

- 4-week-old plants and RG-II was solubilized from their cell walls by treatment with EPG (16).
18. W.-D. Reiter, C. C. S. Chapple, C. R. Somerville, *Science* **261**, 1032 (1993).
19. A. *thaliana* *mur1-1*, *mur1-2*, and wild-type plants were grown as described (17). Plants were sprayed every 7 days with a solution (15 ml per pot) of commercial fertilizer (Jacks 20:20:20, 1 g/liter) containing added L-Fuc (1.64 g/liter). Control plants were sprayed with fertilizer containing no added L-Fuc. Cell walls were prepared from leaves and RG-II were solubilized as described (16).
20. Details of the methods used to determine the in vitro rates of formation and stabilities of wild-type and *mur1* RG-II are available as supplementary material (13).
21. W. D. Reiter, *Plant Physiol. Biochem.* **36**, 167 (1998).
22. E. Zablackis et al., *Science* **272**, 1808 (1996).
23. W. S. York, M. A. O'Neill, unpublished data.
24. R. M. Perrin et al., *Science* **284**, 1976 (1999).
25. G. F. Vanzin, M. Madson, N. Carpita, W.-D. Reiter, personal communication.
26. C. Rayon et al., *Plant Physiol.* **119**, 725 (1999).
27. A. von Schaewen, A. Sturm, J. O'Neill, M. J. Chrispeels, *Plant Physiol.* **102**, 1109 (1993).
28. S. B. Selleck, *Trends Genet.* **16**, 206 (2000).
29. We thank W.-D. Reiter (University of Connecticut) and the Arabidopsis Biological Resource Center at Ohio State University for providing Arabidopsis seeds and J. Rose of Cornell University for critical reading of the manuscript. Supported by grants from the U.S. Department of Energy (DE-FG05-93ER20115 and DE-FG09-93ER20097).

7 May 2001; accepted 27 August 2001

Comparative Genomics of *Listeria* Species

P. Glaser,¹ L. Frangeul,¹ C. Buchrieser,¹ C. Rusniok,¹ A. Amend,⁴
F. Baquero,⁵ P. Berche,⁶ H. Bloecker,⁷ P. Brandt,⁷
T. Chakraborty,⁴ A. Charbit,⁶ F. Chetouani,¹ E. Couvé,¹
A. de Daruvar,⁸ P. Dehoux,² E. Domann,⁴ G. Domínguez-Bernal,⁹
E. Duchaud,¹ L. Durant,⁶ O. Dussurget,² K.-D. Entian,¹⁰ H. Fsihi,²
F. García-Del Portillo,¹¹ P. Garrido,⁹ L. Gautier,⁸ W. Goebel,¹²
N. Gómez-López,¹¹ T. Hain,⁴ J. Hauf,¹⁰ D. Jackson,⁸ L.-M. Jones,³
U. Kaerst,⁷ J. Kreft,¹² M. Kuhn,¹² F. Kunst,¹ G. Kurapkat,⁸
E. Madueño,¹¹ A. Maitournam,¹ J. Mata Vicente,⁵ E. Ng,¹²
H. Nedjari,¹ G. Nordsiek,⁷ S. Novella,⁹ B. de Pablos,¹¹
J.-C. Pérez-Díaz,⁵ R. Purcell,¹ B. Rammel,⁸ M. Rose,¹⁰
T. Schlüter,⁸ N. Simoes,¹ A. Tierrez,⁹ J.-A. Vázquez-Boland,⁹
H. Voss,⁸ J. Wehland,⁷ P. Cossart^{2*}

Listeria monocytogenes is a food-borne pathogen with a high mortality rate that has also emerged as a paradigm for intracellular parasitism. We present and compare the genome sequences of *L. monocytogenes* (2,944,528 base pairs) and a nonpathogenic species, *L. innocua* (3,011,209 base pairs). We found a large number of predicted genes encoding surface and secreted proteins, transporters, and transcriptional regulators, consistent with the ability of both species to adapt to diverse environments. The presence of 270 *L. monocytogenes* and 149 *L. innocua* strain-specific genes (clustered in 100 and 63 islets, respectively) suggests that virulence in *Listeria* results from multiple gene acquisition and deletion events.

Listeria monocytogenes is the etiologic agent of listeriosis, a severe food-borne disease. It survives in the extreme conditions encountered in the food chain, such as high salt concentrations and extremes of pH and temperature. These characteristics are shared by *L. innocua*, a nonpathogenic species often associated with *L. monocytogenes* in food and the environment. The clinical features of listeriosis include meningitis, meningoencephalitis, septicemia, abortion, perinatal infections, and gastroenteritis (1). After ingestion of contaminated food, *Listeria* disseminates from the intestinal lumen to the central nervous system and the fetoplacental unit. The key role of the surface protein internalin (InlA) in the crossing of the intestinal barrier was recently reported (2). Other virulence factors include the invasion protein InlB; the proteins LLO and PlcA, which promote escape from the phagocytic vacuole; and the pro-

teins ActA and PlcB, which are necessary for intracellular actin-based motility and cell-to-cell spread (1, 3). These genes are clustered on a 10-kb virulence locus that is absent from *L. innocua* (3, 4).

Two strains were selected for comparison: *L. monocytogenes* EGD-e (serovar 1/2a), a derivative of strain EGD used by Mackaness in his studies on cell-mediated immunity (5), and *L. innocua* strain CLIP 11262 (serovar 6a), used for heterologous expression of *L. monocytogenes* genes (6). The whole-genome random sequencing method was chosen (7, 8). *Listeria monocytogenes* contains one circular chromosome of 2,944,528 base pairs (bp) with an average G+C content of 39% (Table 1 and Fig. 1) (GenBank/EMBL accession number AL591824). The *L. innocua* chromosome has a similar size (3,011,209 bp) and a similar G+C content (37%) (Table 1 and Fig. 1) (GenBank/EMBL accession number

AL592022). The *L. innocua* strain also contains a plasmid of 81,905 bp (GenBank/EMBL accession number AL592102). We identified 2853 protein-coding genes in the *L. monocytogenes* chromosome and 2973 in that of *L. innocua* (8). Encoded proteins revealed a striking similarity to those of the soil bacterium *Bacillus subtilis*. Genes were thus classified according to the functional categories defined for *B. subtilis* (9) [Web table 1 (8)]. No function could be predicted for 35.3% of *L. monocytogenes* genes and 37% of *L. innocua* genes, a proportion similar to that found in other sequenced bacterial genomes.

Both genomes encoded many putative surface proteins belonging to six families [Web fig. 1 (8)], and expansion of these families seems to be partly due to gene duplications. Internalin and InlB belong to a family of proteins characterized by an NH₂-terminal domain containing leucine-rich repeats (LRRs). Seven other members of this family have already been identified (1, 3). Except for InlB, which is loosely attached to the bacterial surface, and InlC, which is secreted, the five other LRR proteins have a Leu-Pro-X-Thr-Gly (LPXTG) motif that mediates their covalent linkage to peptidoglycan (10). The *L. monocytogenes* genome sequence revealed the presence of 41 proteins containing an LPXTG

¹Génomique des Microorganismes Pathogènes, ²Unité des Interactions Bactéries-Cellules, ³Service d'Informatique Scientifique, Institut Pasteur, 25-28 rue du Dr. Roux, 75724 Paris, France. ⁴Institute for Medical Microbiology, Frankfurterstrasse 107, D-35392 Giessen, Germany. ⁵Servicio de Microbiología, Hospital Ramon y Cajal, 28034 Madrid, Spain. ⁶Unité INSERM U411, Faculté de Médecine Necker-Enfants Malades, 156 rue de Vaugirard, 75730 Paris, France. ⁷Department of Cell Biology and Immunology, Gesellschaft für Biotechnologische Forschung (GBF)-Braunschweig, D-38124 Braunschweig, Germany. ⁸LION Bioscience AG, Im Neuenheimer Feld 517, D-69120 Heidelberg, Germany. ⁹Grupo de Patogénesis Molecular Bacteriana, Departamento Patología Animal I, Facultad de Veterinaria, Universidad Complutense, 28040 Madrid, Spain. ¹⁰Scientific Research and Development GmbH, D-61440 Oberursel, Germany. ¹¹Centro de Biología Molecular "Severo Ochoa," Universidad Autónoma de Madrid, 28049 Madrid, Spain. ¹²TBI-Biozentrum, Lehrstuhl für Mikrobiologie, Universität Würzburg, D-97074 Würzburg, Germany.

*To whom correspondence should be addressed. E-mail: pcossart@pasteur.fr

motif, 19 of which belong to the LRR/internalin family. Eleven of those are absent from *L. innocua* [Web table 2 (8)]. *Listeria monocytogenes* contained more LPXTG proteins than any other Gram-positive bacterium whose genome has been sequenced [13 in *Streptococcus pyogenes* (11), 18 in *Staphylococcus aureus* (12)]. InlB, which is absent from *L. innocua*, and the adhesion protein Ami are attached to lipoteichoic acid via GW modules (13, 14). The *L. monocytogenes* genome contained seven additional members of this family, one of which was absent from *L. innocua* [Web table 3 (8)]. Other surface proteins included proteins that, like ActA (15, 16), have a signal sequence and a hydrophobic COOH-terminal region that may anchor them to the cell membrane [Web table 4 (8)] and p60-like proteins [Web table 5 (8)]. Both *Listeria* spp. encoded 68 putative lipoproteins, predicted on the basis of their characteristic signal sequences [Web table 6 (8)]. Several secreted proteins important for the virulence of *L. monocytogenes* were identified previously, including LLO, PlcA, PlcB, and InlC (1, 3). The genome of *L. monocytogenes* was predicted to encode 86 secreted proteins, some of which have putative degradative functions, like lipases or chitinases. Of these 86 proteins, 23—including three soluble internalins—were absent from *L. innocua* [Web table 7 (8)].

The ability of *Listeria* sp. to colonize and grow in a broad range of ecosystems correlates with the presence of 331 genes encoding different transport proteins (11.6% of all predicted genes of *L. monocytogenes*). Interestingly, 88 (26%) of the 331 transporter genes were devoted to carbohydrate transport, mediated by phosphoenolpyruvate-dependent phosphotransferase systems (PTS) and corresponding to 39 putative complete or incomplete enzyme II permeases [Web table 8 (8)]. *Listeria monocytogenes* has nearly twice as many PTS permeases as *Escherichia coli* and nearly three times as many as *B. subtilis*. Carbohydrates, in particular β -glucosides, have a remarkable impact on the virulence of *L. monocytogenes* [for a review, see (17)]. Eight enzyme II permeases, five of which were predicted to be specific for β -glucosides, were absent from *L. innocua* [Web table 8 (8)]. Thus, in agreement with recent results (18), the *L. monocytogenes* EGD-e-specific PTS could be implicated in virulence.

Given the many environmental conditions that *L. monocytogenes* faces, an extensive regulatory repertoire was expected. We identified 209 and 203 transcriptional regulators, respectively, in *L. monocytogenes* and *L. innocua* [Web table 9 (8)]. This high proportion of regulatory genes

(7.3%) is second only to that of *Pseudomonas aeruginosa* (8.4%) (19), another ubiquitous, opportunistic pathogen. However, *L. monocytogenes* encodes only five sigma factors, versus 18 in *B. subtilis* (9) and 13 in *Mycobacterium tuberculosis* (20). The best characterized regulatory factor of *L. monocytogenes*—PrfA, a member of the Crp/Fnr family—was absent from *L. innocua*. It activates most of the known virulence genes. PrfA binds to a palindromic PrfA recognition sequence (PrfA-box) located in the promoter region (21). Sequence analysis identified genes preceded by a putative PrfA-box in both genomes [Web tables 10 and 11 (8)]. The Crp/Fnr family comprises 15 members in *L. monocytogenes* and 14 in *L. innocua*. The importance of this regulatory family in *Listeria* sp. is highlighted by comparison with other ge-

nomes: *B. subtilis* (9) contains one regulator of this type, *E. coli* two (22), and *P. aeruginosa* four (19). The two largest families of regulatory proteins are the GntR-like regulators and the BglG-like antiterminators, many of which are associated with PTS [Web table 9 (8)]. Both *Listeria* genomes encoded 15 histidine kinases and 16 response regulators constituting two-component regulatory systems. If genome size is taken into account, this number of two-component regulators is similar to that of ubiquitous bacteria such as *B. subtilis* (9) and *E. coli* (22) but higher than that of pathogens with narrow host tropisms, such as *M. tuberculosis* (11 pairs) (20) and *Neisseria meningitidis* (five pairs) (23).

Like *B. subtilis*, the two *Listeria* species express four different classes of stress proteins (HrcA- or sigmaB-dependent, the Clp

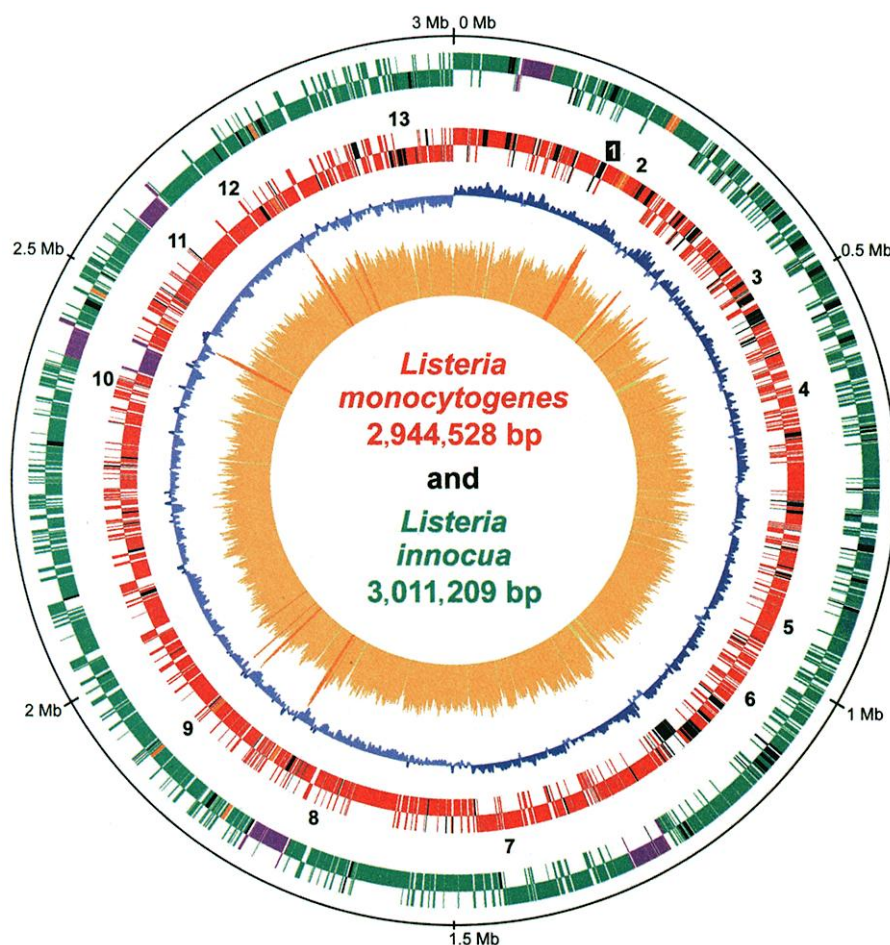


Fig. 1. Circular genome maps of *L. monocytogenes* EGD-e and *L. innocua* CLIP 11262, showing the position and orientation of genes. From the outside: Circles 1 and 2, *L. innocua* and *L. monocytogenes* genes on the + and - strands, respectively. Color code: green, *L. innocua* genes; red, *L. monocytogenes* genes; black, genes specific for *L. monocytogenes* or *L. innocua*, respectively; orange, rRNA operons; purple, prophages. Numbers on the second circle indicate the position of known virulence genes: 1, virulence locus (*prfA-plcA-hly-mpl-actA-plcB*); 2, *clpC*; 3, *inlAB*; 4, *iap*; 5, *dal*; 6, *clpE*; 7, *lisRK*; 8, *dat*; 9, *inlC*; 10, *arpJ*; 11, *clpP*; 12, *ami*; 13, *bvrABC*. Circle 3, G/C bias (G+C/G-C) of *L. monocytogenes*. Circle 4, G+C content of *L. monocytogenes* (<32.5% G+C in light yellow, 32.5 to 43.5% in yellow, and >43.5% G+C in dark yellow). The scale in megabases is indicated on the outside of the genome circles, with the origin of replication at position 0.

REPORTS

family, and the so-called class IV genes) (24) and encode three paralogous cold-shock proteins. Genes involved in acid resistance were also identified [e.g., genes encoding glutamic acid decarboxylases (*gad*)]. Interestingly, one of the three *gad* paralogs of *L. monocytogenes* (*lmo0447*) was missing from *L. innocua*. Furthermore, three genes possibly involved in the degradation of bile salts were present in *L. monocytogenes* but not in *L. innocua* (*lmo2067*, *lmo0446*, and *lmo0754*), probably reflecting the capacity of *L. monocytogenes* to survive in the mammalian gut. One of these genes (*lmo2067*) was preceded by a PrfA-

box, suggesting that it may encode a virulence factor.

The enzymes necessary for glycolysis and the pentose phosphate pathway were present in both *Listeria* genomes. In contrast, the tricarboxylic acid cycle was not complete because α -ketoglutarate dehydrogenase was missing. *Listeria* are able to produce adenosine triphosphate through a complete respiratory chain and contain numerous fermentation pathways. These results are in agreement with the microaerophilic and facultative anaerobic lifestyle of *Listeria*. Growth of *Listeria* in defined medium requires the addition of four vitamins

(riboflavin, biotin, thiamin, and lipoate) and six amino acids (Leu, Ile, Arg, Met, Val, and Cys) (25). Metabolic reconstruction indicated that the biosynthesis pathways for the four vitamins were missing or incomplete. In contrast, all amino acid biosynthesis pathways were identified. Hence, the requirement for amino acids may be due to repression of some amino acid biosynthetic pathways in laboratory growth conditions.

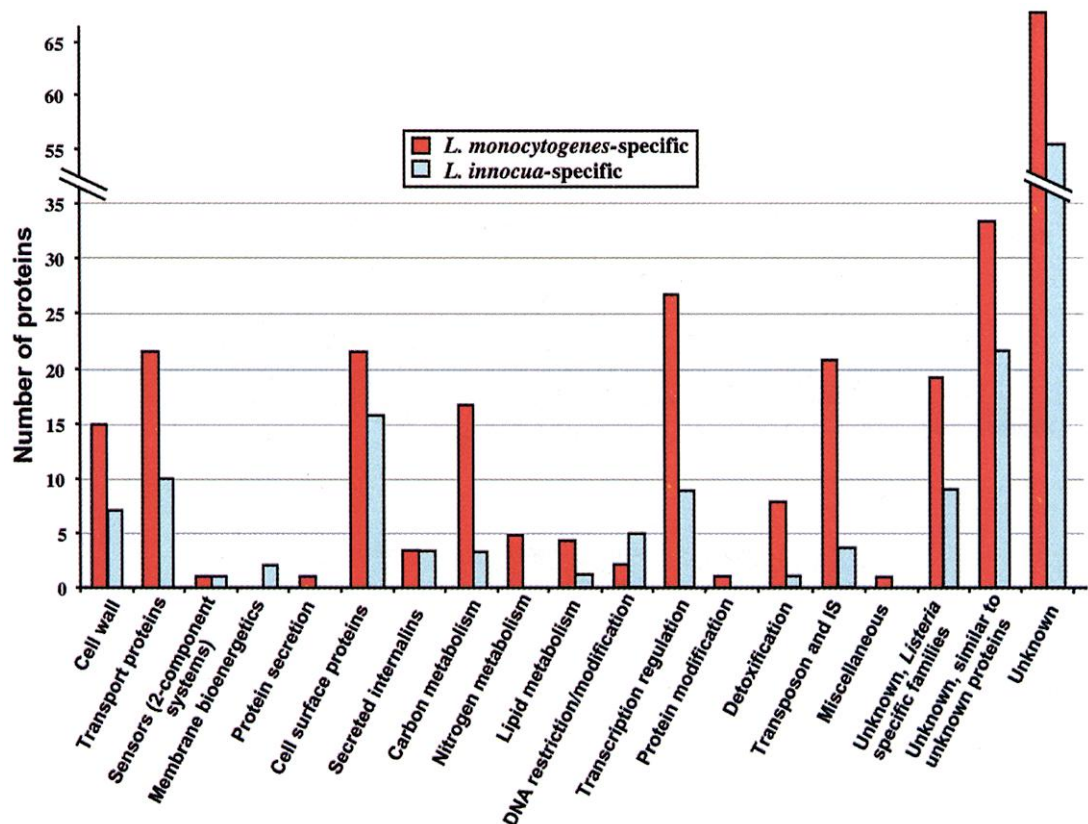
Comparative analysis also revealed a conserved, co-linear organization of the two *Listeria* genomes and an unexpected synteny with the genomes of *B. subtilis* and *S. aureus*, indicating that genomes of this group of bacteria are particularly stable. If prophage genes are excluded, 270 (9.5%) *L. monocytogenes* EGD-e-specific genes and 149 (5%) *L. innocua*-specific genes were identified [Web tables 12 and 13 (8)]. Their different distribution within functional categories—in particular, proteins homologous to known virulence-associated proteins from other bacteria or proteins implicated in adaptation to different environments—reflects the species-specific properties of *L. monocytogenes* (Fig. 2). Genes present in only one species were scattered in multiple regions between 1 and 25 kb on the chromosome (100 in *L. monocytogenes* and 63 in *L. innocua*) (Fig. 1), a situation similar to *E. coli* O 157:H7 compared with *E. coli* K12 (26) but different

Table 1. General features of the two *Listeria* genomes.

	<i>L. monocytogenes</i>	<i>L. innocua</i>
Size of the chromosome (bp)	2,944,528	3,011,209
G+C content (%)	39	37
G+C content of protein-coding genes (%)	38	38
Total number of protein-coding genes	2853	2973
Average length (codons) of protein-coding genes	306	299
Number of rRNA operons (16S-23S-5S)	6	6
Number of tRNA genes	67	66
Percentage coding	90.3%	90.3%
Prophages	1 (60 genes, 41.6 kbp)	5 (301 genes, 219.4 kbp)
Plasmid	0	1 (79 genes, 81.9 kbp)
Number of strain-specific genes*	270 (294 kbp)	149 (195 kbp)
Number of orthologous genes*	2523	2523
Number of transposons	1 (Tn916-like)	—

*Prophage genes excepted.

Fig. 2. Distribution of the 270 *L. monocytogenes* EGD-e-specific and 149 *L. innocua* CLIP 11262-specific genes within the different functional categories. Genes were considered as strain specific for *L. monocytogenes* when they had no ortholog in *L. innocua*, and vice versa. See also Web tables 12 and 13 (8).



from that of *Chlamydia trachomatis* compared with *Chlamydia pneumoniae* (27) where only a few variable regions have been identified. In *L. monocytogenes*, 54 of the 100 specific regions had a significantly lower G+C content than the flanking regions and 6 had a significantly higher G+C content, suggesting recent acquisition by horizontal gene transfer. However, more fragments may have been acquired by horizontal gene transfer from bacteria with a similar G+C content, or over time may have adapted to the *Listeria* genome. This seems to be the case for the "virulence locus," whose G+C content is similar to that of the rest of the chromosome. Comparison of this region among *Listeria* species (4) and with *B. subtilis* indicates

that this gene cluster was probably acquired by a common ancestor of *Listeria* and that *L. innocua* subsequently lost most of it (Fig. 3).

What are the evolutionary forces that led to this mosaic genome structure in *Listeria*? Bacteriophages and plasmids may have played a role in gene acquisition, because the sequenced *L. monocytogenes* and *L. innocua* strains contained one and five prophages, respectively. Furthermore, *L. innocua* contained a plasmid encoding heavy metal resistance. However, the most unexpected finding was that both *Listeria* genomes contain putative DNA uptake genes, homologous to *B. subtilis* competence genes. As *Listeria* are not known to be naturally competent, the *Listeria* DNA

uptake apparatus may have lost its original function. Alternatively, its regulation or the signals that induce competence may differ from those of *B. subtilis*. Gene transfer by transformation could thus explain most of the genomic differences between the two *Listeria* species as well as between *Listeria* and *B. subtilis*.

Sequence analysis of the two *Listeria* genomes revealed a close relationship to *B. subtilis*, suggesting a common origin for the three species. *Listeria* subsequently acquired multiple noncontiguous DNA fragments, including the previously known virulence locus. Characterization of these regions, including analysis of the many surface proteins and adaptation systems, opens new avenues for postgenomic analysis of the life-styles of *L. monocytogenes* in the environment and the infected host.

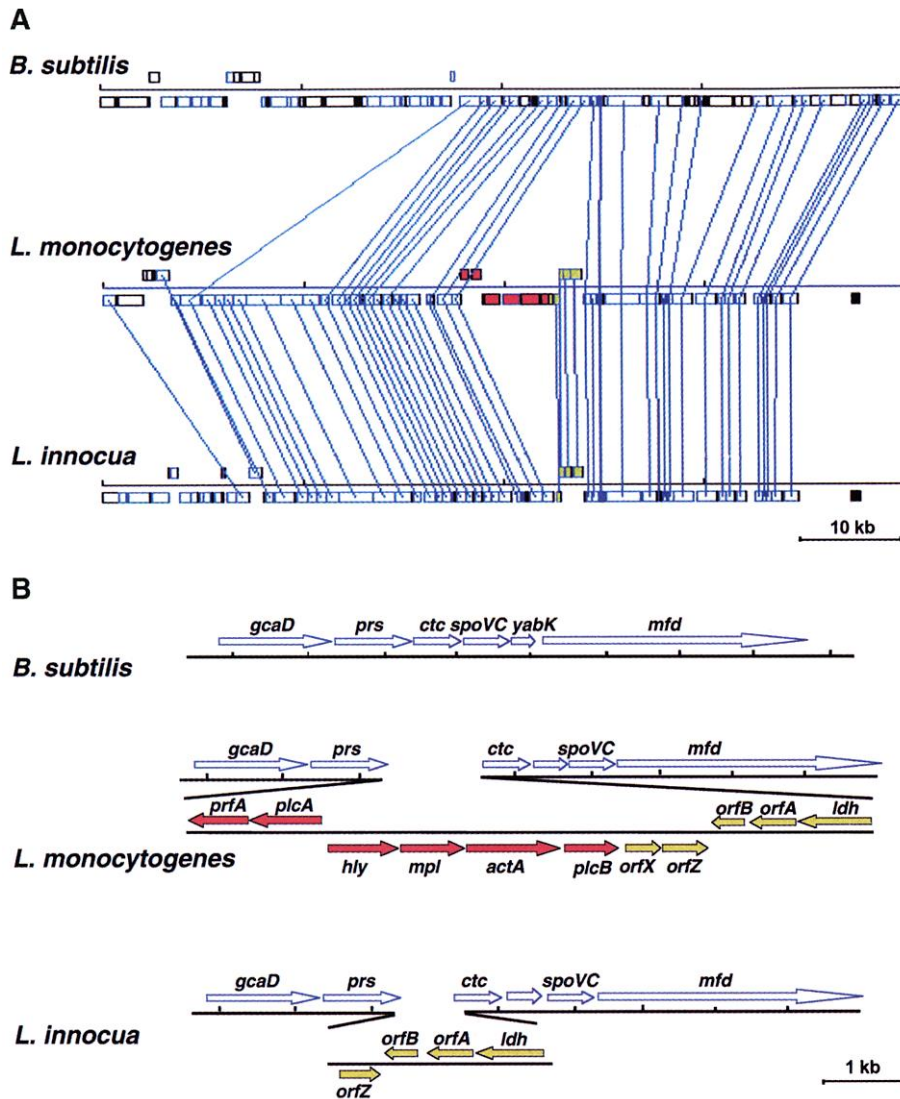


Fig. 3. Comparison of the region containing the "virulence gene cluster" of *L. monocytogenes* and the homologous regions of the *L. innocua* and *B. subtilis* genomes. Open blue boxes and arrows, orthologs among the three genomes; solid red boxes and arrows, virulence gene cluster; solid yellow boxes and arrows, genes absent from *B. subtilis*. (A) Scheme generated by GenomeScout software (LION Bioscience). (B) Enlargement of the region containing the virulence gene cluster.

References and Notes

1. J.-A. Vazquez-Boland et al., *Clin. Microbiol. Rev.* **14**, 548 (2001).
2. M. Lecuit et al., *Science* **292**, 1722 (2001).
3. P. Cossart, M. Lecuit, *EMBO J.* **17**, 3797 (1998).
4. T. Chakraborty, T. Hain, E. Domann, *Int. J. Med. Microbiol.* **2**, 167 (2000).
5. G. B. Mackaness, *J. Exp. Med.* **120**, 105 (1964).
6. J.-L. Gaillard et al., *Cell* **65**, 1127 (1991).
7. R. D. Fleischmann et al., *Science* **269**, 496 (1995).
8. Supplementary Web material, including sequencing, assembly, and annotation methods, is available on Science Online (www.sciencemag.org/cgi/content/full/294/5543/849/DC1).
9. F. Kunst et al., *Nature* **390**, 249 (1997).
10. W. W. Navarre, O. Schneewind, *Microbiol. Mol. Biol. Rev.* **63**, 174 (1999).
11. J. Ferretti et al., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4658 (2001).
12. M. Kuroda et al., *Lancet* **357**, 1225 (2001).
13. L. Braun et al., *Mol. Microbiol.* **25**, 285 (1997).
14. R. Jonquères et al., *Mol. Microbiol.* **34**, 902 (1999).
15. C. Kocks et al., *Cell* **68**, 521 (1992).
16. E. Domann et al., *EMBO J.* **11**, 1981 (1992).
17. J. Kreft, J. A. Vazquez-Boland, *Int. J. Med. Microbiol.* **2**, 291 (2001).
18. K. Brehm et al., *J. Bacteriol.* **181**, 5024 (1999).
19. C. K. Stover et al., *Nature* **406**, 959 (2000).
20. S. T. Cole et al., *Nature* **393**, 537 (1998).
21. W. Goebel et al., in *Gram-Positive Pathogens*, V. A. Fischetti et al., Eds. (ASM Press, Washington, DC, 2000), vol. 1, pp. 439–450.
22. F. R. Blattner et al., *Science* **277**, 1453 (1997).
23. H. Tettelin et al., *Science* **287**, 1809 (2000).
24. S. Nair et al., *Mol. Microbiol.* **35**, 800 (2000).
25. R. J. Premaratne, W. J. Lin, E. A. Johnson, *Appl. Environ. Microbiol.* **57**, 3046.
26. N. T. Perna et al., *Nature* **409**, 529 (2001).
27. T. D. Read et al., *Nucleic Acids Res.* **28**, 1397 (2000).
28. Sequence and annotation data are at <http://genolist.pasteur.fr/listist>. For comparison of *Listeria* sp. and *B. subtilis* genomes, see <http://GS-Listeria.lionbioscience.com>. We thank D. Cabanes, G. Casari, L. Duret, F. Engelbrecht, M. Herler, I. Moszer, G. Perrière, and M. Rauch for help in annotation; J. Arroyo, J. L. Garcia, and C. Jacquet for expertise; and M. Schwartz and A. Danchin for support. P.C. is an international scholar from the Howard Hughes Medical Institute. Supported by the European Commission (contract BIO4CT980036) and the Institut Pasteur.

14 June 2001; accepted 10 September 2001