Genetic Organization of ascB-dapE Internalin Cluster Serves as a Potential Marker for *Listeria monocytogenes* Sublineages IIA, IIB, and IIC

Chen, Jianshun1,2,3, Chun Fang1, Ningyu Zhu1,2, Yonghui Lv4, Changyong Cheng1, Yijiang Bei1,2,3, Tianlun Zheng1,2,3, and Weihuan Fang1,*

1Zhejiang University Institute of Preventive Veterinary Medicine, and Zhejiang Provincial Key Laboratory of Preventive Veterinary Medicine, 388 Yuhangtang Road, Hangzhou, Zhejiang 310058, P.R. China
2Zhejiang Fisheries Technical Extension Center, 18 Maizidian Road, Beijing 100125, P.R. China
3Zhejiang Aquatic Disease Prevention and Quarantine Center, 20 Yile Road, Hangzhou, Zhejiang 310012, P.R. China
4National Fisheries Technical Extension Center, 18 Maizidian Road, Beijing 100125, P.R. China

Received: October 17, 2011 / Revised: December 16, 2011 / Accepted: January 8, 2012

*L. monocytogenes* is an important foodborne pathogen that comprises four genetic lineages: I, II, III, and IV. Of these, lineage II is frequently recovered from foods and environments and responsible for the increasing incidence of human listeriosis. In this study, the phylogenetic structure of lineage II was determined through sequencing analysis of the ascB-dapE internalin cluster. Fifteen sequence types proposed by multilocus sequence typing based on nine housekeeping genes were grouped into three distinct sublineages, IIA, IIB, and IIC. Organization of the ascB-dapE internalin cluster could serve as a molecular marker for these sublineages, with inlGHE, inlGC2DE, and inlIC2DE for IIA, IIB, and IIC, respectively. These sublineages displayed specific genetic and phenotypic characteristics. IIA and IIC showed a higher frequency of recombination (ρ/θ). However, recombination events had greater effect (r/m) on IIB, leading to its high nucleotide diversity. Moreover, IIA and IIB harbored a wider range of internalin and stress-response genes, and possessed higher nisin tolerance, whereas IIC contained the largest portion of low-virulent strains owing to premature stop codons in inlA. The results of this study indicate that IIA, IIB, and IIC might occupy different ecological niches, and IIB might have a better adaptation to a broad range of environmental niches.

**Keywords:** *L. monocytogenes*, lineage II, sublineage, ascB-dapE internalin cluster

*Corresponding author*
Phone: +86 571 8898 2242; Fax: +86 571 8898 2242; E-mail: whfang@zju.edu.cn

# Supplementary data for this paper are available on-line only at http://jmb.or.kr.

The genus *Listeria* comprises eight species: *Listeria monocytogenes*, *L. innocua*, *L. welshimeri*, *L. ivanovii*, *L. seeligeri*, *L. grayi*, *L. marthii*, and *L. rocourtiae* [39]. Only *L. monocytogenes* is considered as a foodborne pathogen of both humans and animals, causing two forms of listeriosis: non-invasive gastrointestinal listeriosis and invasive listeriosis [2]. One salient feature of the population structure of *L. monocytogenes* is the distinction of the four evolutionary lineages I, II, III, and IV with different but overlapping ecological niches [39, 52]. Lineages I and II account for at least 95% of strains from foods and patients, whereas lineages III and IV are rare and seldomly associated with human listeriosis [46]. Furthermore, lineages I and II reflect the distribution of four major serovars, with serovars 1/2b and 4b belonging to lineage I and serovars 1/2a and 1/2c to lineage II [42, 54].

Lineage I (particularly serovar 4b) strains were previously overrepresented when compared with other lineages among strains responsible for listeriosis outbreaks, whereas lineage II strains were usually more frequently recovered from foods and natural environments [9, 25, 29, 46]. Since 2000, the incidence of human listeriosis caused by lineage II strains, particularly those of serovar 1/2a, has increased in many countries. Some examples were the outbreaks occurred with Tomme cheese in Switzerland [7], sandwich in the United Kingdom [20], ready-to-eat meat in Canada [24], and a multinational outbreak due to the consumption of Quargel cheese in Australia and Germany [23]. Characterization of 601 human *L. monocytogenes* isolates during the period 1986–2007 in Sweden reveals that serovar 1/2a has become the predominant serovar (71%) causing invasive listeriosis since 2000 [40]. This is also the case from the retrospective study based on 722 *L. monocytogenes* isolates from Canadian cases and outbreaks of listeriosis [17].
From the evolutionary perspective, lineages I and II constitute distinct species-like lineages, as they correspond to clearly demarcated sequence clusters that fulfill the separateness criteria and divergence levels used in other bacterial groups to distinguish species [4]. Exchange of genetic material is rarely observed between lineages I and II [4, 38]. Recombination, an important evolutionary force that causes higher genetic variability, is more prevalent within lineage II than lineage I [4, 13]. In contrast to lineage I being clonal, lineage II represents a population of high genetic heterogeneity, which is revealed by previous studies [39, 42]. However, detailed knowledge on further subdivision of the lineage II strains is still lacking.

DNA sequence-based subtyping methods, including multilocus sequence typing (MLST), have gained more popularity owing to their reproducibility and discriminatory power [32, 49]. The MLST scheme based on nine housekeeping genes (including gyrB, dapE, hisJ, sigB, ribC, purM, betL, gap, and tuf), proposed by our previous studies, revealed insights on the phylogenetic relationship of the *L. monocytogenes*–*L. innocua* clade [13, 14]. In addition, virulence-associated genes generally evolve rapidly owing to strong selection pressures, and usually serve as sources of clues for discriminating subpopulation structure [56]. Whereas the *Listeria* pathogenicity island I (LIP-I) seems to be conserved in all *L. monocytogenes* strains, the presence of premature stop codons in some virulence-associated genes (e.g., *inlA*) and the absence of other accessory virulence-associated genes (e.g., internalins and stress-response genes) have been found in specific lineages or strains [10, 11, 15, 22, 33].

The purposes of this study were (i) to determine the population structure of *L. monocytogenes* lineage II, and (ii) to characterize the genetic and phenotypic features between distinct subpopulations of *L. monocytogenes* lineage II, through examination of the genetic organization of the *ascB-dapE* internalin cluster.

**Materials and Methods**

**Bacterial Strains and Growth Conditions**

A total of 144 *L. monocytogenes* strains were examined in this study (Table S1). These were subdivided into serovars 1/2a (n = 54), 1/2b (n = 29), 1/2c (n = 18), 4a (n = 10), 4b (n = 31), and 4c (n = 2) by classical serotyping, and lineages I (n = 58), II (n = 72), and III/IV (n = 14) by sequencing and phylogenetic analysis of actA [54] that correlated with serotyping. *Listeria* strains were retrieved from glycerol stocks maintained at −80°C, and cultured in brain heart infusion broth (BHI; Oxoid, England) at 37°C.

**DNA Manipulations**

Genomic DNA was extracted using a previously optimized protocol [12]. Oligonucleotide primers (Tables S2 and S3) were synthesized by Invitrogen Biotechnology (Shanghai, China), and *Taq* DNA polymerase (TaKaRa, China) was used for regular PCR reactions. For products larger than 2 kb, *LA Taq* DNA polymerase (TaKaRa) was employed. The PCR reaction was conducted using the PT-200 thermal cycler (MJ, USA). PCR fragments were purified using the AxyPrep DNA Gel Extraction Kit (Axygen, USA) and ligated into pMD18-T (TaKaRa). The recombinant plasmids were sequenced by the dideoxy method on an ABI-PRISM 377 DNA sequencer.

**Determination of Organization of the *ascB-dapE* Internalin Cluster**

Three sets of PCRs were conducted based on two upstream primers (u1, targeting *inlA*; u2, targeting *inlC2*) and two downstream primers (d1, targeting *inlE*; d2, targeting *inlD*) (Table S2). PCR using primer pair u1/d1 was expected to produce a 4,000 bp fragment from strains harboring *inlGC2DE*, a 2,241 bp fragment from strains harboring the *inlGHE* cluster, and no fragment from those harboring *inlC2DE* or being empty between *ascB* and *dapE*, whereas PCR using primer pair u1/d2 only yields a 2,241 bp fragment from strains harboring *inlGC2DE* [11]. For strains being consistently negative in the above two PCR sets, another PCR using primer pair u2/d2 was performed, which only obtained a 1,782 bp fragment from strains bearing *inlC2DE*. Owing to the conserved repeats present in the internalin multigene family, primers were designed based on the distinguishable regions through sequence comparison.

**Multilocus Sequence Typing (MLST) and Data Analysis**

The MLST scheme was based on the sequence analysis of nine housekeeping genes, including gyrB, dapE, hisJ, sigB, ribC, purM, betL, gap, and tuf (Table S2). The sequences of 29 lineage I strains, 12 lineage II strains, 8 lineage III strains, and 5 lineage IV strains were generated in our previous studies [14], and those of 20 additional lineage II strains were obtained in this study. In addition, the sequences of serovar 1/2a strains 08-5578 and 08-5923 (responsible for the large RTE-meat outbreak in Canada in 2008), and serovar 4b strains F2365 (responsible for the Jalisco cheese outbreak in California in 1985), and CLIP50459 (responsible for the bacteremia cases in France in 1999) were retrieved from GenBank (accession numbers CP001602, CP001604, AE017262, and FM242771, respectively).

For each MLST locus, an allele number was given to each distinct sequence variant, and a distinct sequence type (ST) number was given to each distinct combination of alleles of the 9 genes. MEGA 4.0 [48] was used to construct a neighbor-joining tree of the concatenated sequences of 9 loci with 1,000 bootstrap tests. *L. innocua* was employed as an outgroup species. DNAsc v.5.10.01 [44] was used to calculate the number of alleles, number of polymorphic sites, nucleotide diversity indices (π, mean pairwise nucleotide difference per site), and Tajima’s D [47]. ClonalFrame v.1.1 [21] was used to show the evolution of ρ/θ and r/m as chain run as well as the time to the most recent common ancestor (TMRCA). These two complementary measures were used to assess the relative contribution of recombination and mutation in the creation of the sample population from a common ancestor. Specifically, ρ/θ is the ratio of rates at which recombination and mutation occur, representing a measure of how often recombination happens relative to mutation [34], whereas r/m is the ratio of probabilities that a given site is altered through recombination and mutation, representing a measure of how important the effect of
recombination is in the diversification of the sample population relative to mutation [26].

Detection of Virulence-Associated Genes
Four categories of virulence-associated genes were examined in 144 L. monocytogenes strains using primers listed in Table S3, including (i) stress-response genes conferring tolerance to adverse conditions (e.g., bsh, glutamate decarboxylase system, arginine deiminase system, and agmatine deiminase system); (ii) internalin genes responsible for adherence and invasion of host cells (e.g., inla, inlB, inlC, inlF, inlJ, and lmo2026); (iii) genes involved in escape from vacuoles, multiplication in cytoplasm, and intracellular and intercellular spread (e.g., plcA, lhy, mpl, actA, plcB, and hpt); and (iv) regulatory genes (e.g., prfA). Furthermore, the presence of premature stop codons (PMSC) in inla was assessed in 72 L. monocytogenes lineage II strains through sequencing of the whole length of inla by three sets of primers (Table S3).

Growth Experiments
Stationary-phase bacterial cultures (33 randomly selected strains in total with 11 per lineage II subpopulation) were washed in PBS (0.01 M, pH 7.2), and inoculated (2%) into fresh BHI adjusted to pH 4.8 with lactic acid or supplemented with 200 µM nisin (Sigma, N5764), respectively. Two hundred microliters was transferred into individual wells of a 96-well plate, and cell growth was measured spectrophotometrically (SpectraMax M2, Molecular Device, USA) at 37°C for 10 h. All experiments were performed in triplicate from three separate cultures and repeated three times.

Cell Culture Invasion Assay
The ability of 20 L. monocytogenes isolates (representing three lineage II subpopulations) (Table 4) to invade HeLa epithelial cells was examined. Cell monolayers at 80% confluence in DMEM were inoculated with bacterial suspension (10^7 CFU/ml) to obtain a multiplicity of infection (MOI) of 1:10 for 1 h at 37°C in the presence of 5% CO2. The cell monolayers were washed with PBS (0.01 M, pH 7.2) to remove non-adherent bacteria, and subjected to gentamicin (100 µg/ml) inactivation of extracellular bacteria. The cell monolayers were then lysed with cold deionized water. The CFU values of viable bacteria were determined by plating suitable dilutions of the lysates onto BHI agar. These experiments were repeated twice for each strain. The LD50 values were calculated by using the trimmed Spearman–Karber method on the basis of mouse mortality data recorded during a 10-day post-injection period.

Statistical Analysis
The two-tailed Student’s t-test and χ²-test were used for data analysis, where necessary, and P values ≤0.05 were considered as statistically significant.

RESULTS AND DISCUSSION
L. monocytogenes Lineage II Strains Display Three Patterns of ascB-dapE Internalin Cluster
Five distinct organizations of the internalin cluster between ascB and dapE were identified in 144 L. monocytogenes strains of all four lineages. The lineage I strains carried the inlC2DE cluster, except for one strain (NB9, serovar 1/2b) bearing inlGC2DE; all lineage IV strains harbored no internalin; and lineage III strains contained either inlGC2DE (4/9), inlC2 (2/9), or no internalin (3/9) (Table 1, Table S1). The inlC2 pattern specific to lineage III represents the first to be reported to date. Lineage II strains exhibited diverse internalin profiles in this locus, including inlC2DE (17/72), inlGC2DE (38/72), as well as lineage II-specific structure inlGHE (17/72) (Table 1, Fig. 1, Table S1). Organization

Table 1. Organization of the ascB-dapE internalin cluster in L. monocytogenes lineages and serovars.

<table>
<thead>
<tr>
<th>Lineage</th>
<th>Serovar</th>
<th>No. (%) harboring inlC2DE</th>
<th>No. (%) harboring inlGC2DE</th>
<th>No. (%) harboring inlGHE</th>
<th>No. (%) harboring inlC2</th>
<th>No. (%) harboring nothing</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1/2b</td>
<td>28/29 (96.6%)</td>
<td>1/29 (3.4%)</td>
<td>0/29 (0%)</td>
<td>0/29 (0%)</td>
<td>0/29 (0%)</td>
</tr>
<tr>
<td></td>
<td>4b</td>
<td>29/29 (100%)</td>
<td>0/29 (0%)</td>
<td>0/29 (0%)</td>
<td>0/29 (0%)</td>
<td>0/29 (0%)</td>
</tr>
<tr>
<td></td>
<td>Subtotal</td>
<td>57/58 (98.3%)</td>
<td>1/58 (1.7%)</td>
<td>0/58 (0%)</td>
<td>0/58 (0%)</td>
<td>0/58 (0%)</td>
</tr>
<tr>
<td>II</td>
<td>1/2a</td>
<td>17/54 (31.5%)</td>
<td>35/54 (64.8%)</td>
<td>2/54 (3.7%)</td>
<td>0/54 (0%)</td>
<td>0/54 (0%)</td>
</tr>
<tr>
<td></td>
<td>1/2c</td>
<td>0/18 (0%)</td>
<td>3/18 (16.7%)</td>
<td>15/18 (83.3%)</td>
<td>0/18 (0%)</td>
<td>0/18 (0%)</td>
</tr>
<tr>
<td></td>
<td>Subtotal</td>
<td>17/72 (23.6%)</td>
<td>38/72 (52.8%)</td>
<td>17/72 (23.6%)</td>
<td>0/72 (0%)</td>
<td>0/72 (0%)</td>
</tr>
<tr>
<td>III</td>
<td>4a, 4b or 4c</td>
<td>0/9 (0%)</td>
<td>4/9 (44.5%)</td>
<td>0/9 (0%)</td>
<td>2/9 (22.2%)</td>
<td>3/9 (33.3%)</td>
</tr>
<tr>
<td>IV</td>
<td>4a or 4b</td>
<td>0/5 (0%)</td>
<td>0/5 (0%)</td>
<td>0/5 (0%)</td>
<td>0/5 (0%)</td>
<td>5/5 (100%)</td>
</tr>
<tr>
<td>Subtotal</td>
<td>74/144 (51.4%)</td>
<td>43/144 (29.9%)</td>
<td>17/144 (11.8%)</td>
<td>2/144 (1.4%)</td>
<td>8/144 (5.5%)</td>
<td></td>
</tr>
</tbody>
</table>

*The genetic organization of 58 lineage I and 72 lineage II strains have been reported in our previous publication [16].
of this internalin cluster separated lineage II strains into three potential subpopulations independent of serovars, with \(\text{inlGC2DE} \ (35/54)\) and \(\text{inlGHE} \ (15/18)\) serving as the dominant version of serovars 1/2a and 1/2c, respectively (Table 1).

**Internalin Profiling is Consistent with MLST in Delineating Three Sublineages of \(L.\ monocytogenes\) Lineage II**

In the MLST scheme, there were a total of 791 polymorphic sites (13.54%) in sequences of nine genes in 78 \(L.\ monocytogenes\) strains, with nucleotide diversity \(\pi\) at 0.03934 (Table 2). The 47 resulting unique sequence types (ST) were clustered into four lineages: I, II, III, and IV (Fig. 2). Furthermore, lineage II comprised three subpopulations corresponding to distinct \(\text{ascB-dapE}\) internalin cluster patterns, \(\text{inlGHE}, \text{inlGC2DE},\) and \(\text{inlC2DE}\) (Fig. 2). Notably, seven \(\text{inlGC2DE}\)-containing strains (S3, NB21, S17, SH2, 10403S, SH4, and NB12) appeared to branch off owing to significant mutations in individual genes. Strains S3 and NB21 differed from other \(\text{inlGC2DE}\)-containing strains by their distinct \(\text{ribC}\) gene sequences (Fig. S1A), whereas strains S17, SH2, 10403S, SH4, and NB12 differed by their distinct \(\text{purM}\) gene sequences (Fig. S1B). When we constructed phylogenetic trees excluding \(\text{ribC}\) and \(\text{purM}\), these seven strains consistently fell into the same branch with other \(\text{inlGC2DE}\)-containing strains (Fig. 2 inset).

When sequence data were analyzed after stratification by lineages, the genetic diversity (i.e., number of polymorphic sites and \(\pi\) value) observed within each lineage was significantly lower (Table 2). Within two major lineages, lineage I was less divergent (containing 105 polymorphic sites, 1.80%; \(\pi = 0.00427\)), whereas lineage II was genetically heterogeneous (containing 208 polymorphic sites, 3.56%; \(\pi = 0.00753\)) (Table 2). The number of polymorphisms within each lineage II subpopulation was considerably reduced. The \(\text{inlGC2DE}\)-containing subpopulation appeared to be the most genetically diverse within lineage II, as it harbored the majority of polymorphisms and exhibited significantly higher nucleotide diversity (\(\pi = 0.01157\)) than other subpopulations (\(\pi = 0.00024\) for \(\text{inlGHE}\)-containing strains).

**Table 2. Descriptive analysis of nucleotide sequences of nine genes for \(L.\ monocytogenes\) strains.**

<table>
<thead>
<tr>
<th>Gene</th>
<th>No. strains</th>
<th>Size (bp)</th>
<th>No. alleles</th>
<th>No. (%) polymorphic sites</th>
<th>(\pi)</th>
<th>Tajima’s D</th>
<th>(\rho/\theta)</th>
<th>(r/m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lineage I</td>
<td>31</td>
<td>5,844</td>
<td>20</td>
<td>105 (1.80%)</td>
<td>0.00427</td>
<td>-0.22644</td>
<td>0.043 (0.030–0.056)</td>
<td>0.446 (0.359–0.539)</td>
</tr>
<tr>
<td>Lineage II</td>
<td>34</td>
<td>5,844</td>
<td>15</td>
<td>208 (3.56%)</td>
<td>0.00753</td>
<td>0.65087</td>
<td>0.134 (0.089–0.169)</td>
<td>2.423 (1.860–2.838)</td>
</tr>
<tr>
<td>Sublineage A</td>
<td>10</td>
<td>5,844</td>
<td>3</td>
<td>7 (0.12%)</td>
<td>0.00024</td>
<td>-1.83913*</td>
<td>0.698 (0.319–1.034)</td>
<td>1.240 (0.975–1.548)</td>
</tr>
<tr>
<td>Sublineage B</td>
<td>13</td>
<td>5,844</td>
<td>6</td>
<td>192 (3.29%)</td>
<td>0.01157</td>
<td>0.20340</td>
<td>0.200 (0.079–0.297)</td>
<td>4.324 (2.298–6.335)</td>
</tr>
<tr>
<td>Sublineage C</td>
<td>11</td>
<td>5,844</td>
<td>6</td>
<td>18 (0.31%)</td>
<td>0.00120</td>
<td>0.63866</td>
<td>0.561 (0.281–0.942)</td>
<td>1.237 (0.688–1.989)</td>
</tr>
<tr>
<td>Lineage III</td>
<td>8</td>
<td>5,844</td>
<td>7</td>
<td>210 (3.59%)</td>
<td>0.01330</td>
<td>-0.41214</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Lineage IV</td>
<td>5</td>
<td>5,844</td>
<td>5</td>
<td>129 (2.21%)</td>
<td>0.00987</td>
<td>-0.57336</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Total</td>
<td>78</td>
<td>5,844</td>
<td>47</td>
<td>791 (13.54%)</td>
<td>0.03934</td>
<td>1.08899</td>
<td>0.229 (0.185–0.269)</td>
<td>1.792 (1.551–2.018)</td>
</tr>
</tbody>
</table>

\(\pi\): Nucleotide diversity; \(\rho/\theta\): the ratio of rates at which recombination and mutation occur represents a measure of how often recombination events happen relative to mutations; \(r/m\): the ratio of probabilities that a given site is altered through recombination and mutation represents a measure of how important the effect of recombination is in the diversification of the sample relative to mutation. *, \(p<0.05\). ND, not determined.
subpopulation and $\pi = 0.00120$ for \textit{inlC2DE}-containing subpopulation) (Table 2). In evolutionary terms, a younger bacterial species has a lower level of genetic diversity [49]. The results from this study suggested that \textit{inlGHE}- and \textit{inlC2DE}-containing subpopulations were younger than the \textit{inlGC2DE}-containing subpopulation. This was further confirmed by the determination of the estimated time to the most recent common ancestors (TMRCA) via ClonalFrame. The phylogram based on the analysis with correction for recombination revealed that TMRCA of \textit{inlGC2DE}-containing subpopulation was much shorter than the other two subpopulations (Fig. S2), suggesting that this subpopulation might represent the ancestral subpopulation among three sublineages.

Tajima’s D test not only tests the hypothesis that sequences have evolved according to the neutral theory but
also reveals the demographics of a given population and thus can be used to make inferences on an organism’s population structure [37, 45]. According to the findings by Bakker et al. [4] and this study, *L. monocytogenes* did not evolve under neutrality, as large positive Tajima’s D values became smaller or negative when analysis was performed for each lineage separately, indicating a subdivided population structure for *L. monocytogenes*. Similarly, the large positive Tajima’s D values for all lineage II strains, which were reduced upon stratification by the subpopulation harboring a distinct *ascB-dapE* internalin cluster pattern (Table 2), provided additional evidence that lineage II represented a subdivided population.

In addition, the lineage I strain (NB9) carrying *inlGC2DE* was placed outside the main cluster of lineage I strains bearing *inlC2DE* (Fig. 2). Overall, the organization of the *ascB-dapE* internalin cluster is consistent with MLST. This internalin profiling stands as a potential molecular marker for separation of lineage II into three distinct sublineages, designated as IIA, IIB and IIC, corresponding to subpopulations carrying *inlGHE*, *inlGC2DE*, and *inlC2DE*, respectively.

**L. monocytogenes** Lineage II Shows Sublineage-Specific Recombination Rates

The recombination rate in bacteria can differ widely from one species to another [41]. This study further revealed that the contribution of recombination to genotypic diversity varied with lineages, as shown by the relative frequency of occurrence of recombination versus mutation (r/θ) and the relative effect of recombination versus point mutation (r/m) (Table 2). Lineage II exhibited a considerably higher frequency (r/θ = 0.134) and effect (r/m = 2.423) of recombination than lineage I (r/θ = 0.043; r/m = 0.446) (Table 3).

More remarkably, sublineages IIA, IIB, and IIC also showed different recombination rates. IIA and IIC exhibited a higher frequency of recombination (r/θ = 0.698 for IIA; r/θ = 0.561 for IIC) than IIB (r/θ = 0.200) (Table 3). Based on the concept that an increased selective pressure for rapid diversification in response to various environments results in higher recombination frequency [4, 55], IIA and IIC might have been faced with increased selective pressures posed by environments. However, the higher recombination frequency had not made significant contributions to the genetic diversity of IIA and IIC (Table 2). On the contrary, the recombination events had a greater effect on IIB (r/m = 4.324) than IIA (r/m = 1.240) and IIC (r/m = 1.237). The greater recombination effect of IIB seemed to contribute to its higher nucleotide diversity (Table 2), which might lead to a better adaptability to a broad range of environments.

**Lineage II Subpopulations Harbor Different Compositions of Internalins and Glutamate Decarboxylases**

Among four categories of virulence-associated genes examined (Table S3), the majority of genes were present in all lineage II strains, apart from two internalin genes, *inlf* and *lmo2026*, and the glutamate decarboxylase (GAD) gene family (Table S1).

The multigene internalin family is scattered in *L. monocytogenes* genomes and seems to play broad roles not merely limited to invasion of host cells [6, 35]. InlF is specific to lineage II [11], and mediates increased cell binding and entry when the RhoA/ROCK pathway is inhibited [30]. *inlf* existed in all IIA and IIB strains as well as in a small amount of IIC strains (4/17, 23.5%) (Table 3; Table S1). *lmo2026* is another lineage-II-specific internalin based on our previous report [11], and is possibly involved in listerial multiplication in the brain [3]. *lmo2026* existed in all IIA strains, 34.2% (13/38) of IIB strains, but none of IIC strains (Table 3; Table S1).

The GAD system contributes to the ability of *L. monocytogenes* to tolerate acidic conditions, such as in low pH foods, during gastric transit, exposure to fatty acids in the intestine, and in the phagosome of macrophages during systemic infection [18]. The GAD system comprises three homologs (*gadD1*, *gadD2*, and *gadD3*) located in three distinct loci [18, 19]. Whereas *gadD2* and *gadD3* were identified in all lineage II strains, *gadD1* was specifically present in IIA and IIB strains but absent in all IIC strains (Table 3). In addition, all lineage I strains containing *inlC2DE*-containing lineage I strain (NB9) possessing this gene (Table S1). Consistent with the results in our previous study, the genomic presence of the GAD system correlates with the organization of the *ascB-dapE* internalin cluster in *Listeria* [16]. These results not only demonstrate that lineage II subpopulations exhibited distinct genetic features, but also suggest that some internalin and stress-response genes, which play broad roles in enhancing the adaption to

<table>
<thead>
<tr>
<th>Sublineage</th>
<th>No.</th>
<th>inlA with PMSC*</th>
<th>inlf</th>
<th>lmo2026</th>
<th>gadD1</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIA</td>
<td>17</td>
<td>2 (11.8%)</td>
<td>17 (100%)</td>
<td>17 (100%)</td>
<td>17 (100%)</td>
</tr>
<tr>
<td>IIB</td>
<td>38</td>
<td>0 (0%)</td>
<td>38 (100%)</td>
<td>13 (34.2%)</td>
<td>38 (100%)</td>
</tr>
<tr>
<td>IIC</td>
<td>17</td>
<td>4 (23.5%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Total</td>
<td>72</td>
<td>5 (6.9%)</td>
<td>59 (81.9%)</td>
<td>30 (41.7%)</td>
<td>55 (76.4%)</td>
</tr>
</tbody>
</table>

*PMSC, premature stop codon.*
various environments, might co-evolve with the diversification of sublineages.

Sublineages IIA and IIB Exhibit Higher Nisin Tolerance Than IIC
Since lineage II strains are more frequently recovered from foods, and low pH and nisin (the most extensively used bacteriocin) are often used as part of the “hurdle concept” in the preservation of minimally processed foods [31, 43], we assayed the growth rates of 33 strains representing three sublineages in acidic and nisin-supplemented conditions. IIA, IIB, and IIC strains exhibited similar growth rates on average under standard laboratory condition (BHI, pH 7.0) (Fig. 3) and sublethal acidic condition (BHI, pH 4.8) (data not shown). However, IIA and IIB strains reproducibly grew better on average in the presence of 200 µg/ml nisin than sublineage IIC, with a significantly higher growth maximum (p<0.05 at 7–10 h) (Fig. 3), suggesting IIA and IIB were more tolerant to nisin.

Intracellular ATP levels have been shown to be important for L. monocytogenes in surviving exposure to nisin [8]. A link between amino acid decarboxylation and ATP biosynthesis has been uncovered in bacteria [1, 27, 36]. We thus hypothesized that the GAD system might be implicated in the nisin tolerance. A recent report supported this hypothesis that GadD1 contributed significantly to ATP pools within the cell and subsequently to the nisin tolerance [5]. Given this link, it is tempting to speculate that IIA and IIB containing GadD1 might have an advantage in environments where bacteriocin-producing organisms are abundant.

InlA Premature Stop Codons are Identified in Sublineages IIA and IIC, Leading to Reduced Invasion Efficiency and In Vivo Virulence
InlA is critical for the entry of L. monocytogenes into various non-phagocytic human cells expressing its receptor E-cadherin [6]. Although inlA is present in all lineage II strains, a considerable proportion (>30%) of lineage II strains are reported to harbor premature stop codons (PMSC) in inlA [28, 38, 51, 53]. At least 18 different polymorphisms in inlA leading to PMSC have been observed [39, 50, 51]. In this study, 8.3% (6/72) of lineage II strains contained PMSC mutations in inlA, including 11.8% (2/17) of IIA strains, 23.5% (4/17) of IIC strains, and none (0/38) of IIB strains (Table 3; Table S1). The two IIA strains (C4, P20) had PMSC at nucleotide position 1380 (G→A), whereas the four IIC strains (C2, C6, P7, B4) had PMSC at position 1474 (C→T) (Table 4). Both the frequency and type of inlA mutations were significantly lower than previous results [28, 38, 50, 51, 53]. This is possibly due to the sampling bias, as the majority of L. monocytogenes strains used in this study were isolated from China, whereas those for previous surveys were mostly obtained from American and European countries [42, 50, 51, 53].

The six strains (C4, P20, C2, C6, P7, B4) bearing PMSC in inlA demonstrated significantly attenuated invasion efficiencies (P<0.01) compared with those of other lineage II strains (Table 4). Consequently, the altered InlA impaired virulence of these strains (LD_{50} > 10^{7.5}) in mice relative to other lineage II strains harboring intact InlA (LD_{50} < 10^{6.5}) (Table 4). These results suggested that low-virulent strains due to PMSC mutations in inlA were more common in IIA and IIC (6/34, 17.6%) than IIB (0/38, 0%) (χ^2 = 5.19, p<0.05). Interestingly, almost all the serovar 1/2a strains responsible for documented large listeriosis outbreaks, whose genomic sequences are known, belong to IIB (data not shown).

L. monocytogenes is an opportunistic pathogen as well as a saprotroph, ubiquitous in natural environments such as soil, silage, groundwater, sewage, and vegetation [39]. Removal or inactivation of some virulence factors (e.g., InlA and InlF) in a specific subpopulation could be regarded as adaptive gene loss, which favors its survival in particular environmental niches [12, 22, 42]. Little is known about the ecological mechanisms that drive the evolution of these apparently attenuated strains, but a realistic scenario is that distinct sublineages might be adapted to different ecological niches, and their occurrence as mammalian pathogens may be of limited significance for their evolutionary success in the long term. Determining the natural habitat of distinct sublineages may provide clues to understand why the expression of virulence traits may in fact turn out to be disadvantageous in particular environments.

In conclusion, L. monocytogenes lineage II is a genetically diverse population, encompassing three distinct sublineages, IIA, IIB, and IIC. These sublineages display specific genetic and phenotypic characteristics, and might occupy different ecological niches. IIB appears to be the most genetically diverse subpopulation within lineage II, and contains a relatively complete set of internalin and stress-
response genes, suggesting this sublineage might have a better adaption to a broad range of environmental niches. The organization of the \textit{ascB-dapE} internalin cluster serves as a molecular marker for these sublineages.

\section*{Acknowledgments}
This study was supported by the National Natural Science Foundation of China (31101829) and the Key Project of National Science and Technology Pillar Program (2009BADB9B01). J. C. is supported by the China Postdoctoral Science Foundation (20100481428). We thank Dr. Dongyou Liu at the Royal College of Pathologists of Australasia Quality Assurance Programs, Australia for helpful discussions.

\section*{References}

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
\textbf{Table 4.} HeLa cell invasion ability and \textit{in vivo} virulence in mice of 20 \textit{L. monocytogenes} lineage II strains representing three sublineages. \\
\hline
\textbf{Sublineage} & \textbf{Strain} & \textbf{PMSC in inlA} & \textbf{Relative invasion rate ± SD\(^a\)} & \textbf{log LD\textsubscript{50} in ICR mice} \\
\hline
\textit{IIA} & P19 & None & None & 82.8 ± 13.1 & 5.95 \\
& V1 & None & None & 132.6 ± 18.6 & 6.11 \\
& C10 & None & None & 65.4 ± 15.1 & 6.02 \\
& S9 & None & None & 71.3 ± 8.7 & 6.31 \\
& C4 & 1380 (G→A) & 460 & 18.5 ± 7.3** & 7.87 \\
& P20 & 1380 (G→A) & 460 & 12.2 ± 5.3** & 7.71 \\
\textit{IIB} & S1 & None & None & 76.2 ± 15.7 & 5.53 \\
& 10403S & None & None & 100 ± 0 & 5.49 \\
& M4 & None & None & 72.5 ± 7.2 & 5.45 \\
& M6 & None & None & 82.5 ± 18.6 & 5.55 \\
& S3 & None & None & 76.2 ± 12.2 & 6.19 \\
& C18 & None & None & 80.2 ± 9.9 & 5.98 \\
\textit{IIC} & P3 & None & None & 68.9 ± 21.2 & 6.07 \\
& P6 & None & None & 72.7 ± 12.2 & 5.73 \\
& S11 & None & None & 117.4 ± 28.0 & 6.01 \\
& M8 & None & None & 87.5 ± 6.6 & 5.75 \\
& C2 & 1474 (C→T) & 492 & 7.6 ± 2.6** & 7.55 \\
& C6 & 1474 (C→T) & 492 & 6.7 ± 2.8** & 8.05 \\
& P7 & 1474 (C→T) & 492 & 10.3 ± 6.6** & 8.07 \\
& B4 & 1474 (C→T) & 492 & 6.1 ± 3.0** & 7.80 \\
\hline
\hline
\hline
\end{tabular}
\caption{The invasion efficiencies were normalized to that of a reference strain, 10403S (the level of invasion by this strain was set at 100\%). A two-tailed Student’s t-test was applied to compare the invasion rate of each strain containing PMSC in \textit{inlA} to the average invasion rate of strains bearing intact \textit{inlA}. **p<0.01.}
\end{table}


