nature genetics

Uncovering *Listeria monocytogenes* hypervirulence by harnessing its biodiversity

Mylène M Maury[1–3,11], Yu-Huan Tsai[4,5,11], Caroline Charlier[4–8], Marie Touchon[1,2], Viviane Chenal-Francisque[4,6,7], Alexandre Leclercq[4,6,7], Alexis Criscuolo[9], Charlotte Gaultier[4,5], Sophie Roussel[10], Anne Brisabois[10], Olivier Disson[4,5], Eduardo P C Rocha[1,2], Sylvain Brisse[1,2,12] & Marc Lecuit[4–8,12]

Microbial pathogenesis studies are typically performed with reference strains, thereby overlooking within-species heterogeneity in microbial virulence. Here we integrated human epidemiological and clinical data with bacterial population genomics to harness the biodiversity of the model foodborne pathogen *Listeria monocytogenes* and decipher the basis of its neural and placental tropisms. Taking advantage of the clonal structure of this bacterial species, we identify clones epidemiologically associated either with food or with human central nervous system (CNS) or maternal-neonatal (MN) listeriosis. The latter clones are also most prevalent in patients without immunosuppressive comorbidities. Strikingly, CNS- and MN-associated clones are hypervirulent in a humanized mouse model of listeriosis. By integrating epidemiological data and comparative genomics, we have uncovered multiple new putative virulence factors and demonstrate experimentally the contribution of the first gene cluster mediating *L. monocytogenes* neural and placental tropisms. This study illustrates the exceptional power in harnessing microbial biodiversity to identify clinically relevant microbial virulence attributes.

Mechanistic studies in the field of microbial pathogenesis have greatly contributed to major discoveries in life sciences and led to key diagnostic and therapeutic advances in the field of infectious diseases. Most pathogenesis studies have been conducted with reference strains, which have been exchanged between investigators over the years and for which experimental tools have been developed. The systematic mutagenesis of reference strains filtered by phenotypic *in vitro* and *in vivo* screens has been an exceptionally fruitful approach that led to the discovery and characterization of key genes and gene products involved in microbial virulence[1,2]. As the use of reference strains purposely normalizes the inherent genetic heterogeneity of a given microbial species, this approach intrinsically underestimates biodiversity and the ensuing heterogeneity in the virulence of pathogenic microbial species[3–6].

Many clinical infectious disease phenotypes remain poorly understood at the mechanistic level, and this results in part from the use of clinically irrelevant reference laboratory strains. We hypothesized that widening the focus and considering (i) on the microbial side, the species as a whole and (ii) on the host side, detailed human epidemiological and clinical data would offer a unique opportunity to uncover new virulence attributes associated with clinically relevant microbial phenotypes. We applied this approach to *L. monocytogenes*, a major foodborne pathogen and a widely recognized model microorganism[7–9], which causes two deadly complications—CNS and MN listeriosis.

L. monocytogenes is a highly heterogeneous species: it can be divided into four evolutionary lineages[10–12], 13 serotypes[13] and four PCR serogroups[14]. Multilocus sequence typing (MLST) further subdivides the above categories into clones, which are geographically and temporally widespread[15–17]. Because of the very high morbidity and case fatality rates associated with human listeriosis, extensive surveillance programs that include food control and exhaustive investigation of human cases have been implemented. Presently, all *L. monocytogenes* isolates are regarded as equally virulent by regulatory authorities, although there is evidence against this uniform view: lineage I or serotype 4b occurs more frequently among clinical isolates than lineage II or serotypes 1/2b, 1/2a and 1/2c, relative to the frequency of these categories in food[12,18–21]. Moreover, strains with reduced pathogenicity displaying truncated and non-functional virulence factors such as internalin are commonly isolated from food[21,22]. However, a large-scale and systematic analysis integrating detailed molecular epidemiological data and comparative genomics has not been performed.

RESULTS

Distribution of *L. monocytogenes* clones in food and clinical sources

In France, listeriosis is a notifiable disease, and a single national reference center receives prospectively all isolates of clinical and food origin to which the human population is exposed, ensuring the epidemiological

[1]Institut Pasteur, Microbial Evolutionary Genomics Unit, Paris, France. [2]CNRS, UMR 3525, Paris, France. [3]Paris Diderot University, Sorbonne Paris Cité, Cellule Pasteur, Paris, France. [4]Institut Pasteur, Biology of Infection Unit, Paris, France. [5]INSERM Unit 1117, Paris, France. [6]National Reference Centre for *Listeria*, Paris, France. [7]World Health Organization Collaborating Center for Listeria, Paris, France. [8]Paris Descartes University, Sorbonne Paris Cité, Institut Imagine, Necker–Enfants Malades University Hospital, Division of Infectious Diseases and Tropical Medicine, Assistance Publique–Hôpitaux de Paris (AP-HP), Paris, France. [9]Institut Pasteur, Center of Bioinformatics, Biostatistics and Integrative Biology, Paris, France. [10]Paris-Est University, ANSES (Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'Environnement et du Travail), Food Safety Laboratory, Maisons-Alfort, France. [11]These authors contributed equally to this work. [12]These authors jointly supervised this work. Correspondence should be addressed to S.B. (sylvain.brisse@pasteur.fr) or M.L. (marc.lecuit@pasteur.fr).

Received 6 August 2015; accepted 6 January 2016; published online 1 February 2016; doi:10.1038/ng.3501

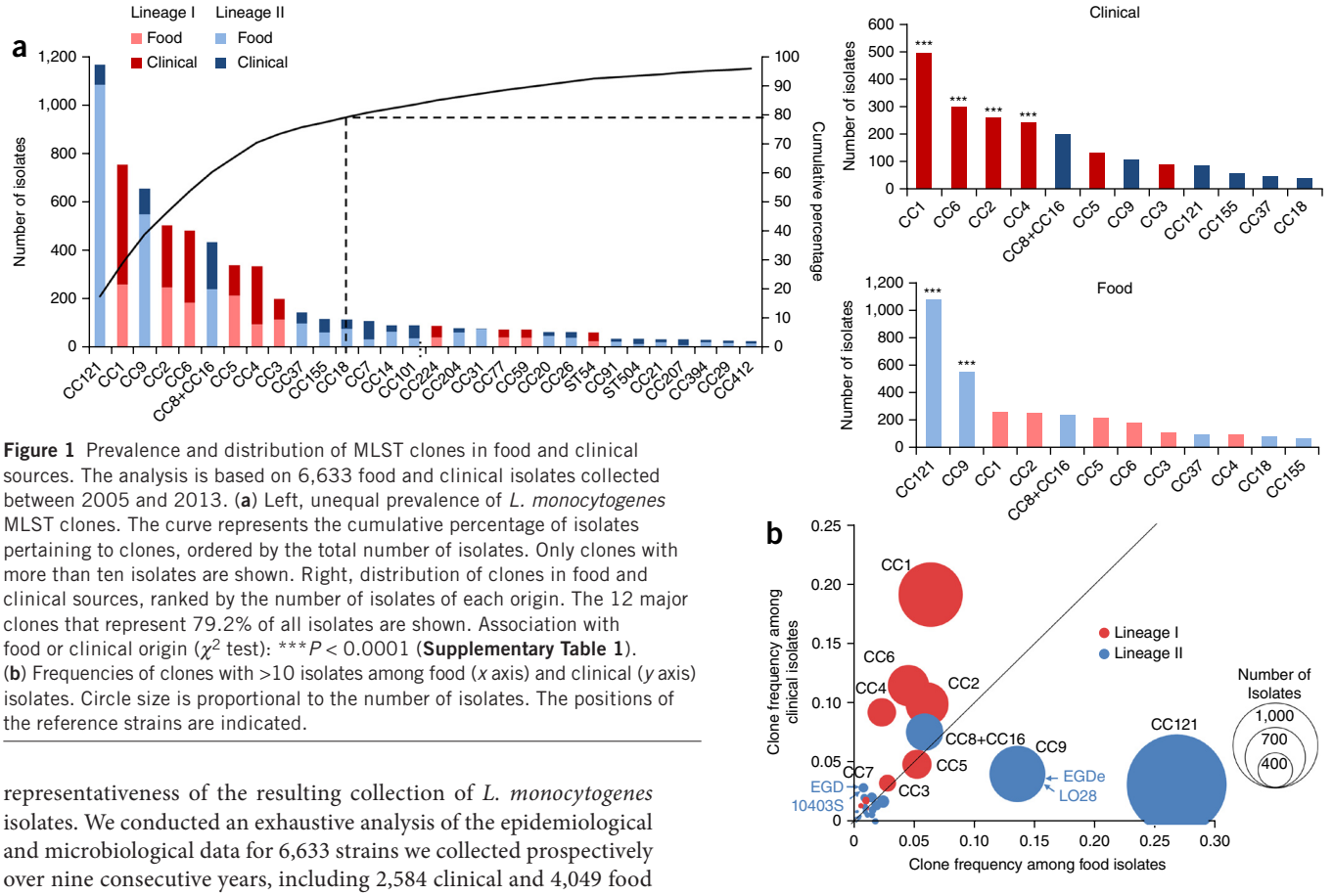

Figure 1 Prevalence and distribution of MLST clones in food and clinical sources. The analysis is based on 6,633 food and clinical isolates collected between 2005 and 2013. (**a**) Left, unequal prevalence of *L. monocytogenes* MLST clones. The curve represents the cumulative percentage of isolates pertaining to clones, ordered by the total number of isolates. Only clones with more than ten isolates are shown. Right, distribution of clones in food and clinical sources, ranked by the number of isolates of each origin. The 12 major clones that represent 79.2% of all isolates are shown. Association with food or clinical origin (χ^2 test): ***$P < 0.0001$ (**Supplementary Table 1**). (**b**) Frequencies of clones with >10 isolates among food (*x* axis) and clinical (*y* axis) isolates. Circle size is proportional to the number of isolates. The positions of the reference strains are indicated.

representativeness of the resulting collection of *L. monocytogenes* isolates. We conducted an exhaustive analysis of the epidemiological and microbiological data for 6,633 strains we collected prospectively over nine consecutive years, including 2,584 clinical and 4,049 food isolates. We ascribed a clone (defined as an MLST clonal complex (CC); Online Methods, **Supplementary Fig. 1** and **Supplementary Note**) to each of these isolates: the 6,633 strains belonged to 63 different clones, with the 12 most prevalent clones representing 79.2% of all strains in this study (**Fig. 1a**). Notably, the frequency distribution of these clones was highly uneven (**Fig. 1a**): the most prevalent clones were CC121 (17.6% of the total isolates), CC1 (11.4%), CC9 (9.9%), CC2 (7.6%), CC6 (7.3%), CC8 and CC16 (which were merged; 6.6%), CC5 (5.1%) and CC4 (5.1%). Remarkably, the clones that predominated in clinical samples were distinct from those in food samples (**Fig. 1a**), leading to striking differences in the relative prevalence of clones among clinical (**Fig. 1b**, *y* axis) and food (**Fig. 1b**, *x* axis) isolates (**Fig. 1a**,**b** and **Supplementary Table 1**). The proportion of clinical isolates per clone differed by one order of magnitude across the clones, ranging from 7.0% (CC121) to 71.3% (CC4) (**Fig. 2a**, *x* axis). Clones CC1, CC2, CC4 and CC6 were strongly associated with a clinical origin ($P < 1 \times 10^{-4}$; 62% of strains from these clones were of clinical origin), whereas clones CC121 and CC9 were strongly associated with a food origin ($P < 1 \times 10^{-4}$; only 10.3% were of clinical origin) (**Fig. 1a**,**b** and **Supplementary Table 1**). These results therefore distinguish three categories of highly prevalent clones: infection-associated clones (CC1, CC2, CC4 and CC6), food-associated clones (CC9 and CC121) and intermediate clones (others) (**Fig. 1a**,**b**). Strikingly, the reference strains EGDe and LO28 belonged to CC9, a food-associated clone very rarely causing human disease, whereas the reference strains EGD and 10403S belonged to CC7, which was very rarely isolated from food (0.9%) and infected patients (2.9%) (**Fig. 1b**).

Listeriosis can manifest as isolated bacteremia (where *L. monocytogenes* has crossed the intestinal barrier), MN infection (where *L. monocytogenes* has crossed the intestinal and placental barriers) and CNS infection (where *L. monocytogenes* has crossed the intestinal and blood-brain barriers). Of note, the most clinically associated clones were also the most associated with MN and CNS infections as opposed to isolated bacteremia (linear regression, $P < 1 \times 10^{-10}$; **Fig. 2a**). Most notably, CC1 and CC4 were most strongly associated with MN and CNS infections and were negatively associated with bacteremia ($P < 1 \times 10^{-5}$; **Supplementary Table 2**), whereas clones CC121, CC9 and CC8+CC16 were associated with low frequencies of MN and CNS infections ($P < 1 \times 10^{-3}$). Considering the three types of infection separately, CC1 was associated with CNS infections ($P < 1 \times 10^{-4}$), CC1, CC2 and CC4 were associated with MN infections ($P < 1 \times 10^{-3}$ for CC1 and CC2, $P < 1 \times 10^{-4}$ for CC4) and CC8+CC16, CC9 and CC121 were associated with bacteremia ($P < 1 \times 10^{-3}$ for CC121 and CC8+CC16, $P < 1 \times 10^{-4}$ for CC9) (**Fig. 2b** and **Supplementary Table 2**). Altogether, these results suggest that the strength of the association of a clone with clinical disease could be causally linked to its virulence.

Ecological distribution and virulence levels of clones

L. monocytogenes is an opportunistic pathogen that infects mostly immunocompromised individuals. We collected detailed clinical and biological data for 812 infected patients enrolled in the MONALISA prospective study on *Listeria* and listeriosis (ClinicalTrials.gov, NCT01520597) and analyzed the distribution of clones as a function of patients' immunosuppressive comorbidities. The food-associated clones CC9 and CC121 were more often isolated in highly immunocompromised patients, whereas CC1, CC2, CC4 and CC6 were more prevalent among patients with few or no immunosuppressive comorbidities (**Fig. 2c**). Strikingly, there was an inverse linear relationship between the predominance of infection-associated clones and the number of immunosuppressive comorbidities (R^2 = 0.96,

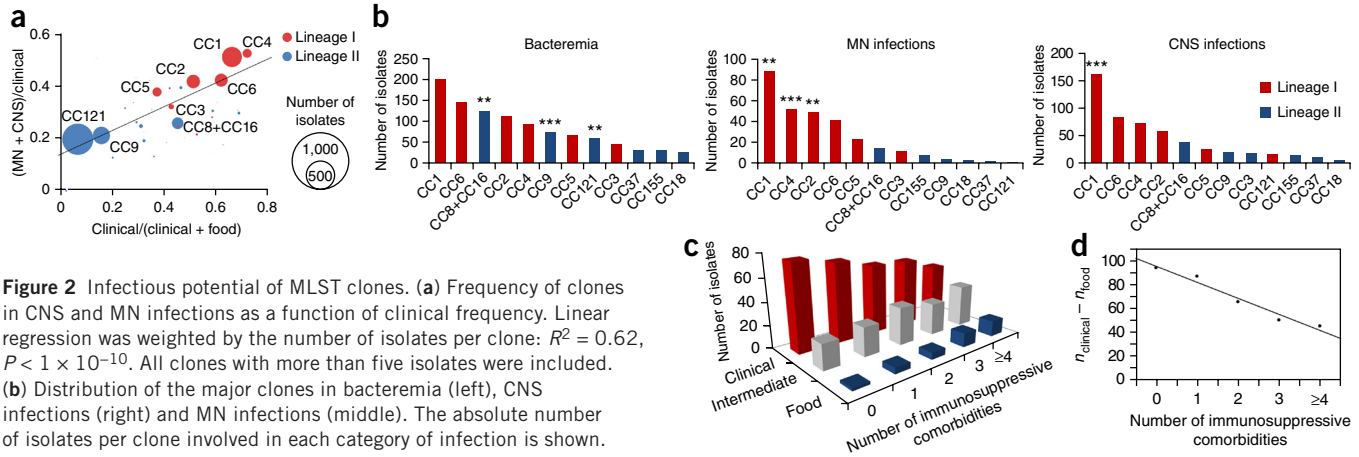

Figure 2 Infectious potential of MLST clones. (**a**) Frequency of clones in CNS and MN infections as a function of clinical frequency. Linear regression was weighted by the number of isolates per clone: $R^2 = 0.62$, $P < 1 \times 10^{-10}$. All clones with more than five isolates were included. (**b**) Distribution of the major clones in bacteremia (left), CNS infections (right) and MN infections (middle). The absolute number of isolates per clone involved in each category of infection is shown. The 12 major clones were ranked by the absolute number of isolates (**Supplementary Table 2**). Association with bacteremia, CNS infections and MN infections (χ^2 test): $^{**}P < 0.001$, $^{***}P < 0.0001$. (**c**) The histograms show the distribution of food-associated (food; CC9 and CC121; blue), infection-associated (clinical; CC1, CC2, CC4 and CC6; red) and intermediate (intermediate; CC8+CC16, CC5, CC3, CC37, CC155 and CC18; gray) clones in patient groups with different numbers of immunosuppressive comorbidities. (**d**) Linear regression ($R^2 = 0.96$, $P = 0.0032$) of the difference between the numbers of isolates belonging to infection-associated clones ($n_{clinical}$) and food-associated clones (n_{food}) (y axis) against the number of immunosuppressive comorbidities of infected patients (x axis).

$P = 0.0032$; **Fig. 2d**). As an infection results from the interplay between host and bacterial factors, these results indicate that specific virulence factors of these invasive clones may compensate for the absence of comorbidities to trigger disease and are in support of the hypothesis that infection-associated clones are hypervirulent.

We therefore assessed the respective virulence of infection-associated and non-infection-associated clones in a humanized mouse model of listeriosis[23], relative to that of the reference strains EGDe (CC9) and 10403S (CC7). Isolates belonging to clones CC1, CC4 and CC6 (**Supplementary Table 3**) induced significantly more body weight loss (**Fig. 3a**) and more efficiently infected the liver (CC1 and CC6) and brain (CC1, CC4 and CC6) than EGDe and 10403S, demonstrating that they are hypervirulent and, most notably, neurotropic in contrast to the reference strains (**Fig. 3b**). In contrast, isolates belonging to CC9 and CC121 (**Supplementary Table 3**), which were epidemiologically strongly associated with food but not with clinical infection, did not induce body weight loss following infection and were less

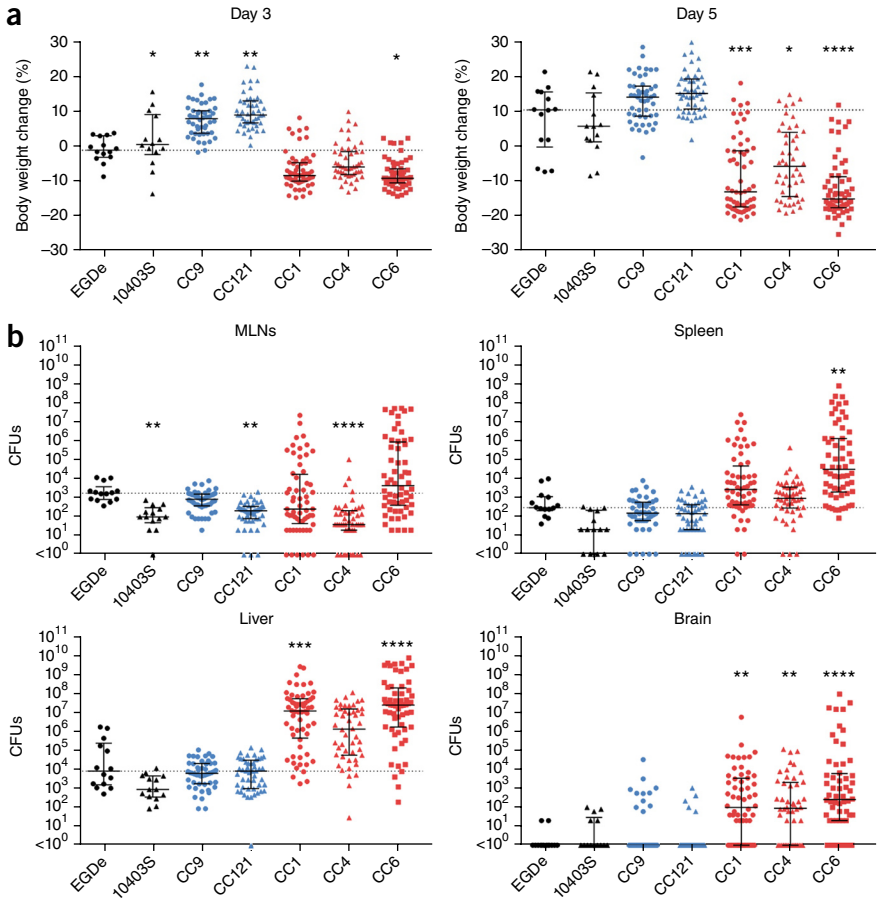

Figure 3 Comparative virulence of the six major clonal complexes. (**a**) Mouse body weight loss on day 3 and day 5 after infection. (**b**) Bacterial load on day 5 after infection is shown as the total number of CFUs recovered from entire organs. Humanized mice were orally inoculated with 2×10^8 CFUs. Results are shown as medians with the interquartile range. The dotted line indicates the median value for EGDe-infected mice. Two isolates from each origin—food, bacteremia, MN infection and CNS infection—were selected for each clone, except for CC1 and CC6, for which there are four human CNS infection isolates and two isolates for each of the other origins (food, bacteremia and MN infection). For CC121, there are two isolates from food, two from bacteremia, one from MN infection and three from CNS infection. Numbers of mice: $n = 14$ for EGDe and 10403S; $n = 48$ for CC9, CC121 and CC4; $n = 60$ for CC1 and CC6. Infection-associated clones are represented in red, food-associated clones are represented in blue and reference strains (EGDe and 10403S) are represented in black. Dunn's multiple-comparison test relative to EGDe-infected mice: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. The difference in comparison to EGDe-infected mice was not significant unless indicated. MLNs, mesenteric lymph nodes.

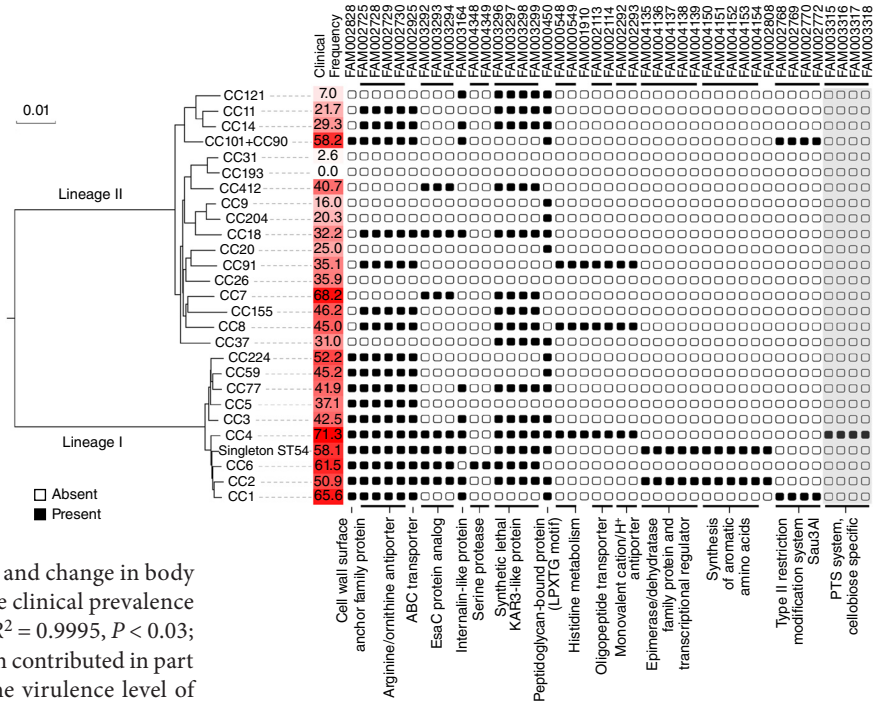

Figure 4 Phylogenetic distribution of new putative virulence factors identified in this study. Genes that were strongly associated with a high frequency of clinical isolates in clones (using generalized estimating equation (GEE) analysis; Online Methods) and had an identified putative function are shown. The recombination-purged phylogeny within each lineage was based on the respective core genome. The percentage of clinical isolates in each clone is highlighted by a red gradient. Gene families are named as in **Supplementary Table 9**. Groups of syntenic genes are indicated by black horizontal lines. Putative functions or pathways are indicated below. Only the four LIPI-4 genes with no paralogs in other genomes are shown (highlighted by gray shading). The scale bar indicates the number of substitutions per site.

invasive, demonstrating that they are hypovirulent (**Fig. 3a**,**b**). Remarkably, colony-forming unit (CFU) numbers in infected organs and change in body weight showed a very strong association with the clinical prevalence of the five clones (stepwise multiple regression, $R^2 = 0.9995$, $P < 0.03$; **Supplementary Table 4**). Although lineage origin contributed in part to this association, these results suggest that the virulence level of *L. monocytogenes* clones directly mirrors their epidemiological association with human listeriosis.

Genomic traits associated with *L. monocytogenes* hypervirulence

To decipher the origin of heterogeneity in virulence among clones, we analyzed the whole-genome sequences of 104 strains representative of the major clonal complexes that compose the *L. monocytogenes* species[15–17,24] (**Supplementary Fig. 2** and **Supplementary Table 5**). In *L. monocytogenes*, most virulence-associated genes identified thus far belong to its core genome[25–28]. We therefore computed the core genome for the *L. monocytogenes* species (**Supplementary Fig. 3** and **Supplementary Table 6**) and analyzed the distribution and variation of known virulence-associated gene products among clones. The distribution of internalin (InlA) truncations was the main feature associated with loss of virulence in hypovirulent clones (**Supplementary Fig. 4**). Additional variation was observed in virulence-associated genes (**Supplementary Figs. 4** and **5**, and **Supplementary Table 7**) and could account, at least in part, for differences in virulence among clones[28].

Putative virulence factors specific to hypervirulent clones (CC1, CC2, CC4 and CC6), including factors involved in *L. monocytogenes* neuroinvasiveness, might have been overlooked by the pathogenesis studies carried out thus far, as these factors are absent in reference strains that happen to belong to clones rarely responsible for human clinical cases (CC9 and CC7). This hypothesis would fit with the observation that reference strains LO28, EGDe, EGD and 10403S, which all belong to these clones, are poorly neuroinvasive (**Fig. 3b** and M.L., Y.-H.T. and O.D., unpublished data). We therefore determined the distribution among clones of all gene families of the *L. monocytogenes* pan-genome (**Supplementary Fig. 3** and **Supplementary Table 8**). To determine the evolutionary pattern of gene families and their correlation with the infection/food ratio of clones, the 1,791 genes of the core genome (**Supplementary Table 6**) were used to construct a recombination-purged phylogeny (**Supplementary Note**). The prevalence of clinical isolates in clones displayed strong phylogenetic inertia ($\lambda = 0.9999$, $P < 0.001$), indicating that the virulence potential of clones is determined, at least in part, by vertically transmitted features (**Fig. 4**). We therefore corrected for this evolutionary inertia[29] to correlate the pattern of the presence and absence of 6,867 gene families of the pan-genome with the infection/food ratio of clones (**Supplementary Table 9**). Strikingly, this analysis identified full-length InlA, *Listeria* pathogenicity island 3 (LIPI-3, or the listeriolysin S cluster) and gene clusters responsible for teichoic acid biosynthesis in serotype 4b strains as being strongly associated with infectious potential at the population level (**Supplementary Table 9**). As these features were previously demonstrated to be involved in *L. monocytogenes* virulence[30–32], this validated our algorithm and indicated that other features strongly associated with infectious potential represent good candidates for novel virulence factors. These features included protein families of defined putative function (**Fig. 4**) and also many uncharacterized putative gene products. CC1- and CC4-specific features scored among the most strongly correlated with infection, as expected by the high infection/food ratio of these hypervirulent clones.

LIPI-4, a locus involved in neural and placental infection

The above epidemiological and experimental data show that CC4 comprises the highest proportion of clinical isolates of all *L. monocytogenes* clones (71.3%; **Figs. 2a** and **4**), is hypervirulent *in vivo* as compared to EGDe and 10403S (**Fig. 3**), and is strongly associated with MN and CNS infections ($P < 1 \times 10^{-5}$; **Supplementary Table 2**). Prominent among 19 CC4-associated genes (**Supplementary Tables 8** and **9**) is a cluster of six genes annotated as a cellobiose-family phosphotransferase system (PTS) (**Supplementary Fig. 6a**). Carbon metabolism modulates virulence in *L. monocytogenes*[33]. We therefore investigated the contribution of this putative sugar transport system to CC4 hypervirulence. A representative CC4 strain, LM09-00558, was chosen to construct a PTS deletion mutant (**Supplementary Fig. 6b**). This mutant was not impaired for growth in culture medium (data not shown). In an oral humanized mouse model of infection, the wild-type CC4 strain was able to infect the CNS more efficiently than EGDe, whereas deletion of the entire

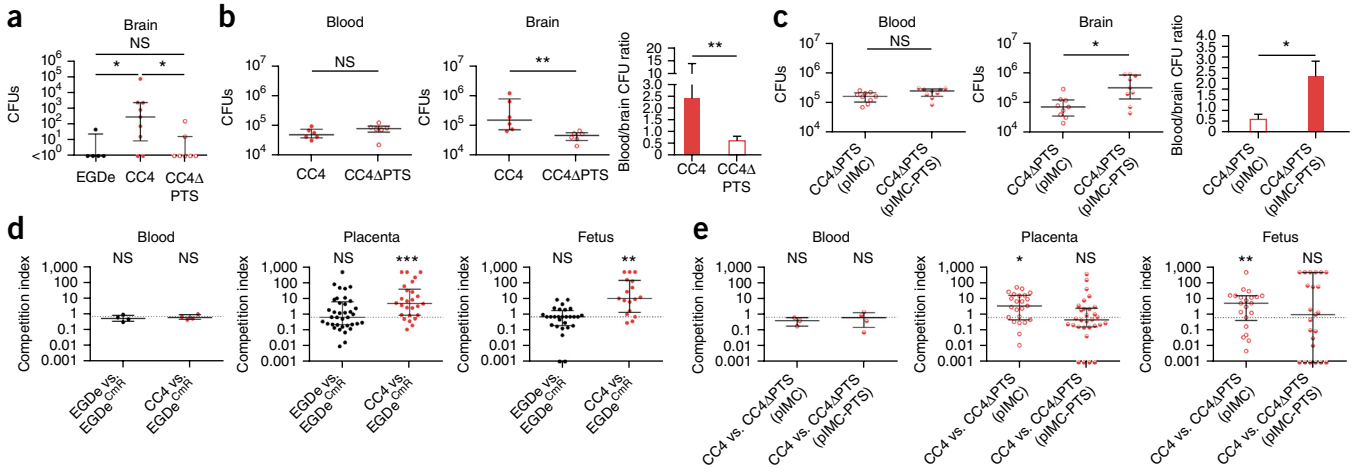

Figure 5 Implication of the PTS cluster (LIPI-4) associated with the hypervirulent clone CC4 in CNS and placental infections. (**a**,**b**) Humanized mice were inoculated orally (dose = 3×10^8 CFUs) (**a**) or intravenously (dose = 5×10^5 CFUs) (**b**) with strain EGDe ($n = 5$ in **a**), CC4 strain LM09-00558 (CC4; $n = 9$ in **a**, $n = 6$ in **b**) or LM09-00558 with the whole PTS cluster deleted (CC4ΔPTS; $n = 7$ in **a**, $n = 6$ in **b**). (**c**) Humanized mice were intravenously infected (dose = 5×10^5 CFUs) by the CC4ΔPTS strain containing either a single copy of pIMC ($n = 9$) or pIMC with the PTS cluster under the control of its native promoter (pIMC-PTS) ($n = 9$ for each group). (**d**) The competition index for wild-type EGDe ($n = 4$) and wild-type LM09-00558 (CC4; $n = 4$) when tested against chloramphenicol-resistant EGDe (containing pIMC; $EGDe^{CmR}$) in pregnant humanized mice. (**e**) The competition index for wild-type LM09-00558 (CC4) when tested against chloramphenicol-resistant CC4ΔPTS (pIMC) ($n = 3$) or CC4ΔPTS (pIMC-PTS) ($n = 4$) in pregnant humanized mice. In **d** and **e**, pregnant mice at day 14 of gestation were intravenously infected with a 1:1 mixture of the two strains as indicated (total dose = 2×10^5 CFUs). Mice were euthanized on day 5 after infection for orally inoculated bacteria (**a**) or on day 2 after infection when intravenously inoculated (**b**–**e**). Results are shown as medians with interquartile range. Each dot represents an organ (**a**–**c**) or blood from one infected mouse (**a**–**e**) or one placenta or fetus (**d**,**e**). Statistical analyses were carried out by Dunn's multiple-comparison test (**a**), Mann-Whitney *U* test (**b**,**c**) or Wilcoxon matched-pairs signed-rank test (**d**,**e**): $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$; NS, not significant.

PTS gene cluster (ΔPTS) considerably reduced the CNS invasion of the CC4 strain without affecting bacterial colonization of other tissues (**Fig. 5a** and **Supplementary Fig. 7a**). In an intravenous infection model where the intestinal barrier is bypassed, we demonstrated that the PTS system directly contributes to neuroinvasion (**Fig. 5b** and **Supplementary Fig. 7b**). Reinserting a single copy of the PTS cluster on the chromosome complemented this phenotype (**Fig. 5c** and **Supplementary Fig. 7c**). These results demonstrate the role of this CC4-associated PTS system in *L. monocytogenes* neuroinvasiveness. We further tested CC4 infectivity in a mouse MN infection model, given the strong association of this clone with human MN infection (**Fig. 2b** and **Supplementary Table 2**), using a competition index method[23]. Remarkably, the representative CC4 strain LM09-00558 infected better than EGDe the placentas and fetuses of pregnant humanized mice but neither maternal organs nor maternal blood (**Fig. 5d** and **Supplementary Fig. 7d**), mirroring the epidemiological association of CC4 with MN infection in humans. To understand the involvement of the CC4-associated PTS system in MN infection, we performed competitive index experiments using either a PTS deletion mutant (ΔPTS) or isogenic PTS-complemented strain as compared to the wild-type CC4 strain. Whereas the ΔPTS strain showed a significant defect in placental and fetal infection as compared to its parental wild-type CC4 strain, the PTS-complemented ΔPTS strain did not (**Fig. 5e** and **Supplementary Fig. 7e**). These results therefore identify the CC4-associated PTS cluster as the first *L. monocytogenes* virulence factor specifically implicated in CNS and MN infection. They indicate that the presence of this PTS locus accounts, at least in part, for the hypervirulence and enhanced CNS and MN tropism of CC4, which is strongly epidemiologically associated with both CNS and MN infection in humans. We propose to name this cluster *Listeria* pathogenicity island 4 (LIPI-4) (ref. 31). The absence of LIPI-4 from CC1 and CC6 suggests that other factors contributing to their hypervirulence remain to be characterized (**Fig. 4** and **Supplementary Tables 8** and **9**).

DISCUSSION

The epidemiological association of serotype 4b and lineage I of *L. monocytogenes* with clinical disease has been documented[18,19]. Here, taking advantage of our fine-grained genetic analysis of *L. monocytogenes* strain biodiversity filtered against exhaustive epidemiological and clinical data, we have been able to (i) establish the stratification of *L. monocytogenes* virulence at the phylogenetic level of clones, (ii) identify a whole series of new putative virulence factors in *L. monocytogenes* and (iii) demonstrate the involvement of one of them, encoding a putative PTS system (LIPI-4), in *L. monocytogenes* hypervirulence. This CC4-associated gene cluster is, to our knowledge, the first *L. monocytogenes* factor specifically implicated in the selective tropism of *L. monocytogenes* for the CNS and fetal-placental unit and is therefore of high clinical relevance. These findings echo the results of the pioneering comparative approach for *Listeria* at the interspecies level[25] and demonstrate the power of integrating within-species biodiversity with epidemiological, clinical and experimental approaches to discover new virulence genes associated with specific and clinically relevant phenotypes[34]. This study establishes that *L. monocytogenes* is a highly heterogeneous species with regard to pathogenicity and is composed of hypervirulent and hypovirulent clones. Moreover, hypervirulent clones are those most likely to cause disease and, in particular, CNS and MN listeriosis. This indicates that the clonal structure of the *L. monocytogenes* species should be taken into consideration in the surveillance of this major foodborne pathogen, which represents nearly half of deaths associated with foodborne infections in Western countries. This study opens a fresh perspective on *L. monocytogenes* pathogenicity and calls for the designation of novel reference strains, representative of the strains involved in actual human infections.

URLs. International MLST database, http://bigsdb.web.pasteur.fr/; guidelines of the European Commission for the handling of laboratory animals (directive 86/609/EEC), http://ec.europa.eu/environment/chemicals/lab_animals/home_en.htm; CLCbio assembler software, http://www.clcbio.com/products/clc-genomics-workbench. All genome sequences were submitted to the European Molecular Biology Laboratory–European Bioinformatics Institute (EMBL-EBI) and are available at http://www.ebi.ac.uk/. All public genomes used in this study are available at ftp://ftp.ncbi.nih.gov/genomes/ (last accessed in February 2013).

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. All genome sequences were submitted to EMBL-EBI (see URLs). Accession codes are listed in **Supplementary Table 5**. An umbrella project was created to group all reads and assemblies associated with the genomes sequenced in this study and is available by using accession PRJEB10817.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank C. Soto Alvarez, G. Pontdeme, T. Cantinelli and L. Diancourt for their contributions to MLST data production and analysis, and S. Roche for providing low-virulence strains for genome sequencing. We also thank D. Mornico (Center of Bioinformatics, Biostatistics and Integrative Biology of the Institut Pasteur) for his help with the submission of genome reads and assemblies. This study was funded by the Institut Pasteur, INSERM, from the French government's Investissement d'Avenir program, Laboratoire d'Excellence 'Integrative Biology of Emerging Infectious Diseases' (grant ANR-10-LABX-62-IBEID), the European Research Council (ERC), ERANET Proantilis, the Programme Hospitalier de Recherche Clinique MONALISA and the Programme de Recherche Translationnelle (PTR) ANSES–Institut Pasteur. Listeriosis surveillance in France is funded by the Institut de Veille Sanitaire (InVS) and the Institut Pasteur.

AUTHOR CONTRIBUTIONS

M.L. and S.B. conceived, supervised and directed the project. *L. monocytogenes* isolates were collected and characterized by A.L. in the context of French National Reference Center for *Listeria* activities, with the help of V.C.-F., as well as A.B. and S.R. Methods for clone identification were developed by M.M.M. and S.B. Epidemiological analyses were performed by M.M.M. and S.B. Clinical data collection and analysis was conducted by C.C. and M.L. Statistical analyses were performed by M.M.M. and E.P.C.R. Comparative genomics analyses were performed by M.M.M., M.T. and E.P.C.R. Phylogenetic analyses were performed by M.M.M., A.C. and M.T. Y.-H.T. generated mutant CC4 strains. *In vivo* experiments were performed by Y.-H.T., O.D. and C.G. M.M.M., S.B. and M.L. wrote the manuscript, with contributions from Y.-H.T., C.C., A.L., M.T. and E.P.C.R.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Falkow, S., Isberg, R.R. & Portnoy, D.A. The interaction of bacteria with mammalian cells. *Annu. Rev. Cell Biol.* **8**, 333–363 (1992).
2. Cossart, P., Boquet, P., Normark, S. & Rappuoli, R. Cellular microbiology emerging. *Science* **271**, 315–316 (1996).
3. Welch, R.A. *et al.* Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **99**, 17020–17024 (2002).
4. Holden, M.T. *et al.* Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. *Proc. Natl. Acad. Sci. USA* **101**, 9786–9791 (2004).
5. Hensel, M. *et al.* Simultaneous identification of bacterial virulence genes by negative selection. *Science* **269**, 400–403 (1995).
6. Parkhill, J. *et al.* Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat. Genet.* **35**, 32–40 (2003).
7. Tilney, L.G. & Portnoy, D.A. Actin filaments and the growth, movement, and spread of the intracellular bacterial parasite, *Listeria monocytogenes*. *J. Cell Biol.* **109**, 1597–1608 (1989).
8. Lecuit, M. Human listeriosis and animal models. *Microbes Infect.* **9**, 1216–1225 (2007).
9. Cossart, P. Illuminating the landscape of host-pathogen interactions with the bacterium *Listeria monocytogenes*. *Proc. Natl. Acad. Sci. USA* **108**, 19484–19491 (2011).
10. Piffaretti, J.C. *et al.* Genetic characterization of clones of the bacterium *Listeria monocytogenes* causing epidemic disease. *Proc. Natl. Acad. Sci. USA* **86**, 3818–3822 (1989).
11. Wiedmann, M. *et al.* Ribotypes and virulence gene polymorphisms suggest three distinct *Listeria monocytogenes* lineages with differences in pathogenic potential. *Infect. Immun.* **65**, 2707–2716 (1997).
12. Orsi, R.H., den Bakker, H.C. & Wiedmann, M. *Listeria monocytogenes* lineages: genomics, evolution, ecology, and phenotypic characteristics. *Int. J. Med. Microbiol.* **301**, 79–96 (2011).
13. Seeliger, H.P.R. & Jones, D. in *Bergey's Manual of Systematic Bacteriology* Vol. 2 1235–1245 (Williams & Wilkins, 1986).
14. Doumith, M., Buchrieser, C., Glaser, P., Jacquet, C. & Martin, P. Differentiation of the major *Listeria monocytogenes* serovars by multiplex PCR. *J. Clin. Microbiol.* **42**, 3819–3822 (2004).
15. Ragon, M. *et al.* A new perspective on *Listeria monocytogenes* evolution. *PLoS Pathog.* **4**, e1000146 (2008).
16. Chenal-Francisque, V. *et al.* Worldwide distribution of major clones of *Listeria monocytogenes*. *Emerg. Infect. Dis.* **17**, 1110–1112 (2011).
17. Haase, J.K., Didelot, X., Lecuit, M., Korkeala, H. & Achtman, M. The ubiquitous nature of *Listeria monocytogenes* clones: a large-scale Multilocus Sequence Typing study. *Environ. Microbiol.* **16**, 405–416 (2014).
18. McLauchlin, J. Distribution of serovars of *Listeria monocytogenes* isolated from different categories of patients with listeriosis. *Eur. J. Clin. Microbiol. Infect. Dis.* **9**, 210–213 (1990).
19. Gray, M.J. *et al. Listeria monocytogenes* isolates from foods and humans form distinct but overlapping populations. *Appl. Environ. Microbiol.* **70**, 5833–5841 (2004).
20. Ward, T.J., Ducey, T.F., Usgaard, T., Dunn, K.A. & Bielawski, J.P. Multilocus genotyping assays for single nucleotide polymorphism–based subtyping of *Listeria monocytogenes* isolates. *Appl. Environ. Microbiol.* **74**, 7629–7642 (2008).
21. Jacquet, C. *et al.* A molecular marker for evaluating the pathogenic potential of foodborne *Listeria monocytogenes*. *J. Infect. Dis.* **189**, 2094–2100 (2004).
22. Nightingale, K.K., Windham, K., Martin, K.E., Yeung, M. & Wiedmann, M. Select *Listeria monocytogenes* subtypes commonly found in foods carry distinct nonsense mutations in *inlA*, leading to expression of truncated and secreted internalin A, and are associated with a reduced invasion phenotype for human intestinal epithelial cells. *Appl. Environ. Microbiol.* **71**, 8764–8772 (2005).
23. Disson, O. *et al.* Conjugated action of two species-specific invasion proteins for fetoplacental listeriosis. *Nature* **455**, 1114–1118 (2008).
24. Chenal-Francisque, V. *et al.* Optimized multilocus variable-number tandem-repeat analysis assay and its complementarity with pulsed-field gel electrophoresis and multilocus sequence typing for *Listeria monocytogenes* clone identification and surveillance. *J. Clin. Microbiol.* **51**, 1868–1880 (2013).
25. Glaser, P. *et al.* Comparative genomics of *Listeria* species. *Science* **294**, 849–852 (2001).
26. Hain, T. *et al.* Pathogenomics of *Listeria* spp. *Int. J. Med. Microbiol.* **297**, 541–557 (2007).
27. den Bakker, H.C. *et al.* Comparative genomics of the bacterial genus *Listeria*: genome evolution is characterized by limited gene acquisition and limited gene loss. *BMC Genomics* **11**, 688 (2010).
28. Kuenne, C. *et al.* Reassessment of the *Listeria monocytogenes* pan-genome reveals dynamic integration hotspots and mobile genetic elements as major components of the accessory genome. *BMC Genomics* **14**, 47 (2013).
29. Paradis, E. & Claude, J. Analysis of comparative data using generalized estimating equations. *J. Theor. Biol.* **218**, 175–185 (2002).
30. Lecuit, M. *et al.* A transgenic model for listeriosis: role of internalin in crossing the intestinal barrier. *Science* **292**, 1722–1725 (2001).
31. Cotter, P.D. *et al.* Listeriolysin S, a novel peptide haemolysin associated with a subset of lineage I *Listeria monocytogenes*. *PLoS Pathog.* **4**, e1000144 (2008).
32. Faith, N. *et al.* The role of *L. monocytogenes* serotype 4b *gtcA* in gastrointestinal listeriosis in A/J mice. *Foodborne Pathog. Dis.* **6**, 39–48 (2009).
33. Eisenreich, W., Dandekar, T., Heesemann, J. & Goebel, W. Carbon metabolism of intracellular bacterial pathogens and possible links to virulence. *Nat. Rev. Microbiol.* **8**, 401–412 (2010).
34. Bille, E. *et al.* A chromosomally integrated bacteriophage in invasive meningococci. *J. Exp. Med.* **201**, 1905–1913 (2005).

ONLINE METHODS

Isolate selection for analysis of the source distribution of clones. Source distribution analysis was performed on a non-redundant collection of 7,342 isolates of food (n = 4,551) and clinical (n = 2,791) origin collected in France between January 2005 and October 2013. The 7,342 isolates were collected by the French National Reference Center (NRC) for *Listeria* (n_{total} = 6,804; n_{food} = 4,013 and $n_{clinical}$ = 2,791) and the French National Reference Laboratory (NRL) for *Listeria* ($n_{total} = n_{food}$ = 538).

The NRL for *Listeria* collects food isolates in the context of targeted controls in food industries. The NRC for *Listeria* collects nearly all isolates involved in human infection cases in France, amounting to an average of 360 strains per year. This high level of exhaustiveness is due to the fact that listeriosis is subject to mandatory declaration in France. The NRC for *Listeria* also collects food isolates involved in nearly 80% of food alerts, which are triggered by food being on the market and presenting a risk because of the presence of *Listeria*. Nearly 700 food or environmental strains are collected each year by the NRC for *Listeria* in the context of food alerts.

Among the food isolates of our collection, 3,143 (69.1% of all food isolates) were isolated from food alerts. Additional food strains were included: the NRC for *Listeria* collected food strains in the context of investigations following neurological forms of listeriosis (n = 178; 3.9%) and strains from checks by food industries (n = 692; 15.2%), and the NRL for *Listeria* collected strains from food surveillance activities (n = 538; 11.8%). The absence of redundancy among strains involved in food alerts was achieved by keeping, from our initial collection, only one isolate of those sharing the same date, food source of isolation, food alert number and MLST clone (after application of the pulsed-field gel electrophoresis (PFGE)-MLST dictionary). Other food isolates were also deduplicated on the basis of date of isolation and food source. In total, 1,772 potential duplicates of food isolates were eliminated to obtain the final data set, which was composed of 7,342 food and clinical isolates (no animal or environmental isolates were included).

Definition of cases. An MN infection was defined as a case of listeriosis in which *L. monocytogenes* was isolated from blood, from a normally sterile site in a pregnant woman or a newborn less than 28 d of age, or from a placenta, a fetus or a stillbirth. Only one isolate was considered when *L. monocytogenes* was cultured from both the mother and newborn. In all other cases, listeriosis was considered non-MN: bacteremia cases were those with detection of *L. monocytogenes* in blood with no evidence of CNS and MN infection, and CNS infections were cases where *L. monocytogenes* was isolated from the cerebrospinal fluid or blood in a patient with CNS clinical symptoms.

Pulsed-field gel electrophoresis. See the **Supplementary Note**.

Multilocus sequence typing, clone definition and lineage assignment. MLST was performed as described by Ragon *et al.*[15]. New alleles and profiles were incorporated into the international MLST database (see URLs). For clonal complex assignments, we pooled the MLST data from this study and from previous ones[15–17,24,35]. Clonal complexes were defined as groups of allelic profiles sharing at least six of seven genes with at least one other member of the group[15]. One exception was made for CC14 and CC91, as they corresponded to two separate clones that were merged as the result of a single intermediate allelic profile (ST206) and as the ApaI and AscI PFGE types of CC14 and CC91 were largely different from each other. ST206 was included in CC14. For public genomes, sequence types were deduced from the genomic sequences. The main phylogenetic lineage of each isolate was defined on the basis of a phylogenetic tree inferred from the concatenated sequences of the seven MLST gene fragments.

PCR serogrouping. See the **Supplementary Note**.

High-confidence identification of MLST clones on the basis of pulsed-field gel electrophoresis profiles. See the **Supplementary Note**.

Statistical tests performed to identify associations of clones with food and clinical sources. To estimate the significance of associations of clones with food, infections or clinical sources, χ^2 tests were used. For each clone, a table of contingency was created (**Supplementary Note**). The thresholds of significance were adapted to take into account multiple tests using sequential Bonferroni correction[36]. The principle is to perform a χ^2 test to obtain a P value for each comparison. The tests are then ordered from those with the lowest to the highest P values. The test with the lowest P value is performed first with a Bonferroni correction taking into account all tests[37]. The second test is performed with a Bonferroni correction involving one fewer test and so on for the remaining tests.

To quantify the relationship between the clinical frequency of *L. monocytogenes* clones and their capacity to cause MN or CNS infections, linear regression analyses were used. These tests were conducted with the lm function implemented in the basic R distribution through comparing, by clone, the frequency of clinical isolates among the total number of isolates and the frequency of isolates involved in MN and CNS infections among the total number of clinical isolates. To avoid biases due to rare clones, we applied weights taking into account the total number of isolates in each clone.

Immunosuppressive comorbidity analysis. MONALISA is a French prospective national cohort that included microbiologically confirmed cases of invasive listeriosis from November 2009 to July 2013 (ClinicalTrials.gov, NCT01520597). For each patient, clinical data, including medical background, and samples, including the clinical isolate, were collected after obtaining written informed consent. The study received institutional review board approval by a local ethical committee (Comité de Protection des Personnes Ile de France 3), according to French legislation.

The immunosuppressive conditions taken into account included reported daily alcohol intake of >3 glasses on any day, cancer, congenital immune deficiency, diabetes, cirrhosis, hemodialysis for end-stage kidney disease, bone marrow transplantation, solid organ transplantation, hematological malignancies, preexisting lymphopenia, preexisting neutropenia, giant cell arteritis, systemic lupus erythematosus, rheumatoid arthritis, spondylarthritis, inflammatory bowel disease, other autoimmune disease, asplenia, age of >70 years, HIV infection, prescription of corticosteroids and prescription of other immunosuppressive treatments in the last 5 years.

To test for an association between the number of immunosuppressive comorbidities and clones associated with clinical or food origin, we performed linear regression between (i) the difference between the number of isolates belonging to infection-associated clones (CC1, CC2, CC4 and CC6) and food-associated clones (CC9 and CC121) and (ii) the number of immunosuppressive conditions of patients infected by these isolates. We performed the same analysis for the infection- associated clones (CC1, CC2, CC4 and CC6) and the intermediate ones (CC8+CC16, CC5, CC3, CC37, CC155 and CC18).

Strains used to assess the virulence levels of major clones *in vivo*. We assessed the virulence levels of the three clones having the highest clinical frequencies (CC1, CC4 and CC6; **Fig. 2a**) as well as the two clones having the lowest clinical frequencies (CC9 and CC121; **Fig. 2a**) in a humanized mouse model. For each clone, we selected a minimum of eight isolates, which included strains from each of the following origins: food, CNS infections, MN infections and bacteremia. Isolates used for the *in vivo* experiments are indicated in **Supplementary Table 3**.

Mouse infection. Mouse experiments were performed as described[38]. We used 7- to 10-week-old mEcad E16P KI female mice in a C57BL/6 genetic background. Six mice were used for each tested strain. As all mice were identical, the allocation of mice to experimental groups by randomization was not relevant, and no blinding to mouse identity was employed. Mice had their food restricted overnight with free access to water. *L. monocytogenes* cultures were prepared as described and used to inoculate mice intragastrically via a feeding needle[38]. Mice were then immediately allowed free access to food and water. Pregnant mice were infected at day 14 of gestation as described[23]. Intravenous infections were performed by injecting 200 µl of bacterial suspension into the tail vein. Competition index experiments were performed by mixing a chloramphenicol-resistant pIMC-containing strain and a chloramphenicol-sensitive strain without pIMC at a 1:1 ratio. At the indicated times, mice were euthanized and organs were homogenized in PBS. Serial dilutions of the homogenates were plated onto brain-heart infusion (BHI) agar plates

doi:10.1038/ng.3501

with or without chloramphenicol. Competition indexes were calculated by dividing the number of chloramphenicol-sensitive CFUs by the number of chloramphenicol-resistant CFUs. All procedures were in agreement with the guidelines of the European Commission for the handling of laboratory animals, directive 86/609/EEC (see URLs), and were approved by the Animal Care and Use Committee of the Institut Pasteur, as well as by the ethical committee of Paris Centre et Sud under 2010-0020. Statistical analysis was performed by Dunn's multiple-comparison test relative to EGDe-infected mice: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

Statistical tests of association between experimental data and clinical frequency. We performed stepwise multiple regression of the average number of CFUs recovered in mesenteric lymph nodes, the spleen, the liver and the brain (transformed by $\log_x + 1$) as well as the average change in body weight on days 3 and 5 after infection of mice against the percentage of clinical isolates per clone. Regression was carried out using the Bayesian information criterion (BIC) to identify which of the variables most explained the clinical frequency of clones.

Strain selection for whole-genome sequencing. A total of 69 strains were selected for Illumina sequencing to represent species diversity on the basis of MLST and PFGE typing (**Supplementary Fig. 1** and **Supplementary Table 5**). A minimum of two isolates with distinct PFGE profiles and isolated from distinct sources were selected for each major clone, when possible. Thirty-seven isolates were from human clinical infections, 16 were from animal infections, 12 were from food sources, one was from the environment and three were of unknown origin. Thirty-five public genomes of good quality available at the time of the study were included in the analysis (**Supplementary Table 5**). Ten additional public draft genomes available at the time of the analysis were excluded because of unsatisfactory quality of the sequences, as assessed by the number of contigs and detected coding sequences, the average size of the coding sequences and the number of coding sequences shared with EGDe. In total, 104 genomes were included, comprising 41 genomes of lineage I, 57 genomes of lineage II, five genomes of lineage III and one genome of lineage IV. These genomes represented five singletons and 34 clonal complexes.

Genome sequencing. Genomic DNA for the 69 strains was extracted using the Promega Wizard genomic DNA purification kit. Genomes were sequenced using the Illumina HiSeq 2000 system with a 2 × 100-nt paired-end strategy. Quality trimming of reads and adaptor clipping were performed using AlienTrimmer[39]. *De novo* assembly was performed on the final set of reads using the CLCbio assembler (see URLs) with a minimum contig size of 500 nt. The Mauve Contig Mover program[40] was used to reorder contigs with completely sequenced genomes as references, using the F2365 genome for lineage I and the EGDe genome for lineage II. Genome sequences were submitted to the MicroScope/MaGe platform (Genoscope, Evry, France)[41] for gene prediction and assignment of gene product functions. To homogenize gene definitions with the newly sequenced genomes, genes that were present but not annotated in the 35 public genomes in GenBank (see URLs; last accessed in February 2013) were added using the MicroScope/MaGe platform.

Core genome definition. The core genome corresponds to the pool of genes ubiquitously found in all strains of the species. A core genome was defined first for the whole species using the 104 genomes, and two additional core genomes were then computed separately for lineage I ($n = 41$ genomes) and lineage II ($n = 57$ genomes). Orthologs were first identified as reciprocal best hits when using end gap–free global alignments between the proteome of a reference genome and the proteome of each of the genomes included in the analysis[42]. The reference genome used for the core genomes of the whole species and lineage II was EGDe, and the F2365 genome was used for the core genome of lineage I. Hits with less than 60% amino acid sequence similarity or that had a more than 20% difference in protein length were discarded. To keep only orthologous genes in the final core genomes and delete all paralogs and xenologs, genes outside blocks of synteny were removed. To do this, a gene was validated as part of the core genome only if, of the five genes downstream and the five genes upstream of it, at least four were at the same location in all the genomes. The core genomes were defined as the intersection of pairwise lists of strict positional orthologs.

Distribution of virulence genes and detection of size variations in virulence gene products. The genome of the EGDe[25] strain was used as the reference for the detection of size variability in the virulence gene products encoded by all the other genomes. Homologs of all the genes in the reference genome were first searched for using nucleotide BLAST against all the other genomes. On the basis of the distribution of all the best *e* values obtained for all BLAST analyses, a gene was considered to be absent in the newly sequenced genome if all the *e* values were greater than 1×10^{-50} and all matches were located in syntenic blocks (delimited by the core genes) distinct from the one for the gene in the EGDe reference. Only hits with *e* values smaller than 1×10^{-50} or located in the same syntenic block as the reference gene were considered to be present and were further analyzed for comparison with EGDe. For this analysis, detected homologous regions were extracted and translated into amino acid sequences. We then compared the size of the reference gene products in EGDe with the size of the translated matches from each genome. A gene product was considered to be shorter than in the reference if it was shorter by at least 20 amino acids. Gene products that were smaller than in EGDe and were encoded by genes located at the ends of contigs were considered to be present, as it was not possible to distinguish biological events from methodological artifacts. Virulence genes of the LIPI-1, LIPI-3 and SSI-1 (stress survival islet 1) islands as well as *inlA* and *inlB* were analyzed in more detail to detect the origins of size variations in gene products. To this end, alignments of the nucleotide and amino acid sequences of all detected matches were performed to detect nonsense mutations and internal deletions in coding sequences.

Phylogenetic analyses based on the core genomes. See the **Supplementary Note**.

Pan-genome definition. The homologous gene families (including paralogs, orthologs and xenologs) of the pan-genome were defined as previously described in Touchon *et al.*[43] except that the Silix parameters were set such that a protein was considered to be a homolog of another protein if the aligned portion had at least 60% similarity and represented more than 80% of the smallest protein. Chromosomal as well as plasmid genes were included in the pan-genome.

Gene presence/absence patterns among clones based on the pan-genome. The exhaustive collection of isolates that was used for the source distribution analysis allowed us to define a frequency of clinical isolates for each clone. We thus needed the presence/absence patterns of genes of the pan-genome by clone. The presence or absence of gene families within a given clone was defined by the majority consensus (>50%) for all genomes of each clone.

Phylogenetic inertia of the clinical frequency of clones and use of generalized estimating equations to identify candidate virulence genes. Some of the genomes were very closely related, whereas others were very distant. We therefore checked for phylogenetic inertia. Measurement of the phylogenetic inertia of the clinical frequency of clones was performed both on the phylogeny of the fused lineages and of separated lineages using the phylosig tool with the lambda method, computed in R and implemented in the phytools package using the lambda method[44]. Phylogenetic inertia was high for the complete data set (lineages I and II, Pagel's $\lambda = 0.9999$, $P = 0.0005$).

To identify the gene families that were most associated with clones frequently involved in clinical infections, comparative analysis of the presence/absence patterns of gene families among clones (according to the $\log_{10}$ values of their clinical frequencies) was carried out taking into account their phylogenetic relationships. This analysis was performed using generalized estimating equations (GEE) computed in R and implemented in the ape package[29]. The estimates of the regression parameters from GEE were calculated for each gene family, reflecting association with clones of high clinical frequency. To identify the gene families that were most associated with clones frequently involved in clinical infections, comparative analysis of the presence/absence patterns of gene families among clones was carried out taking into account the $\log_{10}$ values of their clinical frequencies and their phylogenetic relationships.

Construction of the *Listeria* gene deletion mutant. The *L. monocytogenes* LM09-00558 ΔPTS mutant was obtained by deletion of the entire putative PTS cluster by PCR ligation and amplicon cloning in the suicide vector pMAD, as

previously described[45]. Briefly, primer pair 1 (**Supplementary Table 10**) was used to amplify the 5′ flanking region of the PTS cluster in the *L. monocytogenes* LM09-00558 genome, followed by PCR ligation with the 3′ flanking region amplified with primer pair 2 (**Supplementary Table 10**). The amplified DNA fragments were subsequently cloned into the SalI-BglII site in the pMAD suicide vector. The clones obtained were isolated and sequenced to verify sequence integrity. The plasmids with correct sequence were electroporated into electrocompetent cells prepared with the *L. monocytogenes* LM09-00558 strain on the basis of the method for *L. monocytogenes* 4b strains[46], followed by clone selection as described[45]. The clones obtained were isolated, and the PTS-flanking region was sequenced to verify the deletion.

PTS cloning. The entire PTS gene cluster of *L. monocytogenes* LM09-00558 was amplified with primer pair 3 (**Supplementary Table 10**) followed by blunt-end cloning into pCR-Blunt (Invitrogen). The clones obtained were isolated and sequenced to verify sequence integrity. The insert was digested, purified and cloned into the SalI-NotI site of the pIMC plasmid, which integrates into the *L. monocytogenes* $tRNA^{Arg}$ site in a single copy following conjugation as mediated by the listeriophage PSA integrase encoded by the plasmid[47].

35. Cantinelli, T. *et al.* "Epidemic clones" of *Listeria monocytogenes* are widespread and ancient clonal groups. *J. Clin. Microbiol.* **51**, 3770–3779 (2013).
36. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70 (1979).
37. Bonferroni, C.E. Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* **8**, 3–62 (1936).
38. Disson, O. *et al.* Modeling human listeriosis in natural and genetically engineered animals. *Nat. Protoc.* **4**, 799–810 (2009).
39. Criscuolo, A. & Brisse, S. AlienTrimmer: a tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics* **102**, 500–506 (2013).
40. Rissman, A.I. *et al.* Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics* **25**, 2071–2073 (2009).
41. Vallenet, D. *et al.* MicroScope: a platform for microbial genome annotation and comparative genomics. *Database (Oxford)* **2009**, bap021 (2009).
42. Touchon, M. *et al.* Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* **5**, e1000344 (2009).
43. Touchon, M. *et al.* The genomic diversification of the whole *Acinetobacter* genus: origins, mechanisms, and consequences. *Genome Biol. Evol.* **6**, 2866–2882 (2014).
44. Revell, L.J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
45. Arnaud, M., Chastanet, A. & Débarbouillé, M. New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, gram-positive bacteria. *Appl. Environ. Microbiol.* **70**, 6887–6891 (2004).
46. Monk, I.R., Gahan, C.G. & Hill, C. Tools for functional postgenomic analysis of *listeria monocytogenes*. *Appl. Environ. Microbiol.* **74**, 3921–3934 (2008).
47. Monk, I.R., Casey, P.G., Cronin, M., Gahan, C.G. & Hill, C. Development of multiple strain competitive index assays for *Listeria monocytogenes* using pIMC; a new site-specific integrative vector. *BMC Microbiol.* **8**, 96 (2008).