 89% | 2400 |
| exclusively in SL1 & SL404 (absent in SL150) | group_152 | ID32421_02841 | LMOf2365_0349 | cell wall surface anchor family protein (LPxTG motif) | 293-3221 | 1380 | 69% | 2433 |
| | group_1481 | ID32421_01841 | LMOf2365_2341 | aminotransferase, class I | 221-1166 | 1936 | 97% | 3295 |
| | group_4436 | ID32421_01836 | lmo2375 | hypothetical protein | 263-392 | 2004 | 100% | 3309 |
| | group_378 | ID32421_01835 | I586_01601 | reverse transcriptase | 302-1385 | 1977 | 99% | 3310 |
| | group_1844 | ID32421_00303 | lmo2688 | cell division protein FtsW | 758-1130 | 1009 | 50% | 4177 |
| | group_1501 | ID32421_00308 | LMOf2365_2670 | N-acetylmuramoyl-L-alanine amidase, family 4 | 1100-1775 | 1450 | 72% | 4183 |
| exclusively in SL1 & SL150 (absent in SL404) | group_1153 | ID32421_02877 | lmo0297 | transcriptional antiterminator BglG | 593-1871 | 1991 | 99% | 2373 |
| | sau3AIR | ID32421_02872 | AJXO01000022_949 | Type-2 restriction enzyme Sau3AI | 152-1667 | 1990 | 99% | 2379 |
| | group_4596 | ID32421_02871 | LMOf2365_0326 | transcriptional regulator | 164-206 | 2000 | 99% | 2382 |
| | group_1899 | ID32421_02870 | LMOf2365_0327 | cytosine-specific methyltransferase | 131-1409 | 1988 | 98% | 2384 |
| | group_1900 | ID32421_02869 | LMOf2365_0328 | hypothetical protein | 236-854 | 1952 | 97% | 2386 |
| | group_1572 | ID32421_02868 | LMOf2365_0329 | putative lipoprotein | 197-554 | 1993 | 99% | 2387 |
| | group_1901 | ID32421_02867 | LMOf2365_0330 | threonine aldolase | 305-1079 | 2000 | 99% | 2388 |
| | group_3681 | ID32421_02866 | LMOf2365_0331 | peptidase, M48 family | 464-920 | 2001 | 99% | 2389 |
| | group_1342 | ID32421_02360 | lmo0804 | hypothetical protein | 155-959 | 1997 | 99% | 3058 |
| | group_5703 | ID32037_01341 | not found | hypothetical protein | 89 | 1021 | 51% | 4167 |
| | group_2203 | ID32421_00342 | Bcer98_3982 | hypothetical protein | 329-632 | 1999 | 99% | 4215 |
| | group_596 | ID32421_00343 | not found | hypothetical protein | 248-1286 | 1837 | 91% | 4216 |
| | group_791 | ID32421_00344 | not found | hypothetical protein | 455-983 | 1787 | 89% | 4218 |
| | group_5708 | ID32421_00345 | lwe2675 | hypothetical protein | 299 | 1991 | 99% | 4222 |
| | group_5709 | ID32421_00346 | lwe2675 | hypothetical protein | 359-359 | 1995 | 99% | 4223 |
| | group_2733 | ID32421_00347 | JL53_14750 | hypothetical protein | 407-773 | 1999 | 99% | 4224 |
| | group_3456 | ID32421_00351 | lmo2724 | 3-demethylubiquinone-9 3-methyltransferase | 323-443 | 1986 | 98% | 4229 |
| | group_2255 | ID32421_02154 | LMOf2365_0239 | dihydrouridine synthase family protein | 209-995 | 1585 | 79% | 4765 |
| | group_81 | ID32421_00680 | LMOf2365_0495 | putative lipoprotein | 155-2159 | 1244 | 62% | 6967 |
| | group_2830 | ID32421_00615 | lmo2084 | aminoglycoside phosphotransferase | 455-476 | 1862 | 92% | 7056 |
| | group_3057 | ID32421_01316 | lmo1343 | competence protein ComGE | 284-284 | 2004 | 99% | 8031 |
| | group_1982 | ID32421_01310 | LMOf2365_1365 | glycine cleavage system T protein GcvT | 797-1088 | 2009 | 100% | 8039 |
| | group_555 | ID32421_02213 | lmo1721 | transcriptional regulator | 788-2771 | 1943 | 96% | 8392 |
| | group_2258 | ID32037_00061 | LMOf2365_1741 | transcriptional regulator, TetR family | 188-584 | 1950 | 97% | 8398 |
| | group_1540 | ID32421_02218 | lmo1715 | methyltransferase | 185-668 | 1940 | 96% | 8399 |
| | group_1048 | ID32421_00984 | lmo0738 | PTS beta-glucoside transporter subunit IIABC | 698-1448 | 1879 | 93% | 8654 |
| | group_5905 | ID32421_00985 | lmo0116 | hypothetical protein_lmaC_phageA118 | 167 | 2015 | 100% | 8657 |
| | group_1669 | ID32421_00989 | lmo1655 | vanZ-like protein | 128-563 | 1889 | 94% | 9030 |

**Table S3. SL1 genetic clades-specific genes present in at least 50% of isolates**

Clades-specific genes were only found in GC3 and GC7. Gray shadows highlight genes within the same genomic context.

| Remarks | Roary_family | Reference locus | Ortholog | Annotation | Length (nt) | No. isolates | % Isolates (in GC) | #Order in contig |
|---|---|---|---|---|---|---|---|---|
| exclusively in GC3 | group_1376 | ID106_01313 | lmo1115 | fibrinogen-binding protein | 1562-2693 | 215 | 76% | 7335 |
| | group_5934 | ID106_01309 | lmo1111 | hypothetical protein | 1301 | 271 | 96% | 7345 |
| | group_5933 | ID106_01308 | lmo1108 | hypothetical protein | 485 | 271 | 96% | 7346 |
| | group_4885 | ID106_01307 | JQLU01000002_899 | hypothetical protein | 212-374 | 271 | 96% | 7347 |
| | group_5932 | ID106_01306 | JQLU01000002_900 | hypothetical protein | 1019 | 271 | 96% | 7348 |
| | group_5931 | ID106_01305 | LMOSA_2380 | hypothetical protein | 338 | 271 | 96% | 7349 |
| | group_4884 | ID106_01304 | JQLU01000002_902 | hypothetical protein | 269-686 | 271 | 96% | 7350 |
| | group_3147 | ID106_01302 | lmo1105 | hypothetical protein | 1253-1769 | 269 | 95% | 7356 |
| | group_3146 | ID106_01301 | lmo1104 | P60 protein | 812-1025 | 271 | 96% | 7358 |
| | group_2504 | ID106_01300 | lmo1100 | cadmium resistance protein CadA | 1403-2105 | 268 | 95% | 7359 |
| | group_4029 | ID106_01299 | lmo1102 | cadmium efflux system accessory protein CadC | 236-356 | 267 | 95% | 7360 |
| | group_5930 | ID106_01298 | LMOf2365_0702 | ABC transporter- permease protein | 770 | 270 | 96% | 7361 |
| | group_4883 | ID106_01297 | LMOf2365_0701 | ABC transporter- ATP-binding protein | 278-935 | 271 | 96% | 7362 |
| | group_5929 | ID106_01296 | ATXL01000015_1806 | hypothetical protein | 173 | 269 | 95% | 7363 |
| | group_5928 | ID106_01295 | LMOf2365_1076 | dihydrolipoamide dehydrogenase | 1673 | 271 | 96% | 7364 |
| | acr3_2 | ID106_01293 | AEDB02000104_3834 | Arsenical-resistance protein Acr3 | 1076 | 271 | 96% | 7370 |
| | group_5926 | ID106_01291 | lmo2493 | ArsR family transcriptional regulator | 365 | 271 | 96% | 7379 |
| | arsD_1 | ID106_01290 | HMPREF9320_1672 | Arsenical resistance operon trans-acting repressor ArsD | 371 | 271 | 96% | 7380 |
| | group_5924 | ID106_01289 | lmo1102 | cadmium efflux system accessory protein CadC | 293 | 271 | 96% | 7381 |
| | arsA_2 | ID106_01288 | LMOSA_2220 | Arsenical pump-driving ATPase | 263-1739 | 271 | 96% | 7382 |
| | arsD_2 | ID106_01287 | ADIAL_0134 | Arsenical resistance operon trans-acting repressor ArsD | 311 | 270 | 96% | 7383 |
| | group_5922 | ID106_01286 | lmo1679 | cystathionine beta-lyase | 1142-1142 | 270 | 96% | 7384 |
| | group_5921 | ID106_01285 | LMOSA_2200 | hypothetical protein | 458 | 254 | 90% | 7385 |
| | group_4882 | ID106_01284 | LMOSA_2190 | hypothetical protein | 302-404 | 256 | 91% | 7386 |
| exclusively in GC7 | group_8396 | ID32182_02420 | LMOf2365_0506 | hypothetical protein | 326 | 42 | 91% | 1150 |

**Table S4. Human-associated significant loci**

Determined using treeWAS, using with a significance threshold of $p<10^{-5}$. Gray backgrounds highlight groups of adjacent genes.

| Association type | Roary family | Reference locus | Ortholog | Annotation | #Order in contig | Length (nt) | treeWAS score | No. isolates | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | G1P1 | G0P0 | G1P0 | G0P1 |
| Positive | group_1527 | ID32421_02693 | not found | hypothetical protein | 2637 | 281-305 | 24 | 1007 | 186 | 382 | 446 |
| | group_1926 | ID32421_02503 | LMOSA_15540 | transcriptional regulator | 2878 | 407-647 | 25 | 1024 | 193 | 375 | 429 |
| | group_706 | ID32421_02501 | LMOSA_15560 | hypothetical protein | 2883 | 977 | 25 | 1020 | 191 | 377 | 433 |
| | group_497 | ID32421_02498 | ATXL01000015_1796 | hypothetical protein | 2886 | 560 | 24 | 1029 | 190 | 378 | 424 |
| | group_209 | ID32421_00679 | LMOSA_130 | hypothetical protein | 6973 | 1530 | 24 | 866 | 254 | 314 | 587 |
| | group_10387 | ID31663_01231 | - | tRNA-Glu(ttc) | 7202 | 72 | 25 | 1041 | 181 | 387 | 412 |
| | group_1038 | ID32421_01586 | LMOSA_19640 | N-acetylmuramoyl-L-alanine amidase | 7651 | 218 | 26 | 1265 | 96 | 472 | 188 |
| | group_1361 | ID32421_01563 | I571_03072 | hypothetical protein | 7720 | 870 | 24 | 1303 | 88 | 480 | 150 |
| | group_619 | ID32421_01561 | I580_00527 | hypothetical protein | 7722 | 1262-1313 | 23 | 1175 | 138 | 430 | 278 |
| | group_2465 | ID32421_01095 | LMOSA_24740 | valyl-tRNA synthetase valS | 9148 | 2649 | 22 | 1300 | 89 | 479 | 153 |
| Negative | group_6404 | ID31019_03121 | MGMO_205c00040 | hypothetical protein | 9 | 116 | -22 | 36 | 518 | 50 | 1417 |
| | group_533 | ID31797_00694 | LMOSA_22340 | hypothetical protein | 20 | 245-1199 | -22 | 74 | 508 | 60 | 1379 |
| | group_499 | ID31663_02416 | ATXL01000015_1796 | hypothetical protein | 1102 | 539-2138 | -23 | 404 | 397 | 171 | 1049 |
| | group_10094 | ID31663_01817 | LMOSA_13730 | hypothetical protein | 1157 | 485 | -28 | 492 | 362 | 206 | 961 |
| | group_10662 | ID31663_01818 | LMOSA_13720 | hypothetical protein | 1156 | 473 | -30 | 479 | 358 | 210 | 974 |
| | group_60 | ID31663_01819 | lmo0460 | putative lipoprotein | 1155 | 206-2231 | -27 | 188 | 476 | 92 | 1265 |
| | group_1927 | ID31663_01829 | LMOSA_15540 | transcriptional regulator | 2879 | 209-647 | -25 | 430 | 375 | 193 | 1023 |
| | group_10488 | ID31019_00426 | - | 5S ribosomal RNA | 3190 | 109 | -41 | 282 | 399 | 169 | 1171 |
| | group_10162 | ID31019_00425 | - | tRNA-Asn(gtt) | 3195 | 75 | -29 | 461 | 309 | 259 | 992 |
| | group_10390 | ID31663_02352 | - | tRNA-Glu(ttc) | 3605 | 72 | -33 | 532 | 276 | 292 | 921 |
| | group_6398 | ID31663_02351 | - | tRNA-Val(tac) | 3606 | 75 | -25 | 611 | 246 | 322 | 842 |
| | group_4211 | ID31019_03012 | - | 5S ribosomal RNA (partial) | 8331 | 85 | -37 | 227 | 398 | 170 | 1226 |
| | group_10476 | ID31797_01719 | - | 5S ribosomal RNA | 8587 | 108 | -31 | 520 | 253 | 315 | 933 |
| | group_10432 | ID31797_01722 | - | tRNA-Lys(ttt) | 8593 | 73 | -35 | 499 | 289 | 279 | 954 |

G, genome; P, phenotype; 0 absent, 1 present.

**Table S5. Time and geographic origin of the most recent common ancestors (MRCA) of CC1 clades estimated in this study**

The 95% high posterior density (HDP) confidence intervals of the time of MRCA estimates and the ancestral state probabilities for the most probable origin (Nam, North America, Eur, Europe) are provided in brackets. g200 and g1972 refer to the number of CC1 genomes used in each analysis (*n*=200 and *n*=1,972, respectively). Stars denote clades with different ancestral states depending the reconstruction method (BEAST vs PastML).

| Clade | MRCA [95% HPD] (year) | Most probable ancestor state (probabability) | | |
|---|---|---|---|---|
| | | g200.BEAST | g200.PastML | g1972.PastML |
| CC1 (root) | 197 [860 BC - 1045 AD] | NAm (0.26) | NAm (0.57) | NAm (0.58) |
| *Sublineage* | | | | |
| SL1* | 1857 [1821-1888] | Eur (0.94) | NAm (0.89) | NAm (0.55) |
| SL150 | 1823 [1749-1884] | NAm (0.96) | NAm (0.59) | NAm (0.95) |
| SL404 | 1927 [1873-1968] | NAm \| Eur (0.5) | NAm \| Eur (0.5) | NAm \| Eur (0.5) |
| *SL1 clade* | | | | |
| GC1 | 1920 [1893-1940] | Eur (1.00) | Eur (0.99) | Eur (0.85) |
| GC2* | 1885 [1853-1909] | Eur (0.86) | NAm (0.93) | NAm (0.93) |
| GC3 | 1904 [1838-1905] | Eur (0.71) | Eur (0.96) | Eur (0.98) |
| GC4* | 1888 [1848-1910] | Eur (0.98) | NAm (0.81) | NAm (1.00) |
| GC5* | 1934 [1911-1950] | Eur (0.99) | NAm (0.43) | NAm (0.75) |
| GC6 | 1891 [1873-1908] | Eur (1.00) | Eur (0.82) | Eur (1.00) |
| GC7 | 1917 [1882-1948] | NAm (0.84) | NAm (0.96) | NAm (1.00) |
| GC8 | 1934 [1906-1956] | Eur (1.00) | Eur (0.98) | Eur (1.00) |

## 2. Supplementary figures

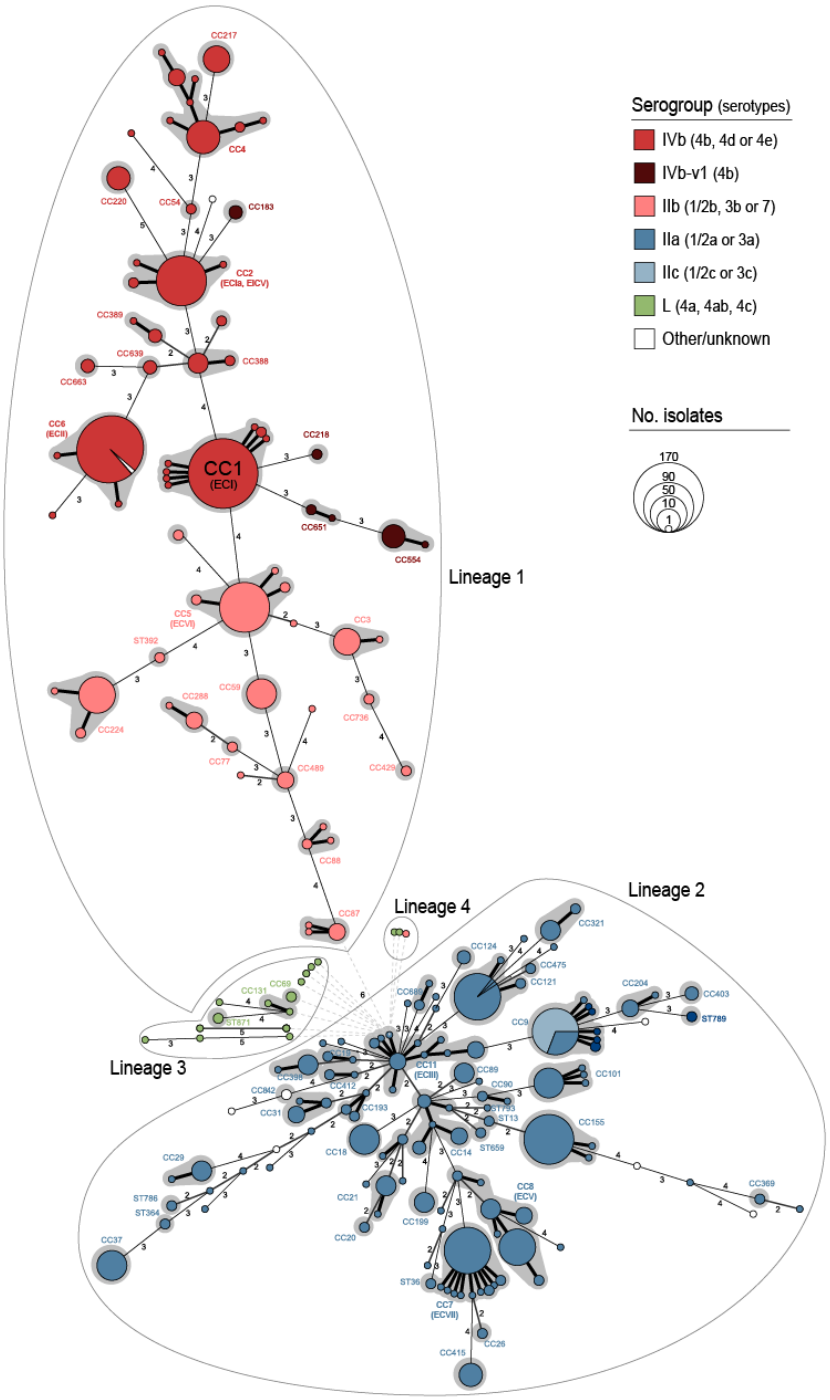


**Figure S1. *Listeria monocytogenes* clonal complex 1 (CC1) placement within the species population diversity**

Minimum spanning tree based on the allelic profiles of the 7-locus multilocus sequence typing scheme (Ragon *et al.*, 2008) obtained in previous published work (Moura et *al.*, 2016). Sequence types are represented by circles proportional to the number of isolates, colored by serogroup. Sequence types comprising more than 2 isolates and belonging to the same clonal complex (CC) are delimited by gray zones and labeled. Numbers in connecting lines denote the number of allelic differences between profiles. For simplicity, allelic differences of 1 (thick branches) and 6 (dotted branches) were omitted.


**Figure S2. Frequency of most prevalent clonal complexes among different environments**

Data collected based on 29,349 *L. monocytogenes* genomes with associated source metadata available on NCBI Sequence Read Archive (as of October 23rd, 2020). MLST typing was performed from reads using the srst2 v.0.1.5 software (http://katholt.github.io/srst2) and the BIGSdb-*Lm* profiles database (https://bigsdb.pasteur.fr/listeria). The prevalence of *Lm*-CC1 in human clinical isolates and animal environments was significantly higher as compared to food and natural environments ($p<1.69x10^{-17}$, Bonferroni-adjusted pairwise comparisons of proportions).

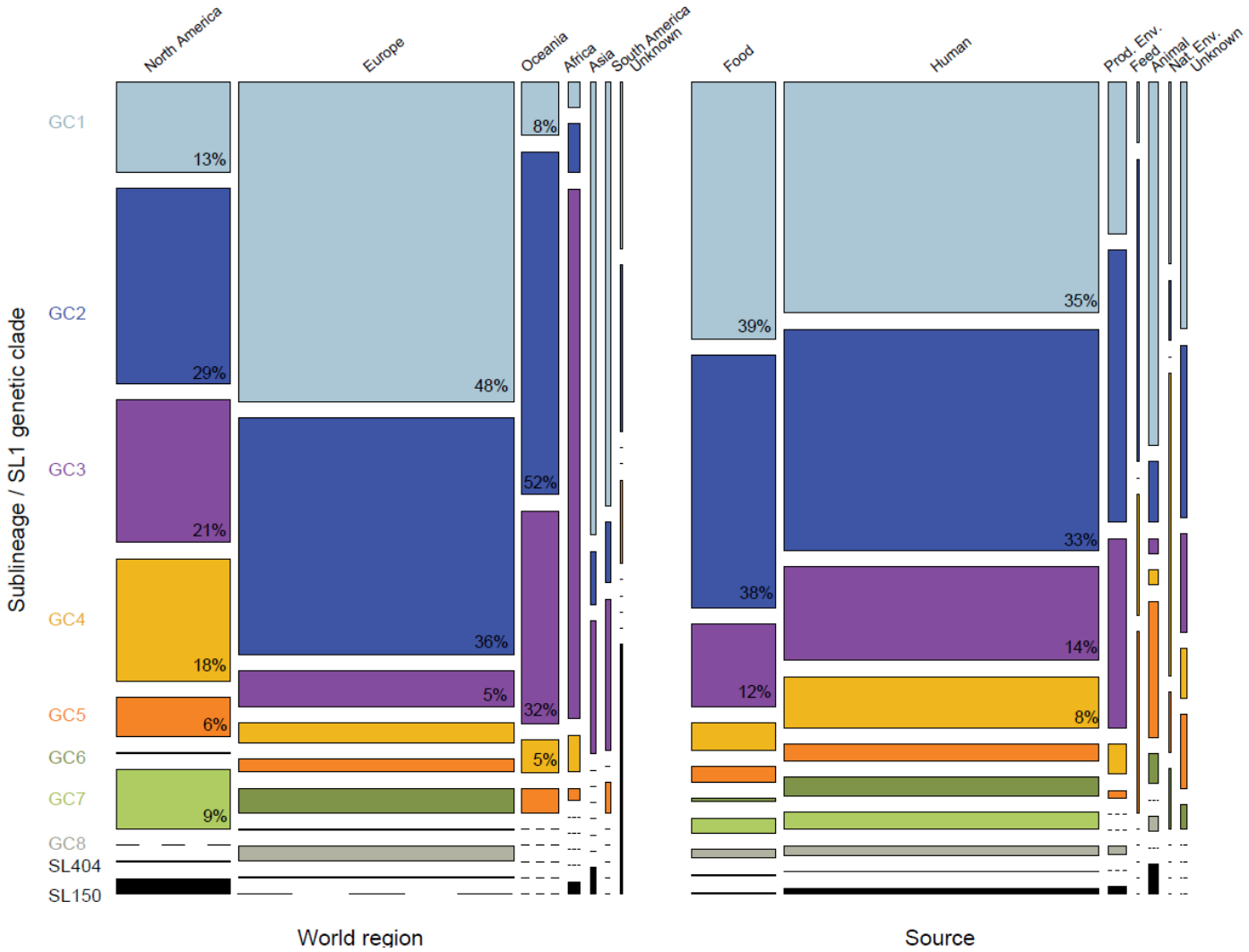


**Figure S3. Distribution of *Lm*-CC1 isolates per clade, world regions and source types (*N*=2,021)**

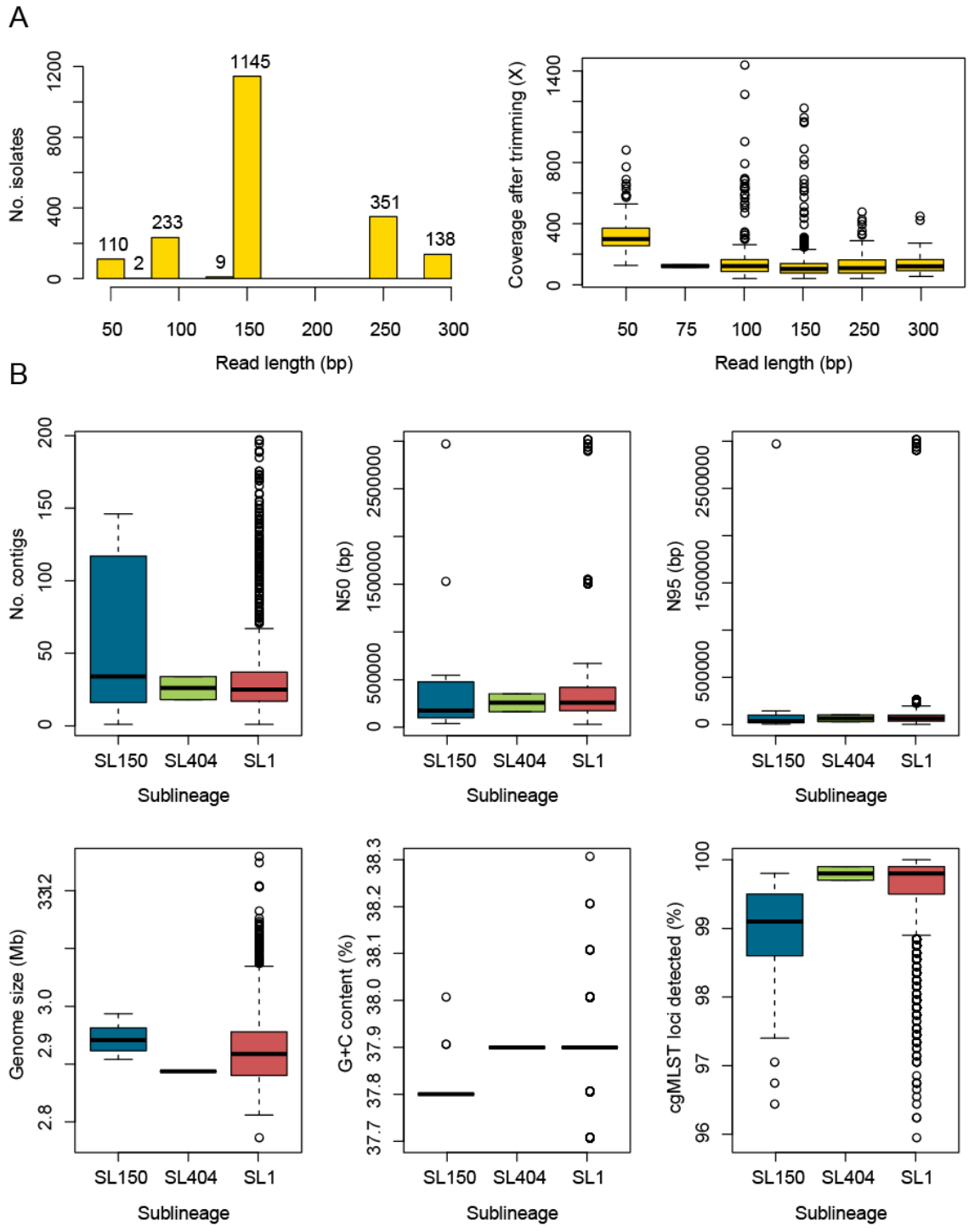


**Figure S4. Genome metrics of isolates included in this study**

**(A)** Distribution of isolates per sequence read length (left) and distribution of sequencing coverages after reads quality trimming (right). **(B)** Assembly metrics per CC1 sublineages, based on the number of contigs, N50 and N95 contig lengths, genome size, G+C content and cgMLST loci detected.

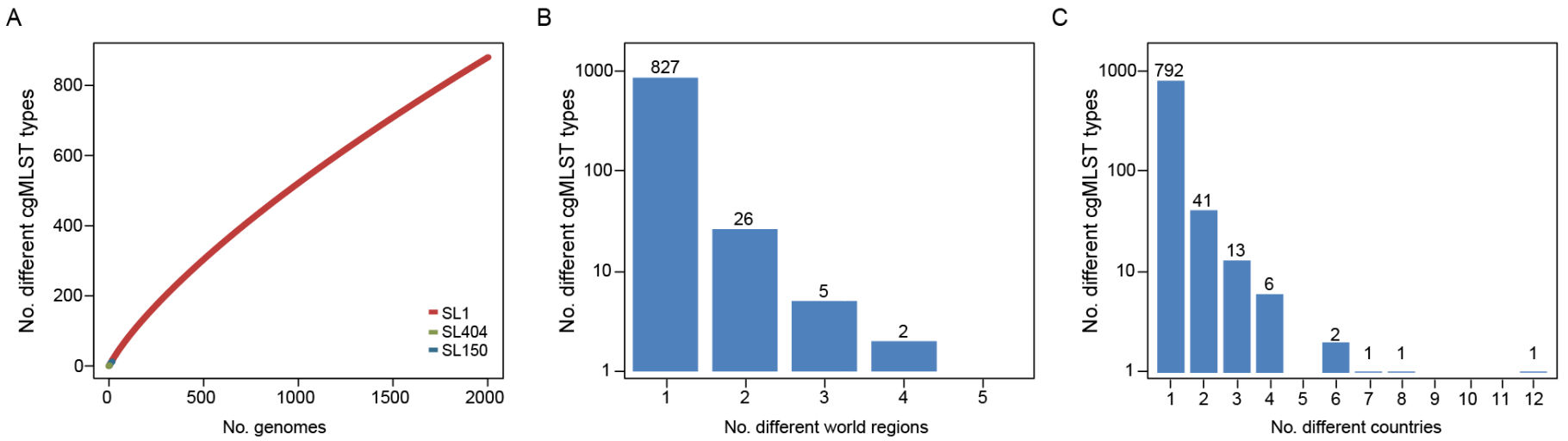


**Figure S5. Core genome multilocus sequence typing (cgMLST) analyses**

**(A)** Rarefaction analysis of cgMLST types sampled per sublineage. **(B)** Number of SL1 cgMLST types per number of different world regions in which they were observed (*n*=860 types with world region information). **(C)** Number of SL1 cgMLST types per number of different countries in which they were observed (*n*=857 types with country information).

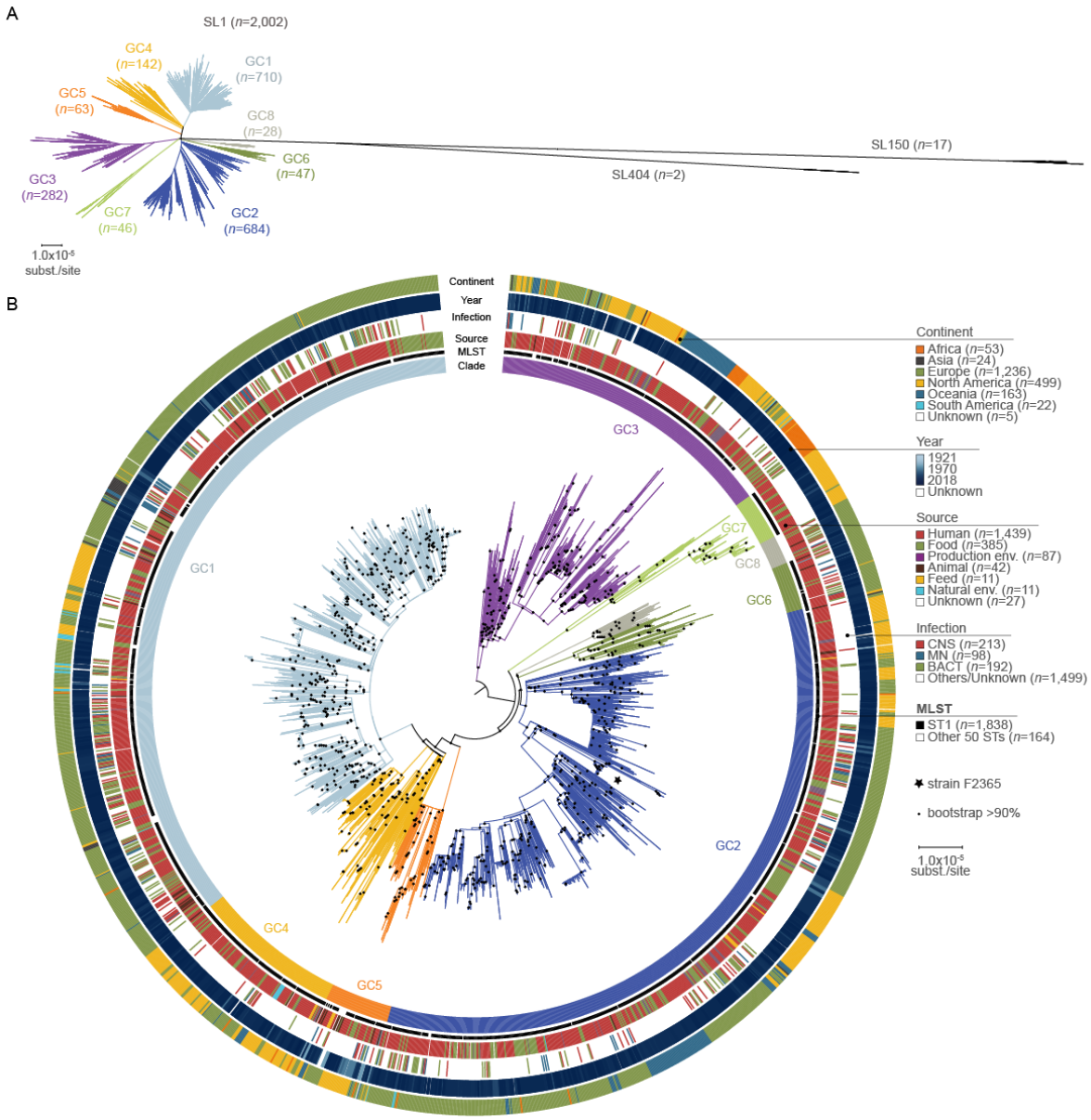


**Figure S6. Phylogenetic analysis based on whole genome SNP analyses**

**(A)** Unrooted maximum-likelihood phylogeny (GTR+F+G4 model, 1,000 ultra-fast bootstraps, using IQ-Tree) of 2,021 CC1 genomes based on the recombination-purged whole genome SNP alignment of 2.28 Mb. **(B)** Midpoint rooted maximum-likelihood phylogenetic tree of 2,002 SL1 genomes based on based on the recombination-purged whole genome SNP alignment of 2.28 Mb. The four external rings indicate the world region, year, type of infection and source type, respectively. The two inner rings indicate ST1 isolates and the 8 SL1 genetic clades identified in this study, respectively. The black star highlights the phylogenetic placement of isolate F2365 (accession no. NC_002973.6), used as reference in whole genome read mapping. Circles represent bootstrap branch support values higher than 90%, based on 1,000 replicates.

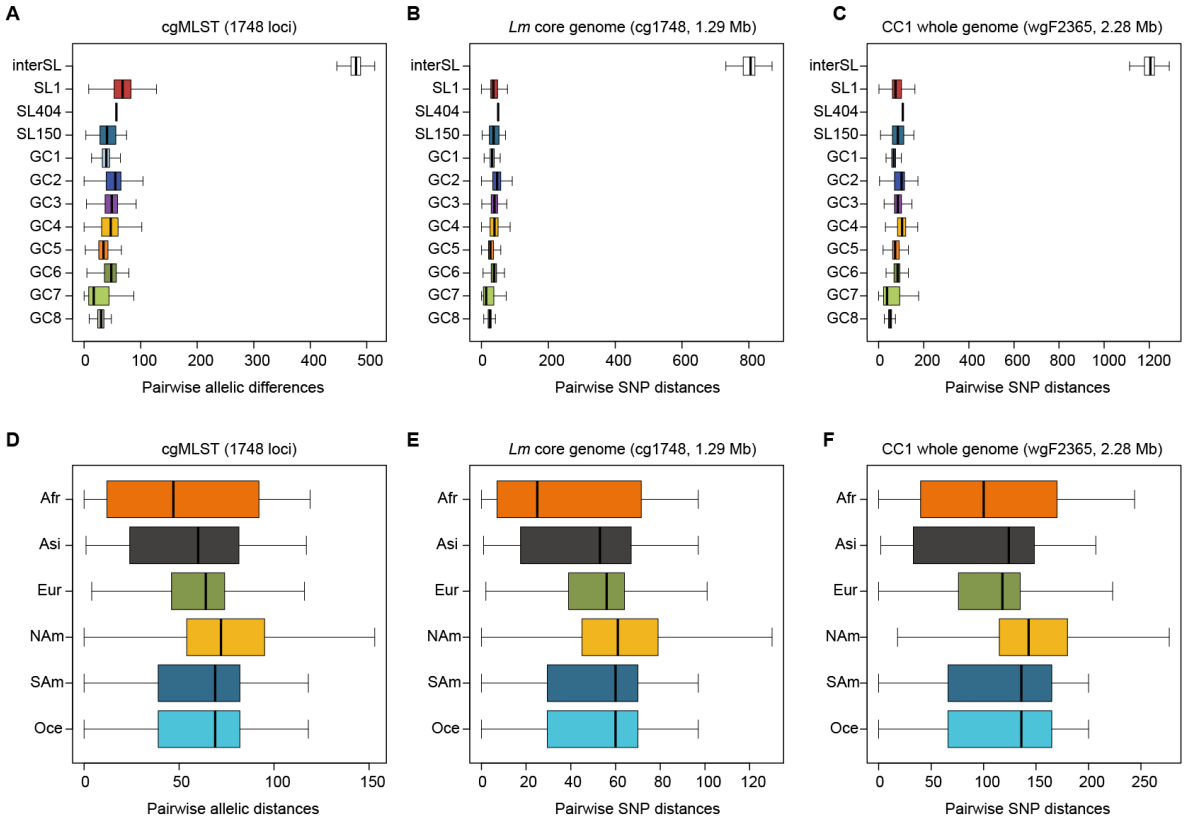


**Figure S7. Genetic diversity among *Lm*-CC1 isolates**

Pairwise isolate distances within CC1 phylogroups (top) and world regions (bottom): **(A, D)** pairwise cgMLST allelic distances; **(B, E)** pairwise SNP distances in recombination-purged *Lm* core genome alignment and **(C, F)** recombination-purged CC1 whole genome alignment. Uncalled alleles, Ns and gap alignment positions were ignored in pairwise comparisons. Each box denotes the 25% and 75% quartiles and lines represent the medians. Inter-SL, inter CC1 sublineages; GC#, within SL1 genetic clades; Afr, Africa; Asi, Asia; Eur, Europe; NAm, North America; Sam, South America; Oce, Oceania.

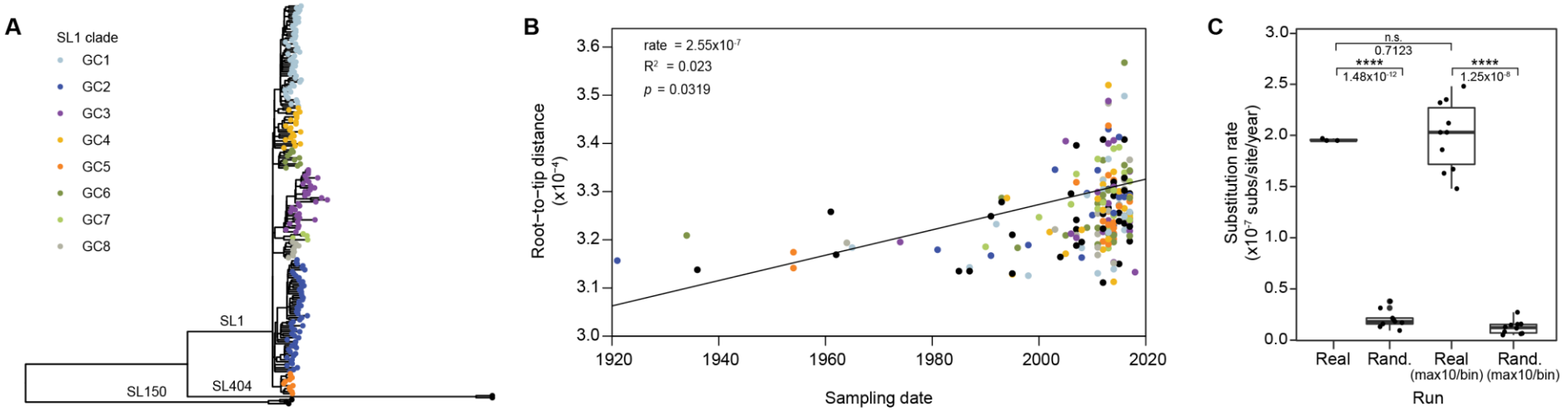


**Figure S8. Temporal analyses on a representative dataset of 200 isolates of *Lm*-CC1**

To control for the over representation of recent isolates, sensitivity analyses were also performed on normalized subsets, allowing only up to a maximum of 10 isolates per bin of 10 years. a) Maximum likelihood (GTR+F+G4) phylogeny of the representative 200 isolates selected randomly across the CC1 phylogeny. Tips are colored by sublineage and SL1 genetic clades as indicated in the legend. b) Regression analyses showing the root-to-tip genetic distance against sampling date (year). Statistical significance was assessed using the F-test. c) Bayesian molecular clock estimations based on real dates (run on 3 and 10 replicates, for the 200 and the normalized datasets, respectively) and randomized tip dates (10 replicates per dataset). Stars denote statistical significance of $p<0.0001$, assessed using t-test; n.s., not significant.

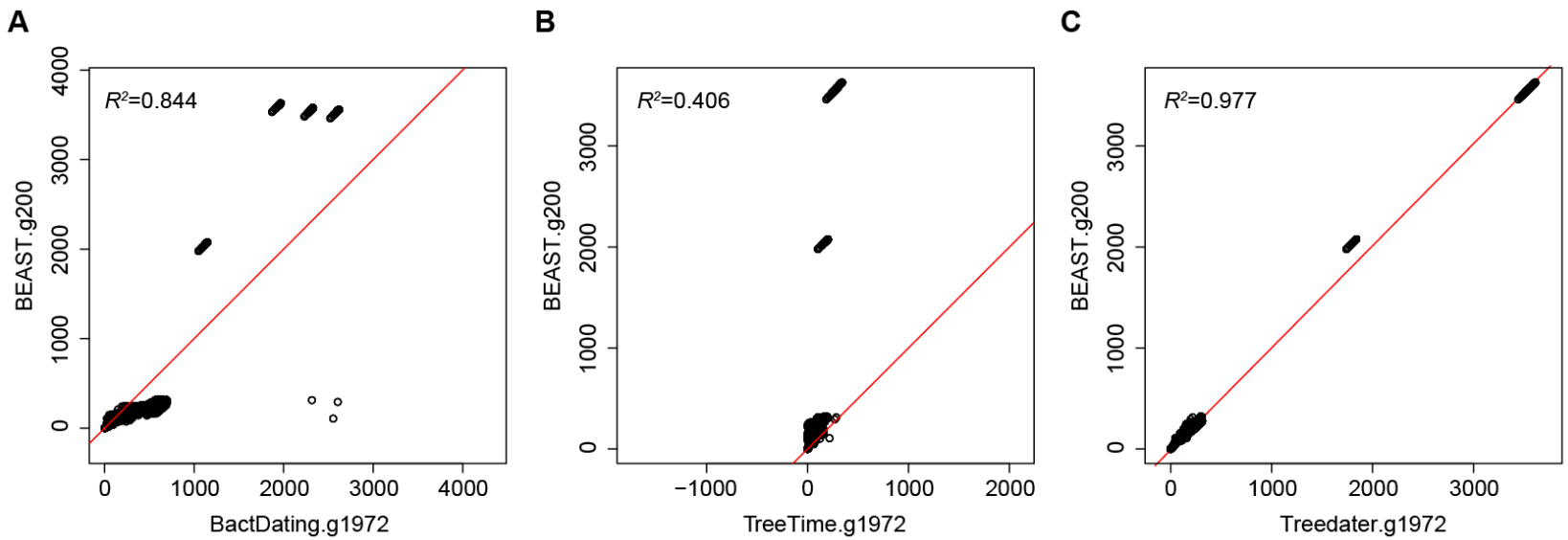


**Figure S9. Benchmarking of dating methods**

Cophenetic distances between isolates dated with BEAST and alternative large-scale methods: a) BactDating v.1.0.1[34], b) TreeTime v.0.5.2[35] and c) Treedater v.0.3.0[36], using the CC1 estimated rate of $1.954 \times 10^{-7} \pm 2.0152 \times 10^{-8}$ substitutions/site/year obtained with BEAST. g200 and g1972 refer to the number of CC1 genomes used in each analyse ($n$=200 and $n$=1,972, respectively). Red lines denote perfect positive correlation coefficients ($R^2$=1).

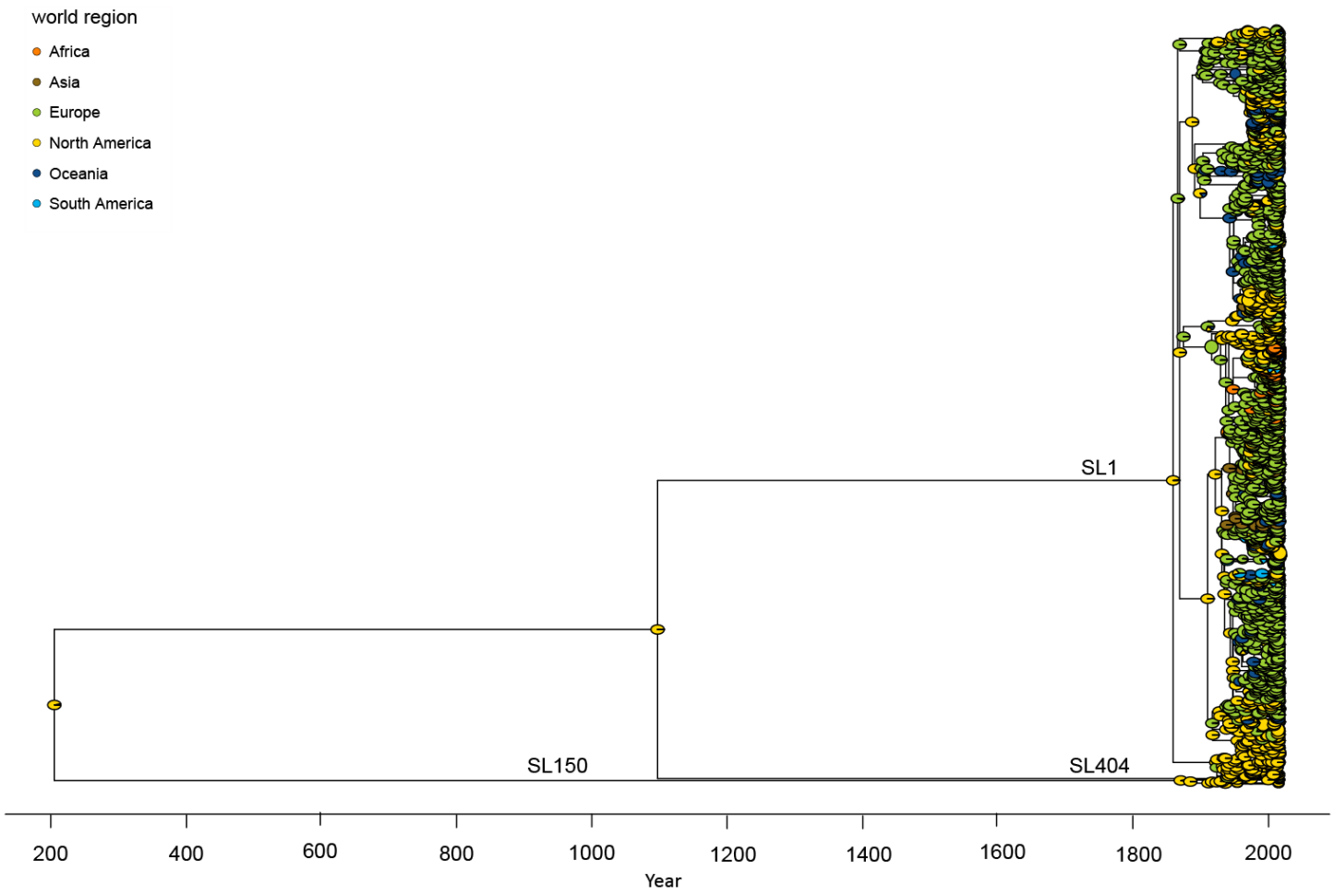


**Figure S10. Phylogeography inference of *Lm*-CC1 based on 1,972 dated genomes**

Pies at the nodes represent the probability of ancestral geographical locations, estimate using PastML using the MPPA method with an F81-like model. The detailed view of SL1 can be found in Figure 4.

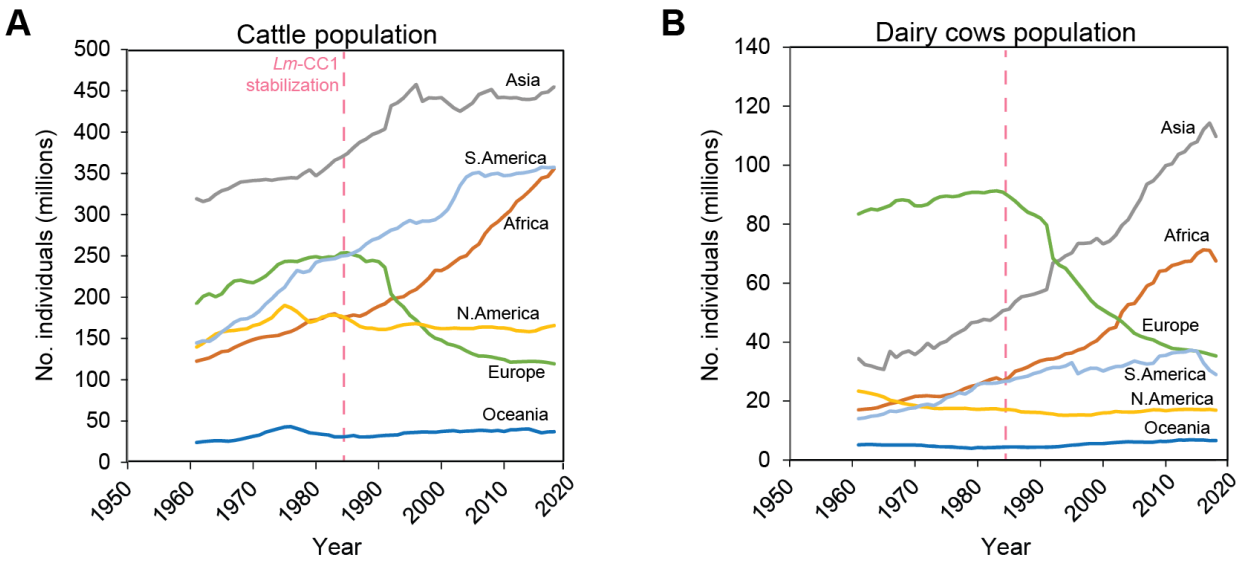


**Figure S11. Cattle demographics**

**(A)** Cattle population per world region; **(B)** Dairy cows per world region. Data available for 1961-2018; source: Food and Agriculture Organization of the United Nations; www.fao.org/faostat). Vertical dashed bars mark the estimated date of the stabilization *of Lm*-CC1 population size.

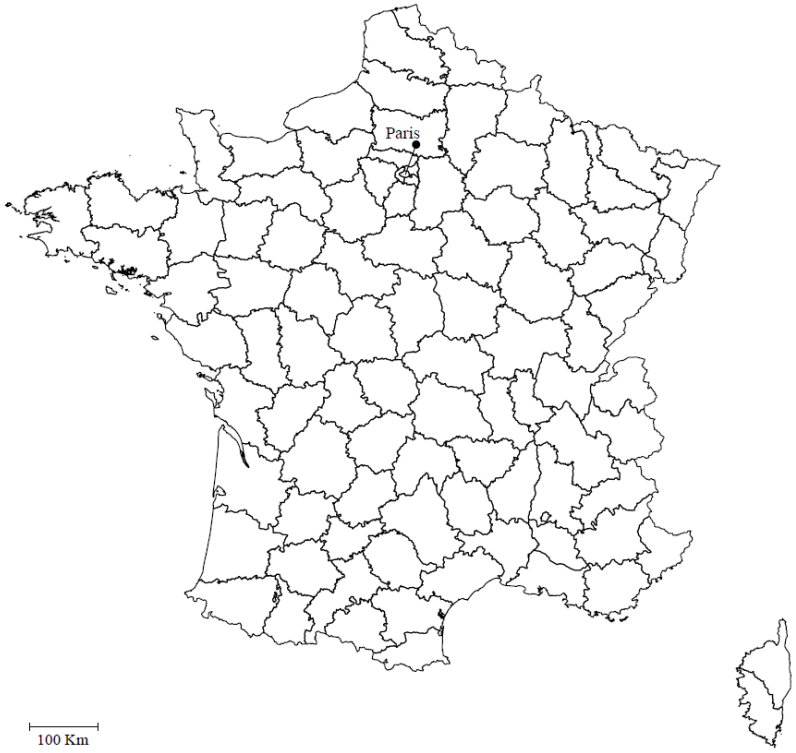


**Figure S12. French administrative departments** (***départements***)
Source: Global Administrative Areas, gadm.org.

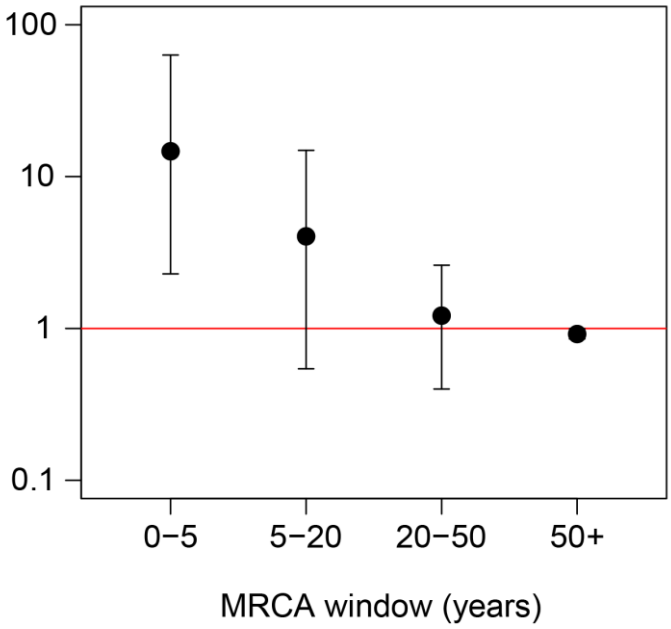


**Figure S13. SL1 transmission dynamics within country (France)**
Relative risk for a pair of isolates to have a MRCA within a defined period when coming from the same Department in France *versus* different ones.