

Comparison of the genome sequences of *Listeria monocytogenes* and *Listeria innocua*: clues for evolution and pathogenicity

Carmen Buchrieser^{a,*}, Christophe Rusniok^a, The *Listeria* Consortium¹, Frank Kunst^a,
Pascale Cossart^b, Philippe Glaser^a

^a Laboratoire de Génomique des Microorganismes Pathogènes, Institut Pasteur, 25, rue du Dr. Roux, 75724 Paris Cedex 15, France

^b Unité des Interactions Bactéries Cellules, Institut Pasteur, 28, rue du Dr Roux, 75724 Paris, France

Received 2 October 2002; accepted 28 October 2002

First published online 22 November 2002

Abstract

Listeria monocytogenes, an invasive opportunistic, food-borne pathogen, remains one of the leading causes of mortality from food-borne infections. The recently determined complete genome sequences of *L. monocytogenes* strain EGDe and of that of the closely related non-pathogenic species *Listeria innocua* strain CLIP11262 enhance our knowledge of the genetic basis of the virulence of *L. monocytogenes* and advance our understanding of the evolution of these *Listeria* species. Both genomes encode a high number of surface, transport and regulatory proteins. Comparison of the genome organisation revealed a perfect synteny between the two *Listeria* genomes. Comparison with other closely related bacteria also showed a high conservation in genome organisation among the *Listeria*, *Staphylococcus* and *Bacillus* group of low G+C content bacteria. Distinct G+C content of a number of strain-specific genes suggests intensive lateral gene transfer. The identification of a 55-kb locus encoding proteins with high homology to *Salmonella enterica* serovar Typhimurium vitamin B₁₂ synthesis proteins as well as those necessary for degradation of ethanolamine and propanediol further indicates acquisition of a complete metabolic pathway by horizontal gene transfer and a probable role of this locus in anaerobic growth in the host.

© 2003 Federation of European Microbiological Societies. Published by Elsevier Science B.V. All rights reserved.

Keywords: *Listeria monocytogenes*; Comparative genomics; Vitamin B₁₂

1. Introduction

The intracellular pathogen *Listeria monocytogenes* is the causative agent of serious epidemic and sporadic listeriosis. The involvement of food as a vector for transmission of *L. monocytogenes* is clearly established in relation to both epidemic and sporadic disease [1,2]. Although rare

when compared to many other food-borne diseases, listeriosis often leads to severe consequences, particularly in susceptible individuals like pregnant women, newborns, people over 65 years and immunocompromised patients. Following ingestion of contaminated food, *Listeria* disseminates from the intestinal lumen to the central nervous system and the foeto-placental unit. Clinical features of listeriosis include meningitis, meningo-encephalitis, septicaemia, abortion, perinatal infections and also gastroenteritis [3].

L. monocytogenes is widely present in nature and it has also been isolated from numerous animals such as cattle, sheep, goats and poultry, but infrequently from wild animals [4]. Furthermore, *L. monocytogenes* has the important capacity to adapt and survive in extreme environments such as high salt concentration (10% NaCl), a broad pH range (from 4.5 to 9.0) and a wide temperature range. The ability to grow between –1°C and 45°C increases the contamination risk in dairy products, meats, seafood and other processed food products via selective enrichment during refrigeration. *Listeria* can also survive

* Corresponding author. Tel.: +33 (1) 45-68-83-72;
Fax: +33 (1) 45-68-87-86.
E-mail address: cbuch@pasteur.fr (C. Buchrieser).

¹ Glaser, P., Frangeul, L., Buchrieser, C., Amend, A., Baquero, F., Berche, P., Bloecker, H., Brandt, P., Chakraborty, T., Charbit, A., Chétouani, F., Couvé, E., de Daruvar, A., Dehoux, P., Domann, E., Domínguez-Bernal, G., Duchaud, E., Durand, L., Dusurget, O., Entian, K.-D., Fsihi, H., Garcia-Del Portillo, F., Garrido, P., Gautier, L., Goebel, W., Gómez-López, N., Hain, T., Hauf, J., Jackson, D., Jones, L.-M., Kärst, U., Kreft, J., Kuhn, M., Kunst, F., Kurapkat, G., Madueño, E., Maitournam, A., Mata Vicente, J., Ng, E., Nordstiek, G., de Pablos, B., Pérez-Díaz, J.-C., Rimmel, B., Rose, M., Rusniok, C., Schlueter, T., Simoes, N., Vázquez-Boland, J.-A., Voss, H., Wehland, J. and Cossart, P.

long periods of drying and freezing with subsequent thawing [5].

During the last 10 years, *L. monocytogenes* has emerged as a model system to study basic aspects of intracellular pathogenesis. The capacity of *L. monocytogenes* to cross the intestinal barrier [6,7] to enter cells, escape from the vacuole, grow in the mammalian cell cytosol, exploit a host system of actin-based motility, and spread from cell to cell, all without contact with the humoral immune system (for review see [3,8]), was elucidated in recent years. However, relatively little is known on the molecular level about for example carbohydrate utilisation, biosynthetic pathways, anaerobic growth and nutritional requirements of *L. monocytogenes*. Numerous biochemical and physiological studies, initiated before gene isolation techniques and genome sequencing became available, have not been pursued and extended to the gene level.

A major step forward was the determination and publication of the complete genome sequence of *L. monocytogenes* and the closely related non-pathogenic species *Listeria innocua* [9]. Sequence analysis and in particular comparative genomics will now help to unravel the molecular basis of the pathogenesis, phenotypic differences and the evolution of listeriae and should allow a better understanding of its biology at the molecular level. A summary of the specific features of the *Listeria* genomes as deduced from their genome sequences and a comparative analysis on genome organisation and gene acquisition events will be presented. These results shed light on pathogenicity and evolution within the genus *Listeria*.

2. A short glimpse of *Listeria*-specific features as deduced from the genome sequence

The *L. monocytogenes* EGDe genome is 2.9 Mb long and has an average G+C content of 39%. The most striking features are an exceptionally large number of surface proteins (4.7% of all predicted genes of *L. monocytogenes* EGDe), an abundance of transport proteins, in particular proteins dedicated to carbohydrate transport (11.6% of all predicted genes of *L. monocytogenes* EGDe), and an extensive regulatory repertoire (7.3% of all predicted genes of *L. monocytogenes* EGDe).

Surface proteins have important roles in the interactions of the micro-organism with its environment, in particular during host infection. Major virulence factors of *L. monocytogenes* are surface proteins, such as internalin and InlB necessary to enter eukaryotic cells, or ActA, playing a key role in actin-based motility. An extensive analysis of the 133 surface proteins of the *L. monocytogenes* EGDe genome was published recently [10]. The largest family are lipoproteins (68 proteins or 2.5% of all genes of the genome) and the second largest family are LPXTG proteins (41 proteins or 1.4% of all genes of the genome). The LPXTG motif is a C-terminal sorting sequence allowing

covalent linkage to the cell wall. The third family are proteins non-covalently attached to the cell surface mediated by their carboxy-terminal domains (24 proteins). This family can be divided into three subfamilies: (i) GW proteins (9) like InlB or Ami which contain in their carboxy-terminal region three modules of ~80 amino acids containing the dipeptide Gly-Trp ('GW modules'), constituting a cell wall adhesion motif; (ii) hydrophobic tail proteins (11) which, like ActA, contain in their carboxy-terminal region a hydrophobic stretch of 22 amino acids followed by a positively charged tail, and (iii) P60-like proteins (4) which are modular proteins containing different domains. Thirty of the 133 surface proteins identified in *L. monocytogenes* EGDe are absent from the *L. innocua* CLIP11262 genome and most interestingly 20 of these 30 *L. monocytogenes* EGDe-specific surface proteins are LPXTG proteins.

The 209 regulatory proteins identified in the *L. monocytogenes* genome should be connected to the capacity of *Listeria* to adapt and respond to a wide variety of different environments and some global regulatory systems might also be associated with virulence. The two largest families of regulatory proteins are GntR regulators (24 proteins) and BglG antiterminators (18 proteins). The family of BglG regulatory proteins seems to be over-represented in *Listeria*, possibly due to the fact that many are associated with PTS systems involved in sugar transport and metabolism which are abundant in *Listeria*. A third large family of regulatory proteins corresponds to two-component systems. *L. monocytogenes* possesses 15 histidine kinases and 17 response regulators.

A third specific feature of the *Listeria* genomes probably also related to its property to colonise a broad range of ecosystems is the presence of a large number of genes encoding different transport proteins. Like in most bacterial genomes the predominant class corresponds to ABC transporters. However, most interestingly 86 (26%) out of these 331 genes are devoted to carbohydrate transport mediated by the phosphoenolpyruvate-dependent phosphotransferase system (PTS). The PTS allows the use of different carbon sources and in many bacteria studied so far the PTS is a crucial link between metabolism and regulation of catabolic operons. For instance, in *Escherichia coli* and *Bacillus subtilis* it was shown that the PTS phosphorylation cascade has a central function in catabolite repression [11,12]. Orthologues of the *B. subtilis* genes involved in this regulation were also identified (Hpr kinase and CcpA), indicating that a similar mechanism for catabolite repression could exist in *Listeria*. The comparison between the *L. monocytogenes* and *L. innocua* genomes reveals the conservation of all ABC transporters but not of all PTS genes.

In contrast to many other bacterial genomes, the *L. monocytogenes* genome contains only three copies of one insertion sequence and that of *L. innocua* contains none. Both genomes contain bacteriophages, but these

do not seem to play a major role in acquisition of virulence genes as has been shown for other bacteria such as pathogenic *E. coli* [13].

3. Genome organisation

The comparison of the organisation of the two sequenced *Listeria* genomes allows insight into evolutionary mechanisms and phylogenetic relationships. Surprisingly, a perfect conservation of the order and the relative orientation of orthologous genes was observed showing a high stability in genome organisation since inversions or shifting of large genome segments have not been observed (Fig. 1A). This conserved genome organisation may be related to the low occurrence of insertion sequence (IS) elements, suggesting that IS transposition or IS-mediated deletions are no key evolutionary mechanisms in *Listeria*. Comparison of the two *Listeria* genomes to that of the non-pathogenic Gram-positive bacterium *B. subtilis* also belonging to the group of low G+C content bacteria reveals a high similarity among proteins. One thousand four hundred and twenty-eight genes were predicted to be orthologues (on the basis of bidirectional Blastp comparisons). Furthermore, a surprising synteny was observed in genome organisation (Fig. 1B). However, an exception is the 616-kb DNA region of *L. monocytogenes*, extending from *rpoC* to *ddlA*, which contains only few orthologues and only one conserved locus of 29 genes encoding functions required for chemotaxis and motility. The same results were obtained when comparing the genome organisation of *L. monocytogenes* and *Staphylococcus aureus* (Fig. 1C). In contrast, comparison of the *L. monocytogenes* genome with *Streptococcus agalactiae*, another phylogenetically closely related member of the group of Gram-positive low G+C content bacteria, or *Lactococcus lactis* showed no synteny (Fig. 1D,E). Orthologous genes were scattered around both chromosomes. Interestingly, in the 616-kb *rpoC*–*ddlA* region again only few orthologous genes were identified. This suggests a specific status for this part of the genome. In addition, this region is enriched in genes specific of each *Listeria* isolate. Thirty seven per cent (96 genes) of the *L. monocytogenes* EGDe-specific genes are clustered within these 616 kb which represent only 21% of the genome. Thus, this region may be a hot spot for horizontal gene transfer or the rest of the chromosome may be less tolerant to insertion of DNA. Furthermore, the whole genome comparisons suggest that high conservation in genome organisation is specific for the *Bacillus*, *Staphylococcus* and *Listeria* group.

4. Species-specific genes – horizontal gene transfer

Generation of genetic variability, occurrence of new phenotypes and selection of variants by environmental

forces represent the key elements of evolution. Horizontal gene transfer is a key factor in these processes. It includes the transmission of phages, transposons and plasmids as well as the uptake of DNA by naturally competent bacteria. Differences between the two listerial genomes seem to be due to all three mechanisms. The *L. innocua* strain sequenced carried an 80-kb plasmid, not present in *L. monocytogenes*, which is predicted to encode in particular resistance to different heavy metals such as arsenic or cadmium. A Tn916-like transposon and one bacteriophage are inserted in *L. monocytogenes* EGDe. *L. innocua* CLIP11626 contains five bacteriophages which represent about 8% of the genome. In both *Listeria* isolates a prophage of the A118 family is inserted in the gene *comK*. The ComK protein acts in *B. subtilis* as a transcriptional activator controlling the late competence genes required for the binding, processing and internalisation of transforming DNA [14]. However, *comK* is not interrupted in all *Listeria* isolates. Analysis of the genome sequence identified a nearly complete set of putative DNA uptake genes homologous to *B. subtilis* competence genes, although *Listeria* are not known to be competent. Only ComQ, ComS, ComX and ComFB, all of which are involved in regulation of competence in *B. subtilis*, are missing. Thus *Listeria* might be competent, but the signals that induce competence have not yet been identified. They may be different since, as mentioned above, the counterparts of several *B. subtilis* regulatory genes required for competence gene expression are missing in *Listeria*.

If prophage genes are excluded 270 (9.5%) *L. monocytogenes* EGDe-specific genes and 149 (5%) *L. innocua* CLIP11626-specific genes were identified. The *L. monocytogenes* EGDe-specific genes are present in 100 DNA fragments scattered throughout the entire chromosome. Interestingly, their G+C content varies from 24 to 46%. The average G+C content of all predicted coding regions is 38%, whereas that of the *Listeria*-specific genes is only 34% (Fig. 2). The G+C content of the genes compared to that of their flanking regions has been used as a marker for phylogenetic origin. In *L. monocytogenes*, 54 of the 100 specific regions had a significantly lower G+C content than the flanking regions and six had a significantly higher G+C content, suggesting recent acquisition by horizontal gene transfer. Fig. 3 shows an example of abrupt changes in G+C content of a *L. monocytogenes*-specific region carrying four genes of unknown function. These specific genes have an average G+C content of 28% whereas the flanking regions have G+C contents of 41 and 42%, respectively. The *L. innocua* CLIP11626-specific genes are clustered in 63 regions containing one to seven genes. Analysis of the G+C content of these regions gave similar results. Thirty have a lower and three a higher GC content than their flanking regions. Interestingly, many of the *L. innocua*-specific surface proteins are located in such regions with distinct G+C content, suggesting that they were acquired by horizontal gene transfer, probably from other bacteria

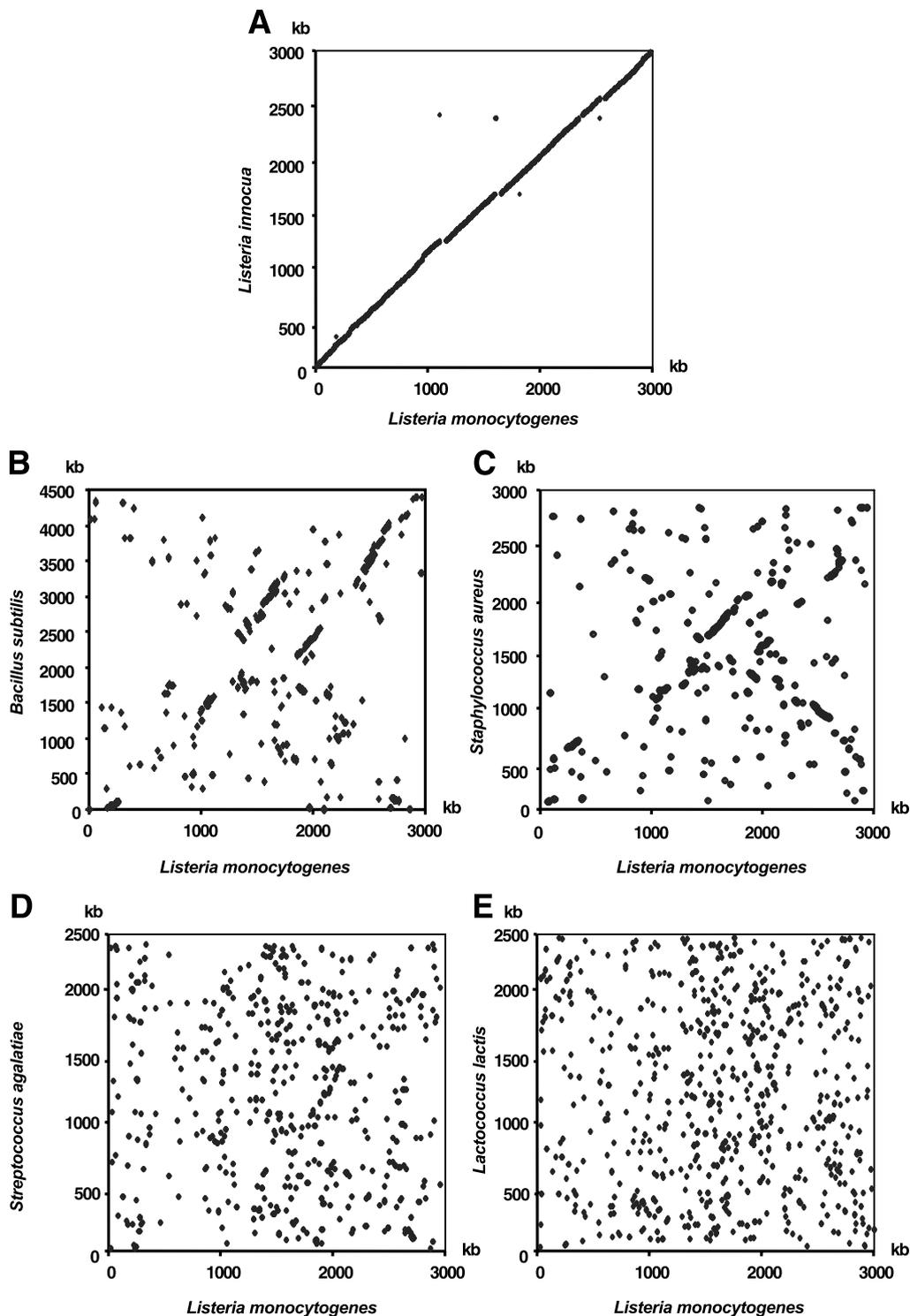


Fig. 1. Synteny between (A) *L. monocytogenes* and *L. innocua*, (B) *L. monocytogenes* and *B. subtilis*, (C) *L. monocytogenes* and *Staphylococcus aureus*, (D) *L. monocytogenes* and *Streptococcus agalactiae*, and (E) *L. monocytogenes* and *Lactococcus lactis*. Regions of synteny appear as continuous lines. Orthologous genes were defined by bidirectional best hits based on Blastp comparisons.

with lower G+C content such as *Streptococcus*. This particular organisation of a number of small regions within the *Listeria* genomes suggests that multiple acquisition but also deletion events have led to the present genome content. However, it can be questioned whether changes in

G+C content are reliable markers for horizontal gene transfer events. It is clear that they are at least not the only markers and that DNA fragments may also have been acquired by horizontal gene transfer but have over time adapted to the *Listeria* genome or originated from

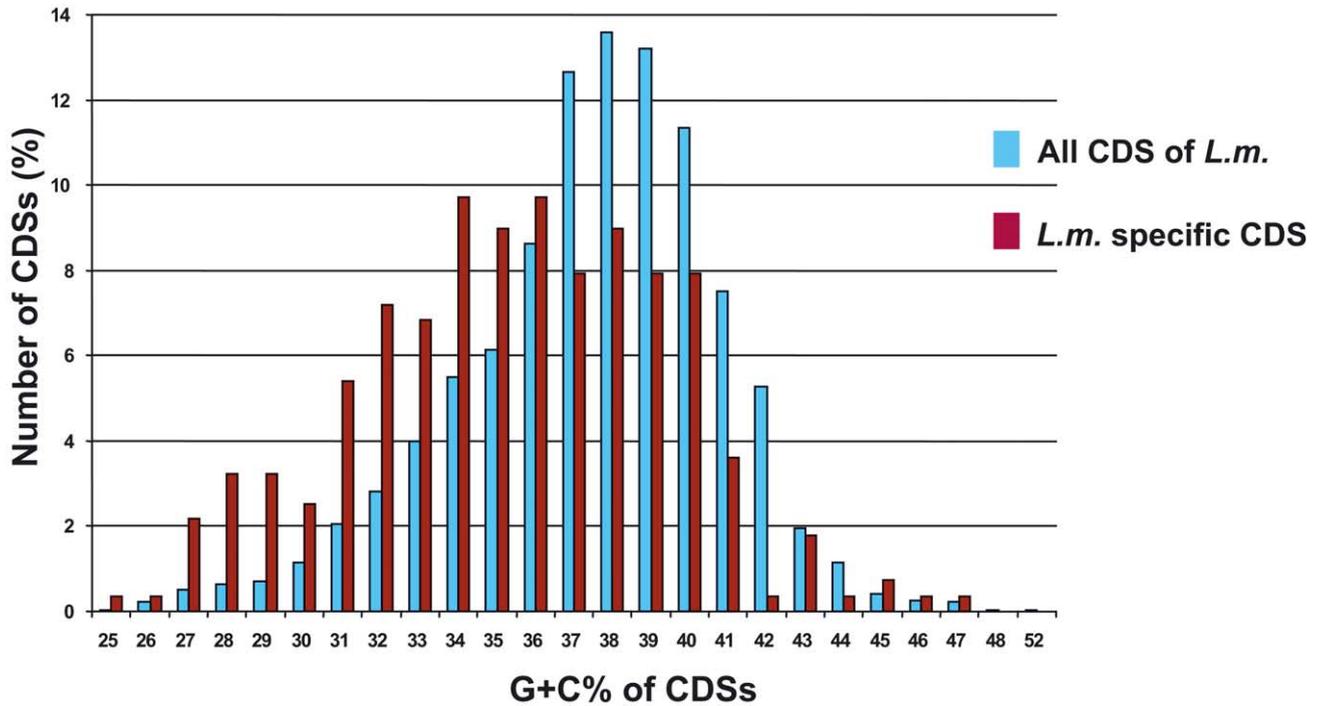


Fig. 2. Number and distribution of the predicted genes of *L. monocytogenes* according to their G+C content. Blue: all genes predicted in the *L. monocytogenes* EGDe genome, brown: genes specific for *L. monocytogenes* EGDe relative to *L. innocua* CLIP11626.

bacteria with similar G+C content. An example of such a region is a 55-kb genome fragment carrying genes necessary for anaerobic vitamin B₁₂ synthesis and the degradation of the carbon sources propanediol and ethanolamine.

5. Vitamin B₁₂ – anaerobic synthesis in *Listeria*?

It is generally accepted that *L. monocytogenes* is an aerobically growing microaerophilic (carbon dioxidophilic)

organism. *Listeria* thrives best at reduced oxygen tension and after replacement of oxygen by carbon dioxide growth is excellent. Under strict anaerobic conditions without addition of CO₂ growth in common culture media is scanty. However, *L. monocytogenes* is able to survive and to colonise the mammalian gut where it encounters anaerobic conditions. Analysis of the genome sequence identified in *L. monocytogenes* and in *L. innocua* a single continuous locus, encoding the proteins necessary for vitamin B₁₂ synthesis. Furthermore, the proteins necessary for degradation of the carbon sources ethanolamine and propanediol

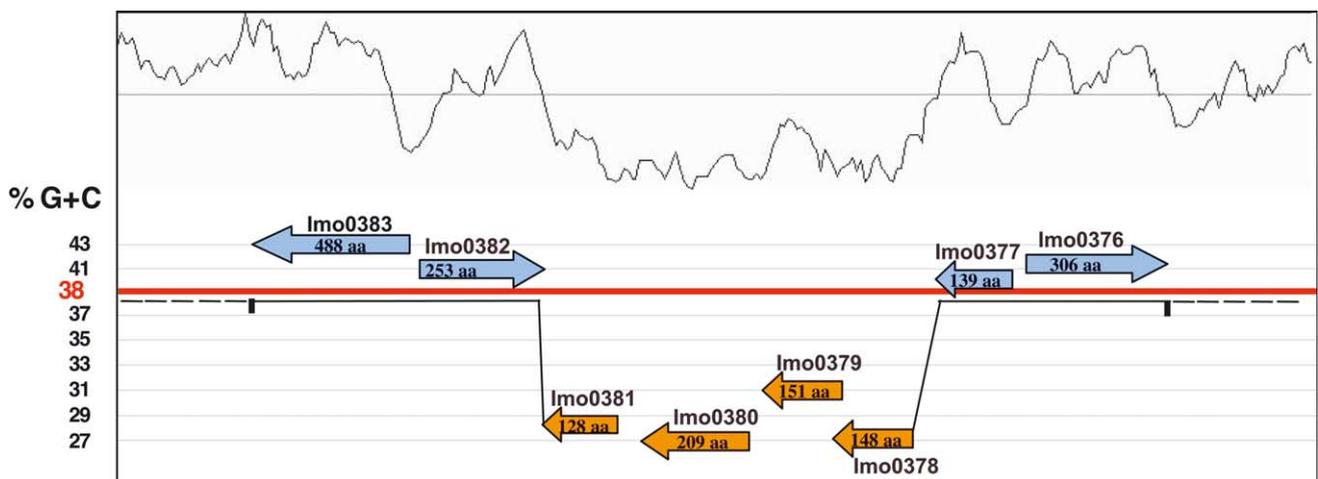


Fig. 3. Schematic representation of a *L. monocytogenes* EGDe-specific genome region. The scale indicates the G+C content of the respective genes. The black curve indicates variations in G+C content of this region. Orange: *L. monocytogenes* EGDe-specific genes, blue: genes present in both sequenced *Listeria* isolates. Numbers give the length of the deduced protein sequence. aa: amino acid, Imo numbers correspond to gene names on the ListiList web server <http://genolist.pasteur.fr/ListiList/>.

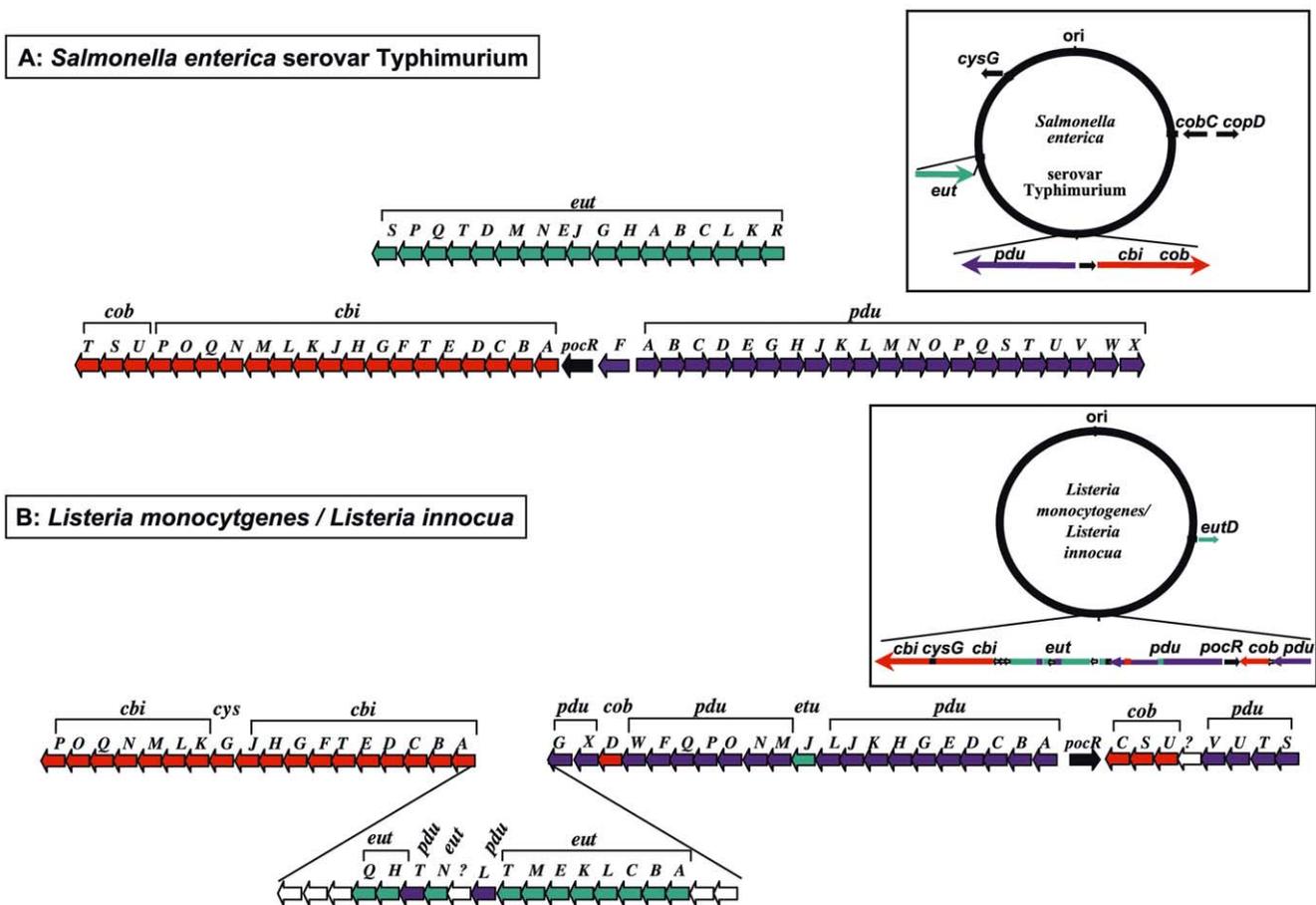


Fig. 4. Schematic representation of the vitamin B₁₂ biosynthesis, ethanolamine and propanediol gene clusters. A: Organisation in *S. enterica* serovar Typhimurium. B: Organisation in *L. monocytogenes* and *L. innocua*. Red: vitamin B₁₂ biosynthesis genes, blue: propanediol (*pdu*) gene cluster, green: ethanolamine (*eut*) gene cluster.

in a coenzyme B₁₂-dependent manner are encoded within the same region.

The ability to synthesise vitamin B₁₂ is unevenly distributed in living organisms. A search in currently available genome sequences reveals that vitamin B₁₂ biosynthesis genes are found in just over one third of the bacteria sequenced so far. An extensive description of the gene organisation in different bacteria is reviewed in Raux et al. [15]. There exist at least two routes for vitamin B₁₂ synthesis: an oxygen-independent pathway present for example in *Salmonella enterica* serovar Typhimurium and *Pseudomonas shermanii* and an oxygen-dependent pathway present in *Bacillus megaterium* or *Pseudomonas denitrificans* [15]. The proteins deduced from the genome sequence of *L. monocytogenes* and *L. innocua* share the highest homology with those of *S. enterica* serovar Typhimurium (38–65% protein identity). Genes encoding CbiD, CbiG and CbiK, specifically associated with the anaerobic pathway, are present in *Listeria*. This suggests that the two *Listeria* species also contain the oxygen-independent pathway like *Salmonella*. Furthermore, like in *Salmonella* this gene cluster is localised close to the predicted terminus of replication and has the same organisation except one gene, *cysG* encoding a putative uroporphyrinogen III methyl-

transferase, which in *Listeria* is localised within the vitamin B₁₂ biosynthesis gene cluster (Fig. 4). Upstream of the cobalamin biosynthesis genes *Listeria* contain orthologues of genes necessary in *Salmonella* for the coenzyme B₁₂-dependent degradation of ethanolamine and propanediol (Fig. 4B). These two gene clusters have a slightly different organisation. In *Salmonella* the ethanolamine gene cluster is located at minute 53 and the propanediol gene cluster is contiguous to the vitamin B₁₂ genes at minute 44 of the genome (Fig. 4A) [16]. In *Listeria* all three gene clusters seem to have been acquired en bloc by horizontal gene transfer as they are organised contiguously around the terminus of replication (Fig. 4B). However, acquisition might have been an ancient event, as the codon usage in this region is similar to that of the rest of the genome (I. Moszer, personal communication).

S. typhimurium synthesises vitamin B₁₂ anaerobically [17] and can use ethanolamine and 1,2-propanediol as the sole carbon and energy source for growth [18]. Anaerobic use of ethanolamine (*eut* operon) may be important for enteric bacteria since this carbon source is a constituent of an abundant class of lipids present in the host's gut. Propanediol (*pdu* operon) is produced by fermentation of the common plant sugars rhamnose and fucose [19]. Fu-

cose is also found in the glycoconjugates of intestinal cells, where it is involved in host–parasite interactions [20]. In vivo expression techniques have indicated that in *S. typhimurium* 1,2-propanediol utilisation (*pdu*) genes may be important for growth in host tissues, and competitive index studies with mice have shown that *pdu* mutations confer a virulence defect [21,22]. Taking into account the results obtained for *Salmonella*, it is tempting to assume that the vitamin B₁₂ synthesis genes and the *pdu* and *eut* operons play a role in anaerobic growth of *Listeria* and perhaps also in virulence of *L. monocytogenes*.

The above-mentioned examples show that the availability of the two *Listeria* genome sequences, their profound analysis and comparative genomics has opened new and exciting possibilities for functional genomics to decipher the molecular basis of the lifestyle of *Listeria*.

Acknowledgements

This work received financial support from EC contract BIO4CT980036, the Ministère de l'Éducation Nationale de la Recherche et de la Technologie (PRFMMIP) and the Pasteur Institut (PTR 6). P. Cossart is an international investigator from the Howard Hughes Medical Institute.

References

- [1] Schuchat, A., Deaver, K.A., Wenger, J.D., Plikaitis, B.D., Mascola, L., Pinner, R.W., Reingold, A.L. and Broome, C.V. (1992) *J. Am. Med. Assoc.* 267, 2041–2045.
- [2] Pinner, R.W. et al. (1992) *J. Am. Med. Assoc.* 267, 2081–2082.
- [3] Vazquez-Boland, J.-A. et al. (2001) *Clin. Microbiol. Rev.* 14, 1–57.
- [4] Farber, J.M. and Peterkin, P.I. (1991) *Microbiol. Rev.* 55, 476–511.
- [5] Lou, Y. and Yousef, A.E. (1998) in: *Listeria*, Listeriosis and Food Safety (Ryser, E.T. and Marth, E.H., Eds.). pp. 131–224. Marcel Dekker, New York.
- [6] Lecuit, M., Vandormael-Pournin, S., Lefort, J., Huerre, M., Gounon, P., Dupuy, C., Babinet, C. and Cossart, P. (2001) *Science* 5522, 1722–1725.
- [7] Pron, B., Boumaila, C., Jaubert, F., Berche, P., Milon, G., Geissmann, F. and Gaillard, J.L. (2000) *Cell Microbiol.* 5, 331–340.
- [8] Cossart, P. and Lecuit, M. (1998) *EMBO J.* 17, 3797–3806.
- [9] Glaser, P. et al. (2001) *Science* 294, 849–852.
- [10] Cabanes, D., Dehoux, P., Dussurget, O., Frangeul, L. and Cossart, P. (2002) *Trends Microbiol.* 5, 238–245.
- [11] Stulke, J. and Hillen, W. (1999) *Curr. Opin. Microbiol.* 2, 195–201.
- [12] Stulke, J. and Hillen, W. (2000) *Annu. Rev. Microbiol.* 54, 849–880.
- [13] Hacker, J., Blum-Oehler, G., Janke, B., Nagy, G. and Goebel, W. (1999) in: *Pathogenicity Islands and Other Mobile Virulence Elements* (Kaper, J.B. and Hacker, J., Eds.). pp. 59–77. ASM Press, Washington, DC.
- [14] Dubnau, D. (1997) *Gene* 192, 191–198.
- [15] Raux, E., Schubert, H.L. and Warren, M.J. (2000) *Cell. Mol. Life Sci.* 57, 1880–1893.
- [16] Roth, J.R., Lawrence, J.G. and Bobik, T.A. (1996) *Annu. Rev. Microbiol.* 50, 137–181.
- [17] Jeter, R.M., Olivera, B.M. and Roth, J.R. (1984) *J. Bacteriol.* 159, 206–213.
- [18] Price-Carter, M., Tingey, J., Bobik, T.A. and Roth, J.R. (2001) *J. Bacteriol.* 183, 2463–2475.
- [19] Obradors, N., Badia, J., Baldoma, L. and Aguilar, J. (1988) *J. Bacteriol.* 170, 2159–2162.
- [20] Bry, L., Falk, P.G., Midtvedt, T. and Gordon, J.I. (1996) *Science* 273, 1380–1383.
- [21] Heithoff, D.M., Conner, C.P., Hentschel, U., Govantes, F., Hanna, P.C. and Mahan, M.J. (1999) *J. Bacteriol.* 181, 799–807.
- [22] Conner, C.P., Heithoff, D.M., Julio, S.M., Sinsheimer, R.L. and Mahan, M.J. (1998) *Proc. Natl. Acad. Sci. USA* 95, 4641–4645.