Gene Fragments Distinguishing an Epidemic-Associated Strain from a Virulent Prototype Strain of *Listeria monocytogenes* Belong to a Distinct Functional Subset of Genes and Partially Cross-Hybridize with Other *Listeria* Species

MARTIN HERD AND CHRISTINE KOCKS*

Institute for Genetics, University of Cologne, D-50674 Cologne, Germany

Received 16 November 2000/Returned for modification 6 February 2001/Accepted 13 March 2001

Most major food-borne outbreaks of listeriosis in Europe and in the United States have been caused by genetically closely related *Listeria monocytogenes* strains of serotype 4b. In order to assess whether genomic loci exist that could underlie this increased epidemic potential, we subtracted the genome of the virulent prototype *L. monocytogenes* strain EGD from a prototype epidemic strain. A total of 39 DNA fragments corresponding to 20% of an estimated total of 150 to 190 kb of differential genome material were isolated. For 21 of these fragments, no function on the basis of homology could be predicted. Of the remaining 18 fragments, 15 had homologies to bacterial surface proteins, some of which have been implicated in virulence mechanisms such as cell invasion, adhesion, or immune escape. Southern hybridization of arrays containing the epidemic-clone-specific DNA segments with genomic DNA of different *L. monocytogenes* strains was consistent with the current lineage division. Surprisingly, however, some of the fragments hybridized in a mosaic-like fashion to genomes of two other *Listeria* species, the animal pathogen *L. ivanovii* and the nonpathogen *L. innocua*. Taken together, our results provide a starting point for the identification of epidemic-trait-associated genes.

The gram-positive, opportunistic bacterium *Listeria monocytogenes* is responsible for severe food-borne infections with a high case fatality rate in susceptible animal and human hosts (16, 44). Human listeriosis occurs as sporadic disease or in the form of outbreaks (epidemics) which can affect up to several hundred people, in particular elderly, very young, or immunocompromised individuals. Symptoms of primary infection range from barely apparent to distinct flu-like symptoms and self-limiting diarrhea. After an extended incubation period, susceptible individuals may develop listeriosis, which may manifest itself as meningitis, meningoencephalitis, or sepsis, or in the case of pregnancy as abortion, miscarriage, or severe generalized infection of the newborn.

Of the 16 or so *L. monocytogenes* serotypes known, three serotypes, 4b, 1/2b, and 1/2a, are responsible for most human cases (16, 44). Molecular and other typing methods divide the species *L. monocytogenes* into distinct phylogenetic lineages separating the serotypes 4b, 3b, and 1/2b from the serotypes 1/2a and 1/2c (6, 38, 42). Serotype 4b strains are responsible for the majority of sporadic disease as well as most major outbreaks of food-borne listeriosis in Europe and North America since 1981. By multiclocus enzyme electrophoresis, a specific 4b strain, representing 1 of 82 electrophoretic types of *L. monocytogenes*, was associated with epidemic disease outbreaks in California in 1985, Switzerland between 1983 and 1987, and France in 1992 (6, 26, 38). A 4b strain with a closely related electrophoretic type was implicated in two outbreaks in Boston in 1979 and 1983 (6, 38).

*L. monocytogenes* possesses a number of well-characterized virulence factors that play a role in its facultative intracellular lifestyle and its capacity to circumvent the humoral immune system by spreading from cell to cell within tissues (11, 45). Although no correlation between epidemic prevalence and increased virulence in a number of animal models and cell culture tests has been found to date for *L. monocytogenes* 4b strains (28, 46), one should nevertheless consider the possibility of genetic loci conferring additional pathogenic traits (12) to the strains prevalent in epidemics. These traits may include properties (not measurable by existing virulence tests) that allow the adaptation to specific environmental conditions and that can influence the course or the outcome of an infection.

One way to identify genetic loci encoding such traits is to employ bacterial genome subtraction (1, 9, 36). Despite the apparent genetic distinctness of *L. monocytogenes* serotype 4b strains, only three genes have been identified to date which are characteristic of serotype 4b strains. These genes were isolated on the basis of their involvement in expression of cell wall teichoic acid-associated, serotype 4b-specific surface antigens (30, 40).

In order to assess whether genetic loci exist that distinguish epidemic *L. monocytogenes* strains, we subtracted the genome of the virulent, experimentally best-characterized strain (EGD serotype 1/2a, whose completed genome sequence is pending publication) from a prototype epidemic strain (F.4565, serotype 4b) that had caused 142 cases of listeriosis including 48 deaths during an outbreak in Los Angeles in 1985 (31). We found that about 5 to 6% of the genome of strain F.4565 does not hybridize to strain EGD DNA. We have isolated 39 such fragments, 35 of which were absent and 4 of which were strongly divergent from the genome of strain EGD. A large portion of these DNA regions code for bacterial surface de-
terminators and hence are likely to play a role in the interaction with the environment or the host organism. We have devised a screening method to test the occurrence of these fragments in the genomes of other *Listeria* species or other *L. monocytogenes* strains, thereby opening the way to search for epidemic-trait-associated genes.

**MATERIALS AND METHODS**

**Bacterial strains.** *L. monocytogenes* strains F.4565 (bacterial collection no. [BAC] 239), F.4566 (BAC 248), F.4567 (BAC 249), F.4568 (BAC 250), and L.2190a (BAC 265), L.2292 (BAC 260), F.1092 (BAC 267), F.3533 (BAC 259), G.0039 (BAC 258), L.5089 (BAC 261), F.7390 (BAC 264), and L.735 (BAC 263) (7; see also Table 2) were provided by J. McLauchlin, Food Hygiene Laboratory, Central Public Laboratory, London, England. *L. monocytogenes* strain LO28 (BAC 089), *L. innocua* Rham’ (BAC 042) CLIP11262T, *L. innocua* (BAC 269) CLIP12511 (type strain), *L. innocua* (BAC 085) CLIP12510, *L. greyi* (BAC 273) ATCC 25401, and *B. subtilis* 168 mpc2 (BAC 274) were obtained from P. Cossart, Institut Pasteur, Paris, France. *L. monocytogenes* strain EGD (BAC 146) was obtained from W. Goebel, Theodor-Boveri-Institut, Würzburg, Germany. The *Escherichia coli* strain used was JM101 (BAC 098). *L. monocytogenes* LO28 ΔactA Δhly (BAC 200) is an isogenic mutant of strain LO28 lacking genes of *actA* and *hly* (C. Kocks, unpublished data).

**Preparation of genomic DNA.** Southern blotting, and “reverse” Southern blotting. Molecular biology methods were carried out using standard procedures (5) unless indicated otherwise. High-molecular-weight DNA was prepared from saturated liquid bacterial culture, taking care to remove all RNA. In order to verify the absence of plasmids, an alkaline lysis type plasmid DNA preparation was carried out with strains F.4565, L.4486a, *L. innocua* BAC 042, and *L. innocua* BAC 085 and the equivalent of up to 2 ml of saturated liquid culture loaded onto an agarose gel. For Southern blotting, 200 ng of RaI-digested genomic DNA was separated and blotted onto positively charged nylon membranes (Hybond N-1) for 4 hr at 40°C. For reverse Southern blotting, plasmid inserts were amplified by PCR using adapter-specific oligonucleotides. Approximately 20 ng of plasmid DNA or PCR products were spotted onto nylon membranes either by hand or by using a VAP08 Multi-Blot replicator (V&AP Scientific, Inc.) and hybridized to 25 ng of radioactively labeled, RaI-digested genomic DNA.

**Establishment of genomic subtraction conditions suitable for *Listeria* genomes.** *Hae*III-digested dX174 replicative-form (RF) DNA (5,389 bp) was added to *L. monocytogenes* EGD DNA (3,150 kb) at a molar ratio of 1:1 and substracted as described below. Assuming a similar fragment length distribution, the initial ratio of phage-specific fragments to other DNA fragments was 1:589. We cloned and sequenced the PCR-amplified material from this subtraction as described below. Four out of 18 randomly selected clones were identified as tester-specific dX174 RF *Hae*III fragments, corresponding to a ratio of 1:4.5. Thus, enrichment for tester-specific fragments was 130-fold. Similarly, enrichment factor (132-fold) was found in a second control subtraction in which DNA of *L. monocytogenes* strain LO28 was used as tester and DNA of an isogenic mutant, LO28 ΔactA Δhly lacking genes of *actA* and *hly*, was used as driver. In this case, the subtraction efficiency was assessed for an *actA* fragment by PCR as described below for *pLeA* up, and *hly*.

**Genome subtraction and plasmid library construction.** Genome subtraction was performed with the PCR-Select Bacterial Genome Subtraction kit (Clontech) according to the manufacturer’s instructions except for the hybridization temperature. Genomic DNA was digested with *RaI* and subtractive hybridization for carried out at 60°C, a temperature adapted to the low GC content of the *Listeria* genome (17). Tester-specific fragments were amplified by two rounds of suppression PCR that yielded a mixture of fragments corresponding in size to *RaI*-restricted genomic DNA. The subtraction efficiency was tested with PCR by comparing the concentration of *RaI*-digested fragments of three known genes, *pLeA* (KOT79/80 5′-TCCGCGTTACCTGGGAAATAGTAGGG-3′), *hly* (K03384 5′-CCGCGCTGCTGCGGAATCCG-3′), and *iap* (KOT77/78 5′-AAGAGGTTGACACTATTTTGGCG-3′-5′-TTTGTTGTGTGATGGTGCGAGG-3′) in equal amounts of PCR-amplified subtracted and unsubtracted material. The PCR products enriched for tester-specific fragments were ligated into plasmid vector pGEM-T Easy (system I; Promega) and transformed into competent *E. coli* cells. Plasmid DNA of 355 clones from the subtracted library and of 125 clones from the unsubtracted library was prepared with the QIAprep 8 Miniprep kit (Qiagen) and analyzed by restriction with EcoRI to test for the presence of inserts and to normalize plasmid DNA concentration. Insert sizes ranged from approximately 0.2 to 2 kb with an average length of 800 bp (consistent with the size distribution of *RaI*-fragments of genomic *L. monocytogenes* DNA).

**Sequence analysis and database searches.** Cycle sequencing was carried out using the BigDye Termination kit (PE Applied Biosystems) and an ABI Prism 377 DNA sequencer. Inserts were sequenced by one pass in each direction using vector-specific oligonucleotides. Homology searches at the nucleotide level (all nonredundant GenBank plus EMBL plus DDBJ plus PDB sequences) or at the amino acid level (all nonredundant GenBank CDS translations plus PDB plus SwissProt plus PIR plus PRF sequences) were carried out at the National Center for Biotechnology Information using BLASTN 2.0.12 and BLASTX 2.1.1 (4). Expect values lower than 0.05 were considered to be potentially significant. If necessary, internal fragment-specific oligonucleotides were used to obtain the complete sequence of fragments with homology to known genes. Fragments with homologies to unknown genes or no homology to database entries were sequenced to about 300 to 400 bp from both ends.

**Nucleotide sequence accession numbers.** The sequences described in Table 1 have been deposited in the EMBL database and assigned accession numbers AJ410352 to AJ410411. They can also be accessed at http://www.unl-koeln.de/~segel14/Kocks/index.html.

**RESULTS**

Five to 6% of the *L. monocytogenes* strain F.4565 genome did not hybridize to the genome of *L. monocytogenes* strain EGD. Genomic subtractions were carried out using a PCR-based subtractive hybridization method (1). Effective subtraction conditions for *L. monocytogenes* were established as described above in Materials and Methods. In order to identify DNA fragments specific for an epidemic clone of *L. monocytogenes* strain F.4565 (serotype 4b) (31) that belongs to a genetically closely related subgroup of strains that has caused major outbreaks of listeriosis in America and Europe (6, 26, 38), we subtracted the virulent prototype *L. monocytogenes* strain EGD (serotype 1/2a) from this epidemic strain. Subtraction efficiency for the *L. monocytogenes* genes *pLeA*, *hly*, and *iap* (all shared by both strains) ranged from 64-fold for *pLeA* to 8-fold for *hly* to no detectable subtraction for *iap* (data not shown).

Plasmids from libraries of PCR-amplified subtracted and unsubtracted material were spotted onto positively charged nylon membranes and hybridized with radioactively labeled whole-genome probes of tester (*L. monocytogenes* strain F.4565) and driver (*L. monocytogenes* strain EGD) DNA under low-stringency conditions. A total of 45 of 270 clones from the subtracted library versus 5 of 90 clones from the unsubtracted library did not hybridize with driver DNA, indicating threefold enrichment. Since the differentially hybridizing fragments from the nonsubtracted library were in the average size range (see Materials and Methods), one can conclude from the frequency of these clones (5 out of 90) that in the order of 5 to 6% of the genome of the epidemic *L. monocytogenes* clone strain F.4565 does not hybridize with the genome of *L. monocytogenes* strain EGD.

A total of 50 differentially hybridizing clones were isolated and sequenced. Seven of these turned out to be isolated more than once. The inserts of the 41 remaining, unique clones were amplified by PCR, cloned onto positively charged nylon membrane, and reanalyzed. Figure 1A shows that 39 clones hybridized differentially with the genome of the tester strain *L. monocytogenes* F.4565. To verify these results, 6 of the 39 fragments
were tested by Southern blotting under low-stringency conditions (data not shown). All six fragments hybridized specifically to the tester strain \((L.\ monocytogenes\ strain\ F.4565)\) and to another epidemic strain of serotype 4b \((L.\ monocytogenes\ strain\ L.4486a\ from\ the\ Swiss\ outbreak\ in\ 1983)\), while none of the fragments showed hybridization to the driver strain \((L.\ monocytogenes\ EGD,\ serotype\ 1/2a)\) or to \(L.\ monocytogenes\ strain\ L028\) (serotype 1/2c). Two of the six fragments cross-hybridized to different extents with \(L.\ innocua\) DNA (U46 and 143; see also Fig. 1 and below).

The absence of the 39 differentially hybridizing fragments from the EGD genome were confirmed by BLASTN and FASTA nucleotide similarity searches against the completed genome sequence of \(L.\ monocytogenes\ strain\ EGD-e\) (which corresponds to our driver strain). Only 4 of the 39 fragments showed significant homologies to EGD-e genome sequences, mainly in coding regions (European Consortium, personal communication). Two of these fragments (125 and 015) had at one end a short sequence region (of about one-sixth of their respective lengths) that had high homology to EGD-e sequences, indicating a junction between conserved and nonconserved F.4565 genome regions. Four differentially hybridizing fragments could be detected with FASTA (162, 60% identity in 394 nucleotides [nt]; 73, 66% identity in 395 nt; 128B, 69% identity in 341 nt; and 157, 71% identity in 416 nt [European Consortium, personal communication]). (Fragment number 157 showed a weak, barely detectable hybridization signal representing our detection limit.) Thus, in these cases, homologous sequences were present in the EGD-e genome but showed a high level of sequence diversity. In contrast to these findings, 13 nondifferential control fragments showed strong homologies to EGD-e sequences with high conservation ranging from 88 to 98% similarity (European Consortium, personal communication).

Many strain F.4565-specific gene fragments had homologies to surface proteins involved in virulence. The results of amino acid homology searches in translated public nucleotide or protein databases with the 39 translated \(L.\ monocytogenes\ strain\ F.4565\)-specific genome fragments are summarized in Table 1. A total of 21 fragments showed no significant homology to database entries (“no hit”) or were homologous to proteins of unknown function. Eleven fragments showed homologies to surface proteins of \(L.\ monocytogenes\) or to other gram-positive bacteria, including one to a leucine-rich protein interaction...
motif of a plant disease resistance gene, three to ABC transporters (fragments 29, 120, and 125), one to a locus involved in surface antigen expression, one to a component of a type II restriction system, and one to a metabolic enzyme. In contrast to these findings, when we analyzed 13 randomly selected non-differential clones (i.e., hybridizing to both the tester and driver *L. monocytogenes* strains) we found that only one was a “no hit” and only two had homology to bacterial surface components, whereas 10 had homology to various enzymes. The distribution of database hits in different protein categories is graphically displayed in Fig. 2. A total of 41% of the differential fragments showed homology to bacterial surface compo-

<table>
<thead>
<tr>
<th>Differential fragments&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Length (bp)</th>
<th>Best matches to homologous proteins (BlastX search)</th>
<th>Amino acid identity (%) (length)</th>
<th>Organism</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>167</td>
<td>378</td>
<td>Involved in expression of surface antigens</td>
<td>100 (378 bp)</td>
<td><em>L. monocytogenes</em> 10527 (serotype 4b)</td>
<td></td>
</tr>
<tr>
<td>157</td>
<td>484</td>
<td>Internalins D, H, A, G, and E (B repeats)</td>
<td>64 (477 bp)</td>
<td><em>L. monocytogenes</em></td>
<td></td>
</tr>
<tr>
<td>29B (←U37)</td>
<td>367</td>
<td><em>L. monocytogenes</em> antigen C</td>
<td>48 (225 bp)</td>
<td><em>L. monocytogenes</em></td>
<td></td>
</tr>
<tr>
<td>162</td>
<td>396</td>
<td>Internalins A, E, H, G, and D (leucine-rich repeats)</td>
<td>45 (393 bp)</td>
<td><em>L. monocytogenes</em></td>
<td></td>
</tr>
<tr>
<td>104 (←U58B)</td>
<td>400</td>
<td>Internalins B, H, E, D, and G (leucine-rich repeats)</td>
<td>33 (390 bp)</td>
<td><em>L. monocytogenes</em></td>
<td></td>
</tr>
<tr>
<td>143</td>
<td>239</td>
<td>Internalins B (C repeats)</td>
<td>33 (222 bp)</td>
<td><em>L. monocytogenes</em></td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>360</td>
<td>Amidase (repeats)</td>
<td>32 (239 bp)</td>
<td><em>L. monocytogenes</em></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>326</td>
<td>Receptor protein kinase-like protein</td>
<td>35 (240 bp)</td>
<td><em>Arabidopsis maniana</em></td>
<td></td>
</tr>
<tr>
<td>33B (←54B)</td>
<td>512</td>
<td>ORF A (repeats)</td>
<td>52 (510 bp)</td>
<td><em>Listeria seeliger</em></td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>523</td>
<td>Immunoreactive 92 kDa antigen PG21</td>
<td>47 (516 bp)</td>
<td><em>Listeria seeliger</em></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>880</td>
<td>Chitinase B (chIB) (cadherin-like repeat)</td>
<td>26 (447 bp)</td>
<td><em>Listeria seeliger</em></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>573</td>
<td>Membrane transporter homologue yeI</td>
<td>65 (567 bp)</td>
<td><em>B. subtilis</em></td>
<td>Foster (18)</td>
</tr>
<tr>
<td>83</td>
<td>720</td>
<td>Internalin B</td>
<td>100 (507 bp)</td>
<td><em>L. monocytogenes</em></td>
<td></td>
</tr>
<tr>
<td>181</td>
<td>1,143</td>
<td>Methyltransferase (R-M system)</td>
<td>57 (471 bp)</td>
<td><em>Lactococcus lactis</em></td>
<td>Twoomey et al. (47)</td>
</tr>
<tr>
<td>48B</td>
<td>525</td>
<td>Transketolase</td>
<td>62 (525 bp)</td>
<td><em>B. subtilis</em></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>565</td>
<td>u-Mannosidase</td>
<td>61 (565 bp)</td>
<td><em>B. halodurans</em></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>929</td>
<td>Catalytic control protein A</td>
<td>98 (558 bp)</td>
<td><em>L. monocytogenes</em></td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>359</td>
<td>UDP-N-acetylglucosamin-2-epimerase</td>
<td>64 (360 bp)</td>
<td><em>B. subtilis</em></td>
<td>Foster (18)</td>
</tr>
<tr>
<td>64</td>
<td>675</td>
<td>Topoisomerase IV subunit</td>
<td>49 (672 bp)</td>
<td><em>B. subtilis</em></td>
<td>Foster (18)</td>
</tr>
<tr>
<td>66</td>
<td>709</td>
<td>PdxT (1,2-propanediol catabolism)</td>
<td>32 (501 bp)&lt;sup&gt;a&lt;/sup&gt;</td>
<td><em>S. enterica</em></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>456</td>
<td>Ethanolamine utilization protein</td>
<td>47 (255 bp)&lt;sup&gt;a&lt;/sup&gt;</td>
<td><em>E. coli</em></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>483</td>
<td>Carbamate kinase</td>
<td>62 (447 bp)</td>
<td><em>Lactobacillus sakei</em></td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>~1,000</td>
<td>β-Glucosidase</td>
<td>56 (315 bp)&lt;sup&gt;a&lt;/sup&gt;</td>
<td><em>Streptomyces coelicolor</em></td>
<td></td>
</tr>
<tr>
<td>104B</td>
<td>533</td>
<td>comG operon protein 1 (DNA uptake)</td>
<td>48 (504 bp)</td>
<td><em>B. subtilis</em></td>
<td>Chung and Dubnau (10)</td>
</tr>
</tbody>
</table>

<sup>a</sup> Alignments at different locations.<br><sup>b</sup> Gene fragments homologous to virulence determinants, surface proteins, or transporters.<br><sup>c</sup> Gene fragments homologous to other proteins.<br><sup>d</sup> Gene fragments without homologies or homologous to proteins with unknown function (indicated with #) as of 29 Sept. 2000 included differential fragment numbers 7 (→1 kb); 15 (787 bp); 16 (192 bp); 17B (→1.7 kb); 18 (→0.9 kb); 55 (531 bp); 67B (734 bp); 73 (478 bp); 78B (→1.1 kb); 80B (→10B; 294 bp); 110 (→0.9 kb); 116 (313 bp); 128B (446 bp); 133 (→113; 748 bp); 136B (625 bp); 144B (532 bp); 149 (393 bp); 161 (→1.1 kb); 166 (697 bp); USB (625 bp); and U22B (→0.9 kb) and the nondifferential fragment number 69 (306 bp).
strains were divided into three groups (1/2a-I, 1/2a-II, and 1/2a-III), serotype 3b and 1/2b strains grouped together, and the 1/2c strain grouped together with the 1/2a-III strain (Table 2 and Fig. 3). Thus, according to the number of shared hybridization signals, 4b strains appeared more related to 3b and 1/2b strains than to 1/2a and 1/2c strains. These results reflect the established division of L. monocytogenes strains into two groups (one comprising 1/2a and 1/2c and one comprising the 4b, 1/2b, and 3b strains) on the basis of multifunctional enzyme electrophoresis, ribotyping, pulse-field gel electrophoresis, and restriction enzyme analysis (6, 8, 21, 38) and, additionally, into subgroups of 1/2a and 4b strains by more resolving methods, such as PCR-restriction enzyme analysis of the in1A-in1B region (14) or sequencing of the in1B gene (15).

Apart from controls, no significant hybridization signal after low-stringency washing was seen with genomic DNA of E. coli, B. subtilis (data not shown), or Listeria grayi, the member of the genus Listeria most distantly related to L. monocytogenes (Fig. 1B). However, DNA from the animal pathogen L. ivanovii and from nonpathogenic L. innocua strongly cross-hybridized to five and nine fragments, respectively, only two of which were common (Fig. 1B and Fig. 3). A second L. innocua strain showed an identical hybridization behavior (data not shown). These latter results raise the possibility that the genomes of at least three different species within the genus Listeria may have a mosaic-like composition with respect to these genes.

Table 2. Hybridization patterns of 39 L. monocytogenes F.4565-specific gene fragments with DNA from different L. monocytogenes strains

<table>
<thead>
<tr>
<th>Original no.</th>
<th>Country</th>
<th>Origin</th>
<th>Serotype</th>
<th>Outbreak</th>
<th>Hybridization pattern(s)</th>
<th>Remark</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>F.4565</td>
<td>USA</td>
<td>Patient, epidemic</td>
<td>4b</td>
<td>Los Angeles 1985</td>
<td>4b-I</td>
<td>Tester strain</td>
<td>Bille and Rocourt (7)</td>
</tr>
<tr>
<td>G.3129</td>
<td>USA</td>
<td>Patient, epidemic</td>
<td>4b</td>
<td>Los Angeles 1985</td>
<td>4b-I</td>
<td>Same outbreak as</td>
<td>F.4565</td>
</tr>
<tr>
<td>F.7441</td>
<td>USA</td>
<td>Patient, epidemic</td>
<td>4b</td>
<td>Lausanne 1987</td>
<td>4b-I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L.4486a</td>
<td>Switzerland</td>
<td>Patient, epidemic</td>
<td>4b</td>
<td>Lausanne 1987</td>
<td>4b-I</td>
<td>Same as L.2192</td>
<td>Bille and Rocourt (7)</td>
</tr>
<tr>
<td>L.2190a</td>
<td>England</td>
<td>Food, sporadic</td>
<td>4b</td>
<td>Boston 1983</td>
<td>4b-II</td>
<td>Same as L.2190a</td>
<td>Bille and Rocourt (7)</td>
</tr>
<tr>
<td>F.8353</td>
<td>USA</td>
<td>Food, sporadic</td>
<td>4b</td>
<td></td>
<td>4b-II</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G.0039</td>
<td>USA</td>
<td>Patient, sporadic</td>
<td>3b</td>
<td></td>
<td>3b and 1/2b</td>
<td></td>
<td>Bille and Rocourt (7)</td>
</tr>
<tr>
<td>L.5089</td>
<td>UK</td>
<td>Patient, sporadic</td>
<td>1/2b</td>
<td></td>
<td>1/2a-I</td>
<td></td>
<td>Bille and Rocourt (7)</td>
</tr>
<tr>
<td>F.7390</td>
<td>USA</td>
<td>Food, sporadic</td>
<td>1/2a</td>
<td></td>
<td>1/2a-II</td>
<td></td>
<td>Bille and Rocourt (7)</td>
</tr>
<tr>
<td>L.735</td>
<td>England</td>
<td>Patient, epidemic</td>
<td>1/2a</td>
<td>Carlisle 1981</td>
<td>1/2a-II</td>
<td>Driver strain</td>
<td>Bille and Rocourt (7)</td>
</tr>
<tr>
<td>EGD</td>
<td></td>
<td>Rabbit</td>
<td>1/2a</td>
<td></td>
<td>1/2a-III and 1/2c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LO28</td>
<td>Spain</td>
<td>Fecal isolate</td>
<td>1/2c</td>
<td></td>
<td>1/2a-III and 1/2c</td>
<td>Healthy pregnant carrier</td>
<td></td>
</tr>
</tbody>
</table>

* Strains with the same electrophoretic type (included for internal control).
* Strains with the same electrophoretic type (included for internal control).
* USA, United States of America; UK, United Kingdom.

**DISCUSSION**

Five to 6% of DNA segments randomly picked from libraries from the unsubtracted genome of L. monocytogenes strain F.4565 did not hybridize to strain EGD DNA, indicating that about 5% of the F.4565 genome is not present in strain EGD. Physical genomic maps are available for L. monocytogenes strains EGD (serotype 1/2a) (49) and Scott A (serotype 4b) (23) and have yielded genome sizes of 3,000 ± 50 and 3,210 ± 60 kb, respectively. Assuming that strain F.4565 has a genome size similar to that of strain Scott A, which has the same serotype, about 150 to 190 kb of its DNA would be absent or substantially different from the genome of strain EGD. With
an average RsaI fragment size of 0.8 kb, we would thus expect about 200 differential fragments to be present in the F.4565 genome. Of these, we have isolated 39 fragments corresponding to about one-fifth of the theoretically expected number. (The latter may explain why we failed to isolate the genes gtcA and gltAB, which are known to be present in serotype 4b but not in serotype 1/2a strains [30, 40].)

Our results are in line with interstrain genome comparison data available for other bacterial species. *E. coli* strains with different pathogenic potential or different host ranges were found to vary by 300 to more than 1,000 kb (9, 12, 37), and the genomes of different *Salmonella* sp. serovars can differ by greater than 900 kb (36). Recent pairwise whole-genome comparisons of *Helicobacter pylori* and *Chlamydia* isolates showed that strain-specific genes tend to cluster in distinct genomic regions of about 50 to 160 kb termed plasticity zones (3, 43), while *E. coli* O157:H7 and *E. coli* K-12 differ by hundreds of islands of DNA that are apparently introgressed into a shared sequence backbone (37). Also, the extent of genome diversification varies: while different *Chlamydia* isolates showed little strain-specific genome material (43), the two unrelated *H. pylori* isolates each had 6 to 7% strain-specific genes (3), and the *E. coli* O157:H7 and K-12 isolates had about 25 and 12% strain-specific genes, respectively (37).

*H. pylori* isolates seem to differ mainly in components of restriction modification systems, insertion sequences, and DNA repeats (1, 3), while the *E. coli* isolates do not seem to differ by functionally distinct groups of genes (37). This is in contrast to the situation in *L. monocytogenes*, where 41% of the strain-specific genes exhibited homology to proteins exposed on the bacterial surface, while for 54% of the genes, no function on the basis of sequence similarity could be predicted (Fig. 2). We found only one DNA segment with homology to a restriction modification system component: a methyltransferase from *L. lactis* sharing homology with the target recognition domain of *M. Sau3A* (47). The presence of this DNA fragment in the genomes of the different *L. monocytogenes* strains used by us (Table 2 and Fig. 3) correlated with resistance of their DNA to *Sau3A* restriction (data not shown). Resistance to *Sau3A* restriction has recently been substantiated for *L. monocytogenes* serotype 4b strains (50).

Based on the sequences of the large and small intergenic spacer regions between the 16S and 23S rRNA genes, *L. monocytogenes* strains of different serotypes are more related to each other than to *L. innocua* (20). The nucleotide diversity of the *sigmaB* and *iap* genes accessible in the databases is consistent with this interpretation (data not shown). Within the genus *Listeria*, the nonpathogen *L. innocua* is the species most closely related to *L. monocytogenes*, while the animal pathogen *L. ivanovii* and nonpathogen *L. grayi* are more distantly related (20). Surprisingly, we found that two partially overlapping sets of nine and five *L. monocytogenes* F.4565-specific fragments strongly hybridized with the genomes of *L. innocua* and *L. ivanovii*, respectively (Fig. 1 and 3). Also, the other serotype 4b-specific genes isolated to date, *gtcA* and *gltAB*, which are loci involved in cell wall teichoic acid-associated surface antigen expression, cross-hybridize to certain strains of *L. innocua* (29, 30, 40). While some of these fragments may have been lost by some lineages of *L. monocytogenes*, others may have been acquired by lateral gene transfer. Thus, our data raise the possibility that lateral gene transfer could play a role in the diversification of *L. monocytogenes* strains. Indeed, this has recently been substantiated for *gtcA* (29).

Repeat regions are preferential subjects of lateral gene

FIG. 3. Summary of hybridization profiles of strain F.4565-specific fragments. Black boxes indicate positive, grey boxes indicate weak, and white boxes indicate negative hybridization signals. The number of hybridizing fragments is indicated below each column, and weakly hybridizing fragments are shown in parentheses. The GC content of hybridizing fragments is indicated in white boxes indicate negative hybridization signals. The number of fragments. Black boxes indicate positive, grey boxes indicate weak, and
transfer events. Many of the F.4565-specific fragments had homologies to repetitive protein elements (Table 1). Repeats facilitate recombination events, which are thought to occur about 50-fold more frequently than point mutations and therefore to have a much higher impact on bacterial genome diversification, in both gram-negative (E. coli, Neisseria meningitidis) and gram-positive (Streptococcus pneumoniae) bacteria (22). In particular, the set of fragments cross-hybridizing with L. in noca shows homology to multidomain, repeat-containing-surface proteins such as internalins and chitinases (Fig. 3). Lateral gene transfer has been implicated in the generation and diversification of chitinases (especially the cadherin-like domain) (35) and proteins containing leucine-rich repeat short sequence motifs thought to be involved in specific protein-protein interactions (27). Moreover, mosaic internalin genes have been described (41). Consistent with this possibility, a large number of transducing Listeria phages have been isolated recently (24). It is also interesting that one of the nondifferential control fragments (no. 104B) shows some homology to a B. subtilis gene involved in DNA uptake (10).

Our hybridization analysis with arrays harboring the 39 F.4565-specific fragments against genomes of different L. monocytogenes strains suggests that L. monocytogenes strains can be subtyped by genomic differences (Fig. 3). Because of its higher resolution, this approach is superior to serotyping for serotype 1/2a and 4b strains. Thus, genomic differences could be exploited for typing L. monocytogenes strains by combinatorial PCRs or by genome hybridizations to premanufactured typing arrays.

Taken together, we have shown that epidemic L. monocytogenes strains differ substantially from the sequenced prototype strain of L. monocytogenes. Many of the isolated gene fragments have homology to bacterial surface components and are therefore likely to confer traits that provide selective advantages in the environment. While this may be a general phenomenon applying to strain diversification of virulent as well as nonviral bacterial strains, our results nevertheless provide a good starting point to search for epidemic-trait-associated genes. For example, the complete set of genes specifically present (or absent, since pathogenicity traits may also be due to lack of specific genes [36]) in an epidemic-trait-associated strain can now be identified and screened against a large number of genomes from different L. monocytogenes isolates and Listeria species. Identification of genes consistently present or absent in epidemic-associated L. monocytogenes strains will open the way for mutational and functional analyses to address the question of whether indeed there is a molecular basis for the increased pathogenic potential of epidemic L. monocytogenes strains.

ACKNOWLEDGMENTS

We are grateful to Jim McLaughlin, London, for providing many of the bacterial strains; to Ali Hemmati-Bivanlou, New York, N.Y., and Chris Wilson, Cambridge, England, for technical advice concerning DNA arrays; to our colleagues from the Institute for Genetics, Uli Böhm, Thorsten Klamp, and Gloria Esposito, for advice with subtractive techniques; to Rita Lange for DNA sequencing; and to Olga Vites and Jonathan Howard for stimulating criticism throughout the work. We thank the European consortium involved in sequencing the L. monocytogenes EGD-e genome for the possibility of analyzing the complete genome sequence prior to its release, especially P. Cossart, Paris, France, who made this possible, and P. Glaser, Paris, France, who kindly carried out the analysis and discussed the results with us. Many thanks also to colleagues whose comments have helped to improve the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft through Sonderschungsforschungszentrum 274.

REFERENCES


