



Abstract

The genus *Listeria* is composed of eight species: two phylogenetically distant species, *L. grayi* and *L. rocourtae*, and six closely related species, *L. seeligeri*, *L. ivanovii*, *L. welshimeri*, *L. innocua*, *L. marthii* and *L. monocytogenes*. Only two species are facultative pathogens (*L. monocytogenes* and *L. ivanovii*), while the other six species represent obligate saprotrophs. In the study presented here we used a comparative genomic approach to characterize the mobilome (plasmids, phages and other mobile elements) of *L. seeligeri*, *L. ivanovii*, *L. welshimeri*, *L. innocua*, *L. marthii* and *L. monocytogenes*. We used three approaches to identify mobile elements: (i) a homology search for mobile element related genes in the primary annotation of the genomes, (ii) a search for phage related genes using Prophinder and (iii) a hidden Markov model search to find putative genes introduced by horizontal gene transfer using the SIGI-HMM software. The SIGI-HMM search identified between 2% (*L. ivanovii*) and 6% (*L. marthii*) of the ORFs in each genome as having been introduced by horizontal gene transfer. Examination of the primary sequence annotation showed that insertion sequence (IS) related transposases were present in all genomes examined, however intact IS (i.e. elements with both ORF A and ORF B) could not be found. Integrative Conjugative Elements (ICE) seem to be rare but wide possibly widespread in the genus *Listeria*; one ICE was found in the genome of an *L. monocytogenes* lineage IIC strain and based on the characteristics of this ICE similar elements could be found in other genomes. A plasmid bearing arsenic- and cadmium-resistance genes was found in an *L. seeligeri* strain. Similar plasmids have been found in *L. innocua* and *L. monocytogenes* genomes, suggesting that plasmids like this are involved in interspecific horizontal gene transfer in the genus *Listeria*.

Introduction

The genus *Listeria* is composed of eight species: *L. monocytogenes*, *L. innocua*, *L. marthii*, *L. welshimeri*, *L. seeligeri*, *L. ivanovii*, *L. rocourtae* and *L. grayi*. Two species, *L. monocytogenes* and *Listeria ivanovii* are known pathogens and cause listeriosis in human and warm-blooded hosts. These pathogenic species are closely related to non-pathogenic species: *L. monocytogenes* is closely related to *L. innocua* and *L. marthii*, and *L. ivanovii* is closely related to *L. seeligeri*.

The advent of next generation sequencing systems has made it possible to rapidly sequence multiple bacterial genomes to extremely high coverage. Here we present data obtained with the Applied Biosystems SOLiD™ system, a high throughput sequencing system that produces 400 million short reads per run. When sequencing eight *Listeria* genomes, each with an average genome size of 3 Mbp, in a single run with 25 bp reads, this throughput translates into a projected coverage of over 200X per genome. We used three approaches to identify mobile elements: (i) a homology search for mobile element related genes in the primary annotation of the genomes, (ii) a search for phage related genes using Prophinder and (iii) a hidden Markov model search to find putative genes introduced by horizontal gene transfer using the SIGI-HMM software.

Materials and Methods

Isolate/Strains sequenced:

- 1. *L. marthii* isolate FSL S4-120:** This isolate has been designated as the type strain of *L. marthii*. *L. marthii* has recently been discovered in several natural areas in the Finger Lake region in NY. It is non-pathogenic and sequence data show that the pathogenicity island is completely missing in this species.
- 2. *L. innocua* (hemolytic) isolate FSL J1-023:** *L. innocua* is a non-pathogenic species and the vast majority of strains lack genes that are involved in pathogenicity. This strain is exceptional in that it contains the pathogenicity island (Johnson et al., 2004) and a homolog of *inIA* (Volokov et al., 2007). The genome sequence of this strain will help to understand the role of horizontal gene transfer and recombination in the evolution of pathogenicity in *Listeria*.
- 3. *L. innocua* (non-hemolytic) isolate FSL S4-378:** Preliminary analyses of MLST data for *L. innocua* suggest that this species shows a high frequency of intra- and interspecific recombination. This strain is distantly related to the already published *L. innocua* genome and the genome sequence of this strain will help us understand the overall importance of homologous recombination in the evolution of *L. innocua* and *L. monocytogenes*.
- 4. *L. seeligeri* (hemolytic) isolate FSL N1-067:** Though *L. seeligeri* is not considered a pathogen, it does contain the pathogenicity island and homologues of other genes involved in pathogenicity in *L. monocytogenes*. Genome sequence data of this species may provide insight why this species is not a pathogen but does have all the genes involved in pathogenicity.
- 5. *L. seeligeri* (non-hemolytic) isolate FSL S4-171:** This isolate is very closely related to FSL N1-067, however it lacks some of the pathogenicity genes like hemolysin. Genome sequence data will help to understand the mechanism behind the loss of these pathogenicity genes.
- 6. *L. ivanovii* subsp. *londoniensis* ATCC 49954:** *Listeria ivanovii* is a pathogen of ruminants and is closely related to *L. seeligeri*. This species seems to be more host-adapted than *L. monocytogenes* as it is only reported in a small number of human listeriosis cases. Two subspecies are recognized within this species; subsp. *ivanovii* and subsp. *londoniensis*. This genome sequence may help us to understand why this species is relatively host specific as compared to *L. monocytogenes*.
- 7. *Listeria monocytogenes* lineage IIC:** *Listeria monocytogenes* can be subdivided into several evolutionary lineages; lineage I, lineage II, lineage III and lineage IV. Lineage III isolates form a distinct clade from the lineage I and II isolates. Sequence analyses have shown that lineage III isolates are frequently involved in intra and inter specific recombination.

Genome sequencing and assembly. Genomes were sequenced using the SOLiD™ system (Applied Biosystems, Foster City, CA). Mate-paired libraries with approximately 1.5 kb insertions were constructed and deposited on one quarter of a flowcell. Then, 25 bp reads were obtained from each of the F3 and R3 tags. Between 27 million and 57 million reads were obtained for each of the genomes. Referenced assembly was performed using the Applied Biosystems *corona_lite* package. After correcting errors in color-space reads using a modified version of the spectral alignment tools from the EULER-USR package (Chaisson, et al., 2009), de novo assembly was performed using the SOLiD™ de novo pipeline, which employs the Velvet assembly engine (Zerbino & Birney, 2008).

In order to identify likely misassemblies, scaffolds were aligned using MUMmer to the most closely related reference genome: *L. monocytogenes* scaffolds were aligned to *L. monocytogenes* F2365, *L. innocua* and *L. marthii* scaffolds were aligned to *L. innocua* CLIP11262, and *L. ivanovii* and *L. seeligeri* scaffolds were aligned to *L. welshimeri* SLC05334. Scaffolds were broken at points where non-contiguous regions of the reference genome were juxtaposed, and then ordered such that they were syntenic with the reference genome. All scaffolds were then concatenated into a single pseudogenome.

Whole genome alignments, Horizontal gene transfer and mobile elements. Whole genome alignments were created using Mauve (Darling et al., 2004). The program SIGI-HMM (Wassik et al., 2006) was used to infer if genes in the genomes were acquired through horizontal gene transfer (HGT). SIGI-HMM uses a codon usage based hidden Markov model to infer if genes are of alien origin (i.e. have been introduced into the genome from a divergent gene pool by HGT), taking into account that highly expressed genes may have codon bias. Genes with more than 10% ambiguous sites were excluded from the analyses. As recommended by the software the detection sensitivity of the program was adjusted to a level that enabled the detection of known foreign genomic elements (this detection sensitivity was 0.95, the highest possible sensitivity). Prophages and prophage derived regions were identified using the online version of Prophinder (Lima-Mendez et al., 2008). Transposons and plasmid related genes were identified with the SIGI-HMM program and through examination of the initial genome annotations.

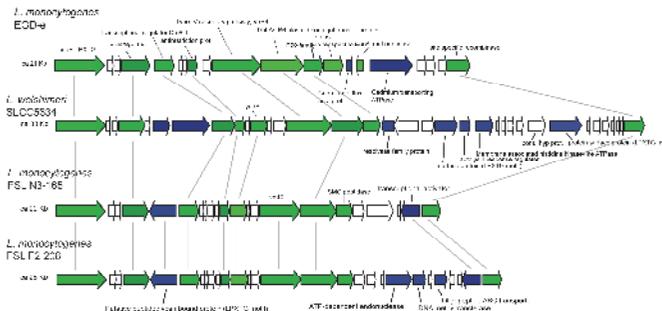


Figure 1. Integrative and Conjugative Elements (ICE) found in *Listeria* spp. Green arrows indicate genes putatively involved in conjugative transfer, blue arrows are accessory genes with a predicted function and white arrows are accessory genes without a known function ('hypothetical proteins'). The back bone of the ICE elements was found in *L. monocytogenes* FSL F2-208 and previously published genomes of *L. monocytogenes* (FSL N3-165, The Broad Institute Genome Sequencing Platform, EGD-e, Glasser et al., 2001) and *L. welshimeri* (Hain et al., 2006) is homologous to ICE t916, however most of the accessory genes of this ICE are not involved in antibiotic resistance, or involved in heavy metal resistance (with the exception of the ICE found in EGD-e). A feature that seems to be conserved in these ICEs is the presence of a *cnab* (Collagen Binding) motif containing protein, which is putatively involved in conjugative transfer.

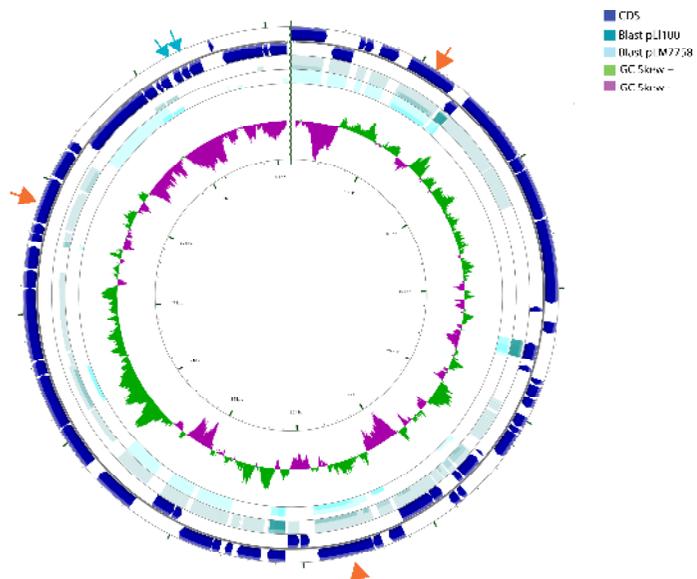


Figure 2. Circular representation of the plasmid found in *L. seeligeri* FSL N1-067 and BLAST similarity to ORFs found in plasmids pL100, *L. innocua* CLIP11262 and pLM7758 (*L. monocytogenes* 08-7758). A large number of genes found in these three plasmids shows a high BLAST identity (>95%), including genes involved in heavy metal resistance (marked with orange arrows). Besides genes involved in heavy metal resistance, the plasmid of *L. seeligeri* N1-067 also contains two ORFs encoding two putative multidrug resistance genes (*erbA* and *erbB*, marked with blue arrows). The outermost two rings show ORFs found in the *L. seeligeri* FSL N1-067 plasmid. The next two rings show the positions of BLAST hits detected by blast searches against the plasmids pL100 (*L. innocua* CLIP11262) and pLM7758 (*L. monocytogenes* 08-7758) comparison protein sets. The height of each arc in the BLAST results rings is proportional to the percent identity of the hit. Overlapping hits appear as darker arcs. This figure was created using the CGView Server of the Stothard Research Group (http://stothard.afms.ualberta.ca/cgview_server/).

Results

Horizontal Gene Transfer

The proportion of ORFs predicted by SIGI-HMM to be introduced into a genome through HGT ranged from 2.0% in *L. ivanovii* to 6.4% in *L. marthii* (table 1). Only two prophage derived regions (one in *L. monocytogenes* HCC23 and one in *L. innocua* FSL J1-023) were identified as regions introduced by HGT, which can be explained by the fact that SIGI-HMM detects genes derived from HGT by their codon usage pattern. Most prophages have a codon usage pattern that is similar to their host, which indicates that these prophages have co-evolved with their *Listeria* hosts as suggested by Dorsch et al. (2009). A large (ca 10 kb) genomic island encoding a lantibiotic synthesis and transport operon was found in *L. marthii*. Although the codon usage bias clearly indicates that this island has been introduced into the genome of *L. marthii* by HGT, the island is probably not exclusive for *L. marthii* and may represent an adaptation of this species to its life as a soil bacterium.

Transposable Elements

ORFs encoding genes associated with two types of transposable elements (i.e., insertion sequence elements and integrative conjugative elements) were found in the primary annotation of the genomes. No evidence for the presence of functional insertion sequence elements was found in any of the seven *Listeria* genomes sequenced for this study. Interestingly, ORFA homologs in the ICE encoding the transcriptional regulator of the transposase encoded by ORFA and ORFB of insertion sequences elements are found in all *Listeria* genomes examined, suggesting that functional insertion sequence elements may have been present in the genome, however ORFB, the second ORF in the insertion sequence element is lost rapidly. A second type of transposable element was found in *L. monocytogenes* FSL F2-208. This transposable element is an integrative and conjugative element (ICE, see figure 1) and is very similar to t916, a classic example of an ICE. An interesting feature of this ICE is that it contains a surface protein with collagen binding motif repeats. This feature seems to be conserved in a group of ICEs found in *Listeria* and other Gram-positive bacteria and we used this feature to successfully find similar ICEs in previously published *Listeria* genomes (*L. welshimeri* SLC05334 and *L. monocytogenes* FSL N3-165, see figure 1). The previously reported ICE in *L. monocytogenes* EGD-e is classical in that it contains cadmium-resistance genes, however the accessory genes found in the other ICEs reported here do not seem to be involved in antibiotic or heavy metal resistance.

Plasmids

Only *L. seeligeri* FSL N1-067, an isolate from a food processing environment, contained a plasmid (figure 2). This plasmid encodes cadmium resistance genes, arsenic resistance genes and antibiotic resistance genes. Some of these genes, the cadmium resistance genes in particular, show a 100% identity at the nucleotide sequence level with the same genes found on plasmids in *L. innocua* and *L. monocytogenes* isolates from food processing environments, suggesting an important role for plasmids in the transfer of resistance genes between highly divergent pathogenic and non-pathogenic *Listeria* species. This is of particular interest since plasmid-borne cadmium resistance has been associated with benzalkonium chloride (Mullapudi et al., 2010), a disinfectant often used in food processing environments.

Table 1. Coverage of prediction of genes derived from HGT of all mobile elements in the *Listeria* genomes.

Strains analyzed are highly sequenced, strains in red are not highly sequenced in this study	% ORFs introduced by HGT	No. of prophages ^a	No. of transposons ^b	Plasmid resistance ^c	No. of transposable elements ^d
<i>L. monocytogenes</i>					
F2-208	2.6	0	1	10	0
EGD-e	3.4	1	1	10	0
CI IP# 150	3.2	0	1	10	0
HCC23	4.8	3	0	10	0
<i>L. marthii</i>					
FSL-2-203	4.5	1	1	10	1
FSL S4-120	2.0	0	1	10	0
<i>L. innocua</i>					
CLIP11262	4.1	0	1	10	0
FSL S4-378	3.7	2	1	10	0
FSL J1-023	4.5	0	1	10	0
<i>L. welshimeri</i>					
SLC05334	3.8	1	0	10	1
<i>L. ivanovii</i>					
FSL-15-018	2.0	1	2	10	0
<i>L. seeligeri</i>					
FSL S4-171	2.0	1	1	10	0
FSL N1-067	4.3	2	1	10	0

^a 12 prophages were detected by coverage from their host genes by using the SIGI-HMM program.

^b 20 mobile elements that are not highly sequenced in this study are included in this table.

^c The number of plasmids encoding antibiotic, prophage or other resistance genes.

^d 10 mobile elements that are not highly sequenced in this study are included in this table.

Conclusions

- Our analyses show that de novo assembly of short read sequences of platforms such as the SOLiD system can be a helpful tool in the identification of mobile elements in bacterial genomes.
- Integrative Conjugative Elements are probably the main group of functional transposable elements in *Listeria*, and while rare, they are found in multiple species.
- Plasmids are likely involved in interspecific horizontal gene transfer of heavy metal resistance genes in the genus *Listeria*.

Acknowledgements

We would like to thank the JCVI for providing the JCVI Annotation Service, which provided us with automatic annotation data and the manual annotation tool Manatee. Barbara Bowen and Esther Fortes for their help with additional Sanger sequencing and handling of the strains; Dima Birney for sharing de novo assembly tools prior to distribution. Part of this work was supported by USDA Special Research Grants (2005-34459-15625, 2006-34459-16952, and 2008-34459-19043).

References

- Chaisson MJ, Brinda D, Pevzner PA. 2009. De novo fragment assembly with short mate-paired reads: Does the read length matter? *Genome Res vol. 19* (2) pp. 336-46.
- Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res vol. 14* (7) pp. 1304-40.
- Dorsch J, Klump J, Bilmeyer R, Schmeicher M, Born Y, et al. 2009. Comparative genome analysis of *Listeria* bacteriophages reveals extensive mosaicism, programmed transpositional frameshifting, and a novel prophage insertion site. *J Bacteriol 191*: 7200-7215.
- Lima-Mendez G, Van Helden J, Traussnitz A, Lepoint R. 2008. Prophinder: a computational tool for prophage prediction in prokaryotic genomes. *Bioinformatics 24*:863-865.
- Mullapudi S, R M Sletsky, S Kahanlou, 2010. Diverse cadmium resistance determinants in *Listeria monocytogenes* isolates from the turkey processing plant environment. *Appl Environ Microbiol 76* (2): 627-30.
- Volokhov DV, Duprier S, Nevelev AA, George J, Buchrieser C, Hiltbrins AD. 2007. The presence of the interlamal gene in natural atypically hemolytic *Listeria innocua* strains suggests descent from *L. monocytogenes*. *Appl Environ Microbiol vol. 73* (6) pp. 1928-20.
- Wassik S, Keller C, Aaser R, Brodt T, Damm C, Fritsche JF, Surwicki K, Henricson P, Menni R. 2006. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics 7*:142.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res vol. 18* (5) pp. 821-9.