GENOMIC ANALYSIS OF *LISTERIA MONOCYTOGENES* FROM POULTRY FURTHER PROCESSING PLANTS

by

LAUREN KATHRYN HUDSON

(Under the Direction of Mark A. Harrison)

ABSTRACT

There have been many reports that some *Listeria monocytogenes* strains are able to persist in food processing facilities, creating opportunities to cause food contamination and foodborne illness. In this study, we used Whole Genome Sequencing (WGS) and subsequent analyses to investigate the genomes of 156 *L. monocytogenes* isolates collected from two previous longitudinal sampling studies of poultry further processing plants. The first objective was to sequence and characterize isolates to determine phylogeny, MLST-types, serotypes, and the presence of virulence and antibiotic-, stress-, and sanitizer-resistance genes. Fifty-six isolates belong to lineage I; 99 belong to lineage II. Eighteen MLST sequence types (ST) were identified, the majority of which were ST321 (n=41), ST5 (n=31), ST155 (n=27), and ST6 (n=20). All genomes contained *fosX*, *lmo0441*, *lmo0919*, *norB*, and *sul* antibiotic-resistance genes, 14.1% (n=22) contained *aacA4*, and 3.2% (n=5) contained *tetM*. Of the genomes, 82.7% (n=129) had 5 total antibiotic resistance (ABR) genes identified, 16% (n=25) had 6 total, and 1.3% (n=2) had 7 total. Of isolates screened, 82.7% (n=129) possessed the following genes related to stress-resistance: *lmo0444*, *lmo0445*, *lmo0446*, *lmo0447*, and *lmo0448*, all contained in a stress survival islet (SSI-1). The *bcrA*, *bcrB*, and *bcrC* genes conferring benzalkonium chloride resistance were

in 73.1% (n=114) of isolates. Results provide insight about *L. monocytogenes* found in poultry further processing plants, including potentially high rates of ABR, stress-resistance, and sanitizer-resistance genes, and can be used to evaluate current and devise more effective *L. monocytogenes* control methods. The second objective of the study was to identify genes that were significantly positively or negatively associated with repeated isolation in those facilities. We found a total of 352 genes or clusters significantly associated with repeated isolation (naive P-value < 0.05 and a Bonferroni-corrected P-value < 0.05), with 180 positively and 172 negatively associated. Some notable genes/gene clusters were annotated as internalin precursors (n=28; *inlJ*, *inlA*, and *inlB*), CRISPR-associated (n=6), and various other genes (*iap*, *cadA*, etc.) These genes can potentially be used to further understanding of *L. monocytogenes* persistence in food processing environments, as markers to differentiate persistent and transient strains, or to improve methods used to clean and sanitize food processing environments.

INDEX WORDS: *Listeria*, *Listeria monocytogenes*, *Listeria monocytogenes* persistence, Whole-genome sequencing, Poultry processing plants

GENOMIC ANALYSIS OF *LISTERIA MONOCYTOGENES* FROM POULTRY FURTHER PROCESSING PLANTS

by

LAUREN KATHRYN HUDSON

B.S., University of Florida, 2012

M.S., University of Georgia, 2014

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2018

GENOMIC ANALYSIS OF *LISTERIA MONOCYTOGENES* FROM POULTRY FURTHER PROCESSING PLANTS

by

LAUREN KATHRYN HUDSON

Major Professor:	Mark A. Harrison
Committee:	Mark E. Berrang Richard J. Meinersmann Xiangyu Deng

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2018

DEDICATION

To my amazing parents, David and Lisa Hudson. You first taught me the value of education and persistence. I would not be where I am today without your constant support, love, and guidance.

To my brother, Garrett Hudson. Though you are not here to see my accomplishments, I know you would be proud. Your memory continues to motivate and inspire me.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Mark Harrison. He has been an outstanding mentor through the last six years, has motivated me to push my limits, and has been an integral part in my intellectual development. I have been fortunate to have the opportunity to work with both a great scientist and person. I would also like to thank my other committee members, Drs. Mark Berrang, Richard Meinersmann, and Xiangyu Deng, and express my appreciation for their support, patience, and guidance. In particular, I'm grateful to Dr. Berrang for his constant positivity. Additionally, I am thankful to Shaokang Zhang, Steven Knapp, Eric Adams, and Linda Genzlinger for your technical help and for graciously sharing your knowledge with me and to Gwen Hirsch for always being willing to lend a helping hand.

Furthermore, I am so very grateful for all the friends I have made throughout my time at the University of Georgia: George and Anuj, who have been here with me from day one; all of my labmates, Stephanie M., Winnie, Lee, Suejee, Charlotte, Laurie, Mikayla, Chase, and Brittany; Audrey and Tobias; Daniel; Bree; Jess; Lisa; Stephanie B.; Alex; Ana; Carlyn; Joe; Jill; and Justin. They have all provided an incredible amount of support and even more memories.

This work was made possible through financial support from the Georgia Agricultural Experiment Stations, the University of Georgia Center for Food Safety, USDA's National Needs Fellowship program, and the University of Georgia Graduate School.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Listeria monocytogenes, the causative agent of listeriosis, is a foodborne bacterial pathogen and was responsible for 116 laboratory-confirmed infections in 2015 *(5)*. Though *L. monocytogenes* infections are relatively rare compared to other foodborne pathogens (like *Salmonella*, *Escherichia coli*, and *Campylobacter*), it has very high hospitalization and death rates, with 111 hospitalizations (95.7% of total infections) and 15 deaths (12.93 Case Fatality Rate [CFR]) reported for 2015 *(5)*.

L. monocytogenes can enter food processing facilities, establish a niche and colonize the food processing environment. Numerous studies have shown that some subtypes of *L. monocytogenes* have the ability to persist in food processing and other food-related environments, while others may only be isolated sporadically in these same types of environments *(1-4, 6-22, 24-31)*. Even though *L. monocytogenes* was recognized as a foodborne pathogen in the 1980s *(23)* and has been extensively studied as a pathogen and environmental nuisance in food processing facilities, little insight has been gained into the reasons why some strains are persistent in processing facilities and some are transient.

A search of "*Listeria monocytogenes* persistence" using Google Scholar, the publically available web search engine of scholarly literature, provides approximately 45,700 search results. This is evidence of the vast amount of research that has been done regarding this topic. However, with the ever increasing availability and decreasing cost to sequence bacterial genomes using Next Generation Sequencing (NGS) techniques and the emergence of more and more methods to

analyze these genomes, we thought it would be valuable to use this technology to further study the genomes of *L. monocytogenes* from food processing facilities (objective one; Chapter 3) and attempt to gain a greater understanding of the differences between persistent and transient strains (objective 2; Chapter 4).

References

1. Autio, T., R. Keto-Timonen, J. Lunden, J. Bjorkroth, and H. Korkeala. 2003. Characterisation of persistent and sporadic *Listeria monocytogenes* strains by pulsed-field gel electrophoresis (PFGE) and amplified fragment length polymorphism (AFLP). *Syst. Appl. Microbiol.* 26:539-545.

2. Berrang, M. E., R. J. Meinersmann, J. F. Frank, and S. R. Ladely. 2010. Colonization of a newly constructed commercial chicken further processing plant with *Listeria monocytogenes*. *J. Food Prot.* 73:286-291.

3. Berrang, M. E., R. J. Meinersmann, J. F. Frank, D. P. Smith, and L. L. Genzlinger. 2005. Distribution of *Listeria monocytogenes* subtypes within a poultry further processing plant. *J. Food Prot.* 68:980-985.

4. Carpentier, B., and O. Cerf. 2011. Review--persistence of *Listeria monocytogenes* in food industry equipment and premises. *Int. J. Food Microbiol.* 145:1-8.

5. Centers for Disease Control and Prevention (CDC). 2017. Foodborne diseases active surveillance network (foodnet): Foodnet 2015 surveillance report (final data). Available at: https://www.cdc.gov/foodnet/reports/annual-reports-2015.html. Accessed 1 April 2018.

6. Cruz, C. D., and G. C. Fletcher. 2011. Prevalence and biofilm-forming ability of *Listeria monocytogenes* in New Zealand mussel (*perna canaliculus*) processing plants. *Food Microbiol.* 28:1387-1393.

7. Fagerlund, A., S. Langsrud, B. C. T. Schirmer, T. Moretro, and E. Heir. 2016. Genome analysis of *Listeria monocytogenes* sequence type 8 strains persisting in salmon and poultry processing environments and comparison with related strains. *PLoS One*. 11:e0151117.

8. Ferreira, V., M. Wiedmann, P. Teixeira, and M. J. Stasiewicz. 2014. *Listeria monocytogenes* persistence in food-associated environments: Epidemiology, strain characteristics, and implications for public health. *J. Food Prot.* 77:150-170.

9. Harvey, J., and A. Gilmour. 1994. Application of multilocus enzyme electrophoresis and restriction fragment length polymorphism analysis to the typing of *Listeria*

monocytogenes strains isolated from raw milk, nondairy foods, and clinical and veterinary sources. *Appl. Environ. Microbiol.* 60:1547-1553.

10. Heir, E., B.-A. Lindstedt, O.-J. Røtterud, T. Vardund, G. Kapperud, and T. Nesbakken. 2004. Molecular epidemiology and disinfectant susceptibility of *Listeria monocytogenes* from meat processing plants and human infections. *Int. J. Food Microbiol.* 96:85-96.

11. Keto-Timonen, R., R. Tolvanen, J. Lunden, and H. Korkeala. 2007. An 8-year surveillance of the diversity and persistence of *Listeria monocytogenes* in a chilled food processing plant analyzed by amplified fragment length polymorphism. *J. Food Prot.* 70:1866-1873.

12. Knudsen, G. M., J. B. Nielsen, R. L. Marvig, Y. Ng, P. Worning, H. Westh, and L. Gram. 2017. Genome-wide-analyses of *Listeria monocytogenes* from food-processing plants reveal clonal diversity and date the emergence of persisting sequence types. *Environ. Microbiol. Rep.* 9:428-440.

13. Korenova, J., K. Oravcova, A. Veghova, R. Karpiskova, and T. Kuchta. 2016. Biofilm formation in various conditions is not a key factor of persistence potential of *Listeria monocytogenes* in food-processing environment. *J. Food Nutr. Res.* 55:189-193.

14. Leong, D., K. NicAogáin, L. Luque-Sastre, O. McManamon, K. Hunt, A. Alvarez-Ordóñez, J. Scollard, A. Schmalenberger, S. Fanning, C. O'Byrne, and K. Jordan. 2017. A 3-year multi-food study of the presence and persistence of *Listeria monocytogenes* in 54 small food businesses in Ireland. *Int. J. Food Microbiol.* 249:18-26.

15. Lunden, J., T. Autio, A. Markkula, S. Hellstrom, and H. Korkeala. 2003. Adaptive and cross-adaptive responses of persistent and non-persistent *Listeria monocytogenes* strains to disinfectants. *Int. J. Food Microbiol.* 82:265-272.

16. Lunden, J., R. Tolvanen, and H. Korkeala. 2008. Acid and heat tolerance of persistent and nonpersistent *Listeria monocytogenes* food plant strains. *Lett. Appl. Microbiol.* 46:276-280.

17. Lunden, J. M., T. J. Autio, A. M. Sjoberg, and H. J. Korkeala. 2003. Persistent and nonpersistent *Listeria monocytogenes* contamination in meat and poultry processing plants. *J. Food Prot.* 66:2062-2069.

18. Lunden, J. M., M. K. Miettinen, T. J. Autio, and H. J. Korkeala. 2000. Persistent *Listeria monocytogenes* strains show enhanced adherence to food contact surface after short contact times. *J. Food Prot.* 63:1204-1207.

19. Møretrø, T., and S. Langsrud. 2004. *Listeria monocytogenes*: Biofilm formation and persistence in food-processing environments. *Biofilms*. 1:107-121.

20. Morganti, M., E. Scaltriti, P. Cozzolino, L. Bolzoni, G. Casadei, M. Pierantoni, E. Foni, and S. Pongolini. 2016. Processing-dependent and clonal contamination patterns of *Listeria monocytogenes* in the cured ham food chain revealed by genetic analysis. *Appl. Environ. Microbiol.* 82:822-831.

21. Murugesan, L., Z. Kucerova, S. J. Knabel, and L. F. LaBorde. 2015. Predominance and distribution of a persistent *Listeria monocytogenes* clone in a commercial fresh mushroom processing environment. *J. Food Prot.* 78:1988-1998.

22. Palma, F., F. Pasquali, A. Lucchi, A. De Cesare, and G. Manfreda. 2017. Whole genome sequencing for typing and characterisation of *Listeria monocytogenes* isolated in a rabbit meat processing plant. *Ital. J. Food Saf.* 6:6879.

23. Ray, B., and A. Bhunia. 2014. Fundamental food microbiology CRC Press, Boca Raton, FL.

24. Rorvik, L. M., B. Aase, T. Alvestad, and D. A. Caugant. 2003. Molecular epidemiological survey of *Listeria monocytogenes* in broilers and poultry products. *J. Appl. Microbiol.* 94:633-640.

25. Rørvik, L. M., D. A. Caugant, and M. Yndestad. 1995. Contamination pattern of *Listeria monocytogenes* and other *Listeria* spp. In a salmon slaughterhouse and smoked salmon processing plant. *Int. J. Food Microbiol.* 25:19-27.

26. Senczek, D., R. Stephan, and F. Untermann. 2000. Pulsed-field gel electrophoresis (PFGE) typing of *Listeria* strains isolated from a meat processing plant over a 2-year period. *Int. J. Food Microbiol.* 62:155-159.

27. Stasiewicz, M. J., H. F. Oliver, M. Wiedmann, and H. C. den Bakker. 2015. Whole-genome sequencing allows for improved identification of persistent *Listeria monocytogenes* in food-associated environments. *Appl. Environ. Microbiol.* 81:6024-6037.

28. Vangay, P., J. Steingrimsson, M. Wiedmann, and M. J. Stasiewicz. 2014. Classification of *Listeria monocytogenes* persistence in retail delicatessen environments using expert elicitation and machine learning. *Risk Anal.* 34:1830-1845.

29. Verghese, B., M. Lok, J. Wen, V. Alessandria, Y. Chen, S. Kathariou, and S. Knabel. 2011. Comk prophage junction fragments as markers for *Listeria monocytogenes* genotypes unique to individual meat and poultry processing plants and a model for rapid niche-specific adaptation, biofilm formation, and persistence. *Appl. Environ. Microbiol.* 77:3279-3292.

30. Wang, J. J., A. J. Ray, S. R. Hammons, and H. F. Oliver. 2015. Persistent and transient *Listeria monocytogenes* strains from retail deli environments vary in their ability to adhere and form biofilms and rarely have inla premature stop codons. *Foodborne Pathog. Dis.* 12:151-158.

31. Williams, S. K., S. Roof, E. A. Boyle, D. Burson, H. Thippareddi, I. Geornaras, J. N. Sofos, M. Wiedmann, and K. Nightingale. 2011. Molecular ecology of *Listeria monocytogenes* and other *Listeria* species in small and very small ready-to-eat meat processing plants. *J. Food Prot.* 74:63-77.

CHAPTER 2

LITERATURE REVIEW

There are currently 18 recognized species in the genus *Listeria (18, 37, 39, 52, 56)*. *L. monocytogenes* is a mesophile and can grow between 0-45°C, with growth rates slowing at lower temperatures *(59)*. Optimal growth is at 30-37°C *(46, 47, 59)* and pH 6-9 *(46)*. Cells can be killed at >50°C *(47)*. *L. monocytogenes* can grow in vacuum- and modified-air packaged food products at 0-1°C *(59)*. *L. monocytogenes* is relatively resistant to freezing, drying, high salt (>10%), and pH 5.0 and above (up to pH 10-11) and is acid tolerant *(46, 59, 72)*. Freezing does not significantly reduce cell numbers, but the ability to survive during freezing and injury to cells that occurs during freezing are dependent on the freezing rate and the food matrix *(47)*. *L. monocytogenes* can grow at pH values between 4.4-9.6 *(47, 62)*. Growth in low pH environments is influenced by temperature and the type of acid present *(62)*. At pH 4.3 and below, cells may still survive, but will not proliferate *(47)*. Optimal water activity for growth of *L. monocytogenes* is ≥0.97 *(47, 62)*. The minimum a_w is 0.93 for most strains, but as low as 0.90 for some *(47)*. It is also able to survive for extended periods at an a_w as low as 0.83 *(47, 62)*. Research has shown that as a_w of the medium *L. monocytogenes* is suspended in decreases, thermal resistance increases (inverse relationship) *(47, 62, 69)*. *L. monocytogenes* can grow at 10-12% sodium chloride *(47, 62)*. In addition, it can also survive for long periods at high salt concentrations, with survival being significantly increased at lower temperatures *(47)*.

L. monocytogenes is broadly distributed in the environment, hardy in adverse environmental conditions, and can proliferate in phagocytes *(47)*. The bacterium can persist for

lengthy times in/on food, soil, processing plants, and hard surfaces *(47)*. These qualities make controlling *Listeria* in food production facilities particularly challenging. Many previous studies have demonstrated the ability of *L. monocytogenes* to persist in food production and other food-associated environments *(8, 20, 65, 67, 70)*. Due to *L. monocytogenes* being resistant to many environmental conditions (low temperatures, pH, salt, low water activity, etc.), controlling the pathogen in foods must be achieved through a multiple hurdle approach and through preventative measures (such as hygienic design and good sanitation practices) and environmental testing and monitoring.

L. monocytogenes can be found in a large variety of food products (both of animal and plant origin) and has been isolated from raw milk, soft cheeses, fresh and frozen meat, poultry, and seafood products, eggs, and fruit and vegetable products (especially leafy vegetables and tubers like potatoes and radishes) *(26, 59)*. Thermally processed food products that contain *L. monocytogenes* after processing, either have not been properly treated or were contaminated after thermal processing.

Outbreaks and Recalls

L. monocytogenes was first recognized as a contaminant of plant- and animal-based foods and as the cause of foodborne illnesses in the 1980s, though human listeriosis has been known about for much longer *(59)*. The first confirmed foodborne listeriosis illness occurred in 1983. In contrast to many other foodborne illnesses, listeriosis is sporadic and rare, but it can be very severe *(47)*. The first documented foodborne listeriosis outbreak was identified in 1991 and was linked to coleslaw *(47)*. Since then, *Listeria* contamination of deli meats is estimated to cost $1.1 billion yearly in the United States *(47)*. In addition to monetary costs, 4,000 quality-adjusted life years are also lost *(47)*.

The USDA Food Safety Inspection Service (FSIS) issues approximately 30 product recalls due to *L. monocytogenes* contamination yearly, and this number has increased *(47)*. This increase is likely due to increased food testing and environmental monitoring done by food processors leading to increased finding of *L. monocytogenes.* Often products are recalled as a precaution (i.e., *L. monocytogenes* may be found in a facility but not in the actual food product) *(47)*. Some believe that it is impossible to have foods free of *L. monocytogenes* due to its ubiquitous nature *(59)*. Regulatory agencies have relied heavily on strong *Listeria* control programs in place at food production facilities to reduce *L. monocytogenes* contamination and the resulting illnesses, outbreaks, and deaths *(59)*.

There are a few key reasons *L. monocytogenes* has an increasing role as a foodborne pathogen in the last few decades: *L. monocytogenes* can contaminate and grow in ready-to-eat (RTE) foods; the increased use of refrigeration to preserve foods; the increased consumption of RTE products, especially those that are stored for extended time periods; and the situation that many RTE products are consumed without proper heating or microwave heating *(47, 59)*. Some newer technological methods used for the production of RTE products have many points in processing in which *L. monocytogenes* may contaminate product at low levels *(59)*. Low-level contamination events can become more serious if temperature abuse occurs and the pathogen is allowed to grow to higher concentration during refrigerated storage *(59)*. Centralization and consolidation of food production facilities can make sanitation practices more difficult, which may aid in the emergence of listeriosis prevalence *(47)*.

Stress Adaptation

When exposed to stress conditions, such as high salt (>10%) or high temperature (>45^{o}C), *L. monocytogenes* may adapt *(59)*. Acid-adapted *L. monocytogenes* has been

demonstrated to survive when exposed to pH environments as low as 3.5 and has been associated with the development of nisin-resistance *(59)*. Compared to unadapted control cells, *L. monocytogenes* cells temporarily exposed to 0.1% H_2O_2 also showed evidence of cross-resistance to successive exposure to 0.5% H_2O_2, 5% ethanol, 7% NaCl, pH 5.0, or 45°C *(59)*. *L. monocytogenes* cells that were acid adapted by exposure to pH 5 and then suspended in low-pH fruit juices have also demonstrated increased heat resistance (increase in D-value) *(59)*. This indicates that stress adaptation can also occur in food systems, and not just culture media. When exposed to sublethal stresses, specifically a high temperature (45°C) or 5.5% NaCl, *L. monocytogenes* cells have been shown to elongate or associate in chain formations *(59)*. It is thought that stress can affect the morphology of cells though the down-regulation of gene products related to cell division (muramidase, autolysin amidase, etc.) *(59)*.

Listeria contamination of food products can typically occur in two ways: primary (occur in raw materials and pass through to the final product) or secondary (contamination of the product during or after processing) *(35, 59)*. Secondary contamination of food products is typically thought to occur by strains of *L. monocytogenes* that are resistant to cleaning and sanitation and may persist in the food processing environment *(35)*.

Listeria can enter a processing plant a variety of different ways, including via soil on items like workers' clothes or shoes, transport equipment, contaminated machinery, animals/pests that either excrete the bacterium or have contaminated external surfaces, raw plant material, raw meat foods, or possibly even asymptomatic human carriers *(62)*. *Listeria* will often find niches in processing plants that are moist with sufficient nutrients to support growth, most often floor drains, condensate, standing/stagnant water, floors, and areas on processing equipment with organic residue *(59, 62)*. On equipment, small parts, unreachable portions, and

some types of materials may not be able to be effectively cleaned and sanitized, and can subsequently serve as niches for *L. monocytogenes (59)*.

Persistence in Food Processing and other Food-Related Environments

Numerous studies have shown that some subtypes of *L. monocytogenes* have the ability to persist in food processing and other food-related environments, while others may only be isolated sporadically in these same types of environments *(3-5, 8, 14, 19, 20, 22-24, 27, 33-35, 38, 40-43, 48, 49, 51, 55, 60, 61, 65, 67, 71, 74, 75, 78, 79)*. One flaw in the study of persistence of *L. monocytogenes* in food-associated environments is the lack of a standard and universal definition of "persistence" and formal (statistical) criteria for identifying persistent strains *(20)*. When identifying persistent strains, most researchers use repeated isolation of a subtype from a specific environment over a given amount of time as the criterion *(20)*. It is also important to note that terms like "persistent" and "nonpersistent" are not absolutes, as strain types could have a different persistence status in different facilities, environments, conditions, etc *(27)*.

In addition, a variety of different subtyping methods have been used in these persistence-related studies, including both phenotypic and genotypic methods, such as PFGE typing, MLST typing, ribotyping, serotyping, single-locus typing, SNP-based subtyping, etc. *(20)*. Different subtyping methods have different discriminatory power, so some may not be as appropriate as others to use for persistence studies *(20)*, and the variety of typing methods used makes it difficult to make comparisons between studies. In addition, the historical frequency of isolation of a specific subtype should play a role in determining persistence *(20)*. For example, repeated isolation of a historically rarely isolated subtype provides more evidence of persistence than repeated isolation of a historically commonly isolated subtype *(20)*. Malley, et al. developed a statistical-based approach to determine relative frequency of a subtype by comparing the

observed frequency to the historical frequency *(45)*. One caveat to this relative frequency-based approach to determining persistence is that obtaining the appropriate frequency information for different subtypes may be expensive and time-consuming and may not be possible for studying datasets of previously isolated *L. monocytogenes (20)*.

The threshold for repeated isolation to determine persistence often varies between studies as well, with some studies defining persistent strains to be subtypes found more than once over a given period and other studies using a higher threshold, such as being isolated three or more times over a given period. It is also difficult to distinguish between isolates repeatedly introduced (i.e., from raw ingredients, surrounding environments, etc.) in an environment and truly persistent *L. monocytogenes* isolates *(20)*.

Epidemiology and Public Health Surveillance

In the United States, listeriosis is nationally notifiable, which means that confirmed infections must be reported to local, state, territorial, or federal public health authorities, who monitor the disease and detect outbreaks *(12)*. *L. monocytogenes* was responsible for 116 laboratory-confirmed infections in 2015 *(11)*. Though *L. monocytogenes* infections are relatively rare compared to other foodborne pathogens (like *Salmonella*, *Escherichia coli*, and *Campylobacter*), it has very high hospitalization and death rates, with 111 hospitalizations (95.7% of total infections) and 15 deaths (12.93 Case Fatality Rate [CFR]) reported for 2015 *(11)*.

***Listeria* Whole Genome Sequencing Project.** Routine surveillance of *L. monocytogenes* isolates using whole genome sequencing began in 2013 as a collaboration project between the Centers for Disease Control and Prevention (CDC), U.S. Food and Drug Administration (FDA), U.S. Department of Agriculture (USDA), National Center for Biotechnology Information

(NCBI), and state and local health departments *(13)*. During foodborne illness investigations, data from PulseNet, the *Listeria* initiative, and GenomeTrakr are used. Through the use of both whole genome sequencing data (generated from patient, food, and environmental isolates) and data obtained from infected patients about foods eaten, public health officials aim to detect more *L. monocytogenes* infection clusters (possible outbreaks), attribute cases of listeriosis to likely sources, identify previously unknown or emerging sources of *L. monocytogenes*, and end listeriosis outbreaks while they are relatively small *(9, 13)*. After whole genome sequencing began being used to aid in the detection and investigation of listeriosis outbreaks, the total number of outbreaks solved has increased and the number of cases per outbreak have decreased *(9, 13)*.

Government Regulation

Due to the aforementioned characteristics that make *L. monocytogenes* particularly hardy and difficult to control in food production environments and in food products, many experts believe that it is an impossible task to produce "*Listeria* free" food *(47, 59)*. However, because listeriosis is so severe and has such a high case-fatality rate, it must be controlled *(47)*. This has created quite a conundrum for those in charge of food safety in industry and those charged with creating regulations involving the control of *L. monocytogenes* in foods and food processing environments.

In the United States, a "zero-tolerance" of *L. monocytogenes* in RTE products policy is in place *(59)*. Conversely, other countries, including Canada and some European countries, have established an acceptable level of 100 *L. monocytogenes* cells per 25 g RTE food product *(59)*. The U.S. FDA conducted a risk assessment of *L. monocytogenes* in different foods in 2003 *(59)*.

This consisted of the analysis of many RTE foods and categorizing them into five groups based on risk: very high, high, moderate, low, and very low risk *(59)*.

Table 2.1. *L. monocytogenes* Risk Categories and Associated Food Groups (From Ray, et al. *(59)*)

Risk Category	Foods
Very High	Deli meats and frankfurters (not reheated)
High	High fat dairy products, pasteurized fluid milk, pâté and meat spreads, soft unripened cheeses, smoked seafood, unpasteurized fluid milk
Moderate	Cooked RTE crustaceans, deli salads, dry/semidry fermented sausages, fresh soft cheese, fruit, semisoft cheese, soft ripened cheese, vegetables, frankfurters (reheated)
Low	Preserved fish, raw seafood
Very Low	Cultured milk products, hard cheese, ice cream and frozen dairy products, processed cheese

Regulatory agencies have used consumer education campaigns to help reduce foodborne listeriosis *(59)*. These campaigns include messages about thoroughly cooking raw meat, poultry, and seafood products; thoroughly washing raw fruits and vegetables; keeping uncooked meats separate from vegetables, cooked foods, and RTE foods; avoiding raw milk or products made with raw milk; and washing hands, knives, and cutting boards after preparing or handling uncooked foods *(59)*. Special dietary recommendations have also been specified for listeriosis-

susceptible populations, including: avoid soft cheeses (Mexican style, feta, Brie, Camembert, blue-veined, cream, or cottage cheeses); reheat (until steaming) all refrigerated leftover foods and RTE foods before eating; and pregnant women, the elderly, and the immunocompromised should avoid foods from retail delicatessens, such as RTE meat or poultry products *(59)*.

Listeria monocytogenes Genome

L. monocytogenes contains one circular chromosome and is considered a low G+C content *(73)* bacterium with a G+C content of 36-42% *(62)*. The genome of *L. monocytogenes* is relatively stable *(36)*, and consistent in size, within a range of 2,893 to 3,011 kb (*25*). However, it is open and able to generate or include new genetic material, which may explain the organism's ability to adapt to new niches in food processing facilities. Kuenne, et al. *(36)* examined both the core and accessory genome of 16 *L. monocytogenes* strains (with representatives from each serotype) and found that the core genomes of all strains were highly syntenic and that accessory genes were either scattered along the chromosomes (46%) or found inside hyper variable regions (20% when excluding MGE). These gene-level differences resulted from nine hyper variable hotspots, eight different prophages, three transposons (Tn916, Tn554, IS3-like), and two mobilizable islands.

Evolution and Lineages

L. monocytogenes strains can be grouped into four evolutionary lineages *(36, 77)*. Originally, three lineages were defined by both ribotype patterns and the allelic analysis of *hlyA* and *actA*, two virulence genes *(77)*. Later, lineage III was divided into two the sublineages IIIA and IIIB and more recently, IIB has been re-assigned as lineage IV *(62)*. Lineage IIIA is rhaminose-positive and lineage IIIB is rhaminose-negative *(59)*. Lineage I contains mostly epidemic clones *(59)* and the strains are typically responsible for human listeriosis cases and

outbreaks. Lineage II strains are more sporadically isolated from humans and animals *(59)*. Lineage III and IV isolates are rare in humans and are typically only isolated from animals *(36, 62)*.

Table 2.2. Classification of *Listeria monocytogenes* Based on Genomic Fingerprinting and Pathogenic Potential (Epidemic Data). From Ray, et al. *(59)*

Lineage	Outbreaks	Predominant Serotypes
I	Epidemic clones and responsible for most human listeriosis outbreaks	1/2b, 3b, 4b, 4d, 4e, 7
II	Sporadic listeriosis cases	1/2a, 3a, 1/2c, 3c
III and IV	Rarely cause human listeriosis	4a, 4c

Characterization, Phylogenetic, and Subtyping Methods

L. monocytogenes can be characterized below the species level through genetic or serotyping methods. Typing is necessary for epidemiological studies and outbreak investigations *(29)*. Serovar designations for *L. monocytogenes* are based on the immunoreactivity of the O (somatic) and H (flagellar) antigens, two surface cell structures *(10, 58, 64)*. *L. monocytogenes* strains can be divided into at least 13 serotypes *(26, 36, 47, 59, 62, 73)*: 1/2a, 3a, 1/2c, 3c, 4b, 4d, 4e, 1/2b, 3b, 7, 4a, 4ab and 4c. All thirteen serovars can cause disease, but most (>90%) isolates from humans with listeriosis are 1/2a, 1/2b, and 4b *(47, 62)*, with 1/2a and 1/2b being the most commonly isolated from human listeriosis patients in Europe and 4b most common in the

United States and Canada *(59)*. Serovar 4b strains cause an estimated 33-50% sporadic listeriosis cases worldwide *(62)*. In contrast, 1/2a and 1/2c are the most common serovars isolated from foods *(62)*. Three serovars (1/2a, 1/2b, and 4b) cause most (95%) human listeriosis and serovar 4b is most commonly associated with outbreaks *(10)*. *L. monocytogenes* PCR-serogroups can be deduced *in silico* from WGS data using BIGSdb-Lm platform *(28, 50)*.

For many years, Pulsed-Field Gel Electrophoresis (PFGE) has been the primary method of molecular typing for *L. monocytogenes (7, 21, 29, 32)*. PFGE is difficult to standardize and provides limited information about phylogenetic relationships among isolates *(58)*. PFGE is no longer the main method used for outbreak investigations and listeriosis surveillance and is being replaced by WGS-based methods.

Prior to WGS being widely available and affordable, typing based on single genes or gene clusters (single locus and multilocus sequence typing) was common. Some commonly used genes included *ActA (4, 5)*, the *prfA* virulence gene cluster *(76)*, and *inlA (58)*.

Multilocus sequence typing (MLST) is based on sequencing and analyzing seven housekeeping genes: *acbZ* (ABC transporter), *bglA* (beta-glucosidase), *cat* (catalase), *dapE* (succinyl diaminopimelate desuccinylase), *dat* (D-amino acid aminotransferase), *ldh* (lactate deshydrogenase), and *lhkA* (histidine kinase) *(44, 58, 63)***.** For each MLST locus, an allele number is given to each distinct sequence variant**.** A sequence type (ST) is given to each distinct combination of alleles for all of the MLST loci. MLST typing can be performed *in silico* using the BIGSdb-Lm platform *(28, 50, 57)*.

A variation of MLST, Multi-Virulence-Locus Sequence Typing (MVLST), is also available to type *L. monocytogenes* based on virulence genes. MVLST is performed similarly to MLST typing and the scheme is based on three virulence genes (*prfA*, *inlB*, and *inlC*) and three

virulence-associated genes (*dal, lisR,* and *clpP*) *(80)*. Like with MLST, an allele number is given to each distinct sequence variant and a sequence type (ST) is given to each distinct combination of alleles for all of the MVLST loci.

Whole-Genome Sequencing Data Analysis. Next-generation sequencing allows for multiple types of analyses of bacterial whole-genome sequencing (WGS) data. These types of analyses typically can fall into two categories: single nucleotide polymorphism (SNP) analysis and gene-based analyses of whole genomes (like whole-genome multi locus sequence typing [wgMLST]) *(31, 66)*. Initial steps include DNA extraction and isolation, library preparation and quality control, and sequencing, which can be done on platforms like Illumina, PacificBiosciences (PacBio), MinIon sequencers. Once sequencing data is obtained, the next steps include sequencing quality control, optional genome assembly, variant calling and subtyping based on variants *(66)*.

In SNP-based methods, single nucleotide changes are used to infer phylogenetic relatedness between isolates *(31, 58)*. The steps in SNP-based methods are typically as follows: mapping raw reads onto a reference genome, identifying SNPs, removing lower-quality SNPs, creating a multiple sequence alignment (MSA) from selected SNPs, and inferring phylogeny from the MSA *(31)*.

SNP-calling can be done with or without a reference genome.

In addition, SNP-calling methods come in two main categories: basic SNP-based methods and high-quality SNP-based (hqSNP) methods. In hqSNP-based methods, SNPs are filtered by sequence depth and quality and those with less support (i.e., having few mapped raw reads, having low consensus in the raw reads, being located in high mutation regions like phage regions, etc.) are removed from analysis *(25, 31)*. hqSNP analysis offers a high level of detail,

but can be computationally intensive for large datasets (i.e., may require high computing power and is time consuming) and may require a skilled bioinformatician *(25)*. There are currently two publically available hqSNP pipelines available: the FDA's Center for Food Safety and Applied Nutrition (CFSAN) Single Nucleotide Polymorphism (SNP) Pipeline *(15, 68)* and Lyve-SET (the LYVE version of the SNP Extraction Tool [SET]) *(30, 31)*. A proprietary hqSNP pipeline is also currently available through BioNumerics.

The wgMLST typing scheme is based on 4804 loci. For each genome, all loci are compared against a database of known loci *(31)*. For each loci, if it matches a known allele, it is assigned an identifier, and if it does not match a known allele, it is given a new identifier. All loci between multiple genomes can be compared and a distance calculated. These distances can be used to infer phylogeny *(31)*.

Some other analyses that can be performed on WGS data are Core-genome Multilocus Sequence Typing (cgMLST), Extended Multilocus Sequence Typing (eMLST), virulence and resistance gene profiles, phage analysis, and mobile genetic element (MGE) profiling. The cgMLST scheme is based on 1,791 loci in the core genome and is performed similarly to wgMLST *(50)*. It can be performed using the BIGSdb-Lm platform *(57)*. Similarly, eMLST is based on whole core genome (which would include all genes present in all isolates of a species) *(29)*. Virulence and resistance gene profiles can be constructed based on the presence of 76 loci involved in virulence or resistance using BIGSdb-Lm platform *(28, 50)*. MGEs, such as phages, can be identified and characterized in genomes using programs like Phaster *(2, 81)*.

Virulence and Pathogenicity Factors

L. monocytogenes has many virulence factors that aid in pathogenesis. Most virulence genes are found on the pathogenicity island (PAI), a 9.0 kb region of the chromosome, and are

regulated by two regulatory factors (protein regulatory factor [PrfA] and sigmaB) *(59)*. Expression of virulence genes is differentially regulated by available carbon sources, temperature, bile salts, salt, and the presence of acidic and anaerobic environments *(59)*. Listeriolysin O (LLO) is a beta-hemolysin (causes beta-hemolysis of erythrocytes and destroys phagocytic cells that engulf them) and is typically produced by pathogenic/virulent strains of *L. monocytogenes*. It is encoded by the chromosomal gene *hly*.

L. monocytogenes is capable of intracellular invasion and can invade tissues, including the placenta in pregnant women, and can enter the blood stream and can then invade other susceptible cells. Once inside a cell, *L. monocytogenes* may replicate. There are two steps of cellular entry: directly into phagosomes and then from the phagosomes into the cytoplasm of the phagocyte. Non-phagocyte cell entry requires surface-bound proteins, including Internalin (In1A; for which the mammalian surface receptor is *E-cadherin*; required for entry into epithelial cells), In1B (required for entry into hepatocytes as seen in mouse models), p60 (encoded by *iap* gene), ActA (required for actin polymerization and intracytoplasmic movement of cells; encoded by *actA* gene), and Ami (a bacteriolysin).

Facilitated by LLO, *L. monocytogenes* cells escape the phagolysosomal membranes into the cytoplasm (cytosol). From there, ActA is involved in the formation of actin tails that move the cell toward the cytoplasmic membrane. The cell can then move into neighboring cells by pushing the membrane out into a projection called a filopodium, which is then absorbed by the neighboring cell, where double membrane vacuoles form around the bacteria. LLO and two bacterial phospholipases (phosphatidylinositol-specific phospholipase C [encoded by *plcA*] and the broad-range phospholipase C [encoded by *plcB*]) facilitate escape from the vacuoles. The entire process can be continually repeated in adjacent cells.

Table 2.3. Major Virulence Proteins in *Listeria monocytogenes* (adapted from Ray et al. *(59) (80)*)

Virulence Factor	Function
Protein regulatory Factor (PrfA)	Listeriolysin positive regulatory protein; Regulation of virulence protein expression
Internalin (InlA)	Responsible for invasion into intestinal epithelial cells and placenta during pregnancy
Internalin B (InlB)	Entry into hepatocytes and hepatic phase of infection
Virulence invasion protein (Vip)	Invasion of epithelial cells
Listeria adhesion protein (LAP)	Adhesion to intestinal epithelial cells
Listeria adhesion protein (LapB)	Adhesion and invasion of epithelial cells
Autolysin amidase	Adhesion to host cells
Listeriolysin (LLO)	A hemolysin responsible for lysis of red blood cells. Aids in bacterial escape from vacuole inside the cell.
Actin polymerization protein (ActA)	Nucleation of actin tail for bacterial movement inside the cytoplasm
Bile salt hydrolase (BSH)	Survival in gut
Phosphopilase C (PLC)	Lysis of vacuole membrane
Metalloprotease	Helps synthesis of PLC

fibronectin binding protein (Fbp)	Adhesion to intestinal epithelial cells
amidase	Adhesion to intestinal epithelial cells
Alanine racemase (*dal*)	
Two-component response regulator (*lisR*)	
clpP	Clp proteolytic subunit

Persistence and Related Phenotypes

Once inside facilities, *L. monocytogenes* may contaminate food products and/or establish a niche and become colonized in the food processing environment. If allowed to colonize in the plant environment, *L. monocytogenes* may attach to surfaces and/or form biofilms, which make them more resistant to removal through cleaning and sanitation. Some common niches for *L. monocytogenes* are moist and contain nutrients, most often floor drains, condensate, standing/stagnant water, floors, and areas of processing equipment with organic residue *(42, 59, 62)*. On equipment, small parts, unreachable portions, and some types of surfaces may not be effectively cleaned and sanitized, and can subsequently serve as harborage sites of *L. monocytogenes (42, 59)*

Repeatedly isolated *L. monocytogenes* strains may indicate persistence in the processing facility (i.e., the strain has created a niche in the plant) or repeated introduction (i.e., the strain may be frequently re-introduced into the plant from sources such as raw material or the surrounding environment). Theoretically, strains isolated after cleaning and sanitation are more

likely to be persistent strains than strains isolated during production (before cleaning and sanitation).

It is of interest to the food industry to determine which, if any, phenotypic characteristics of *L. monocytogenes* aid in allowing strains to establish a niche and persist in food processing environments. Some hypothetical phenotypic characteristics that may correlate with persistence are biofilm formation, attachment to surfaces, stress resistance (heat, drying and desiccation, acid, etc.), disinfectant and sanitizer resistance, and slower growth rates *(20, 24, 35, 79)*. Attached cell and biofilms could potentially aid in persistence because they are more difficult to remove by mechanical cleaning and sanitation and because of their increased resistance to environmental stress, including disinfectants, low temperatures, decreased nutrients, and increased salt concentrations *(43)*.

Some research has shown that persistent strains of *L. monocytogenes* have phenotypic characteristic that might aid in their ability to establish and maintain persistence in food processing environments *(1, 6, 27, 43, 53)*. In contrast, other research has shown the opposite, that persistent strains do not significantly differ from nonpersistent strains in phenotypic characteristics that may be associated with persistence *(16, 17, 23, 24, 27, 35)*.

Korenova et al. (*33*) showed no correlation between biofilm formation and *L. monocytogenes* persistence. They tested biofilm forming ability of 33 different *L. monocytogenes* strains (originally isolated from various food samples and food processing environments) under various conditions (including exponential and stationary growth phases, various temperatures, under limited nutrient conditions, and under high-salt conditions) on polystyrene plates. Biofilm formation was quantified by crystal violate staining. They found that biofilm formation was variable for different strains under the various conditions tested and that

no statistically significant differences in biofilm formation by persistent and sporadic strains under the tested conditions could be found *(35)*.

Lunden, et al. *(43)* measured the adherence after various contact times (1, 2 and 72 h) of persistent (n=3) and non-persistent (n=14) *L. monocytogenes* strains isolated from poultry plants and an ice cream plant. They found that persistent strains from poultry processing plants had 2- to 11-fold greater adherence than nonpersistent strains after 1 and 2 h contact times. The adherence of a persistent strain from a ice cream plant was also higher than most nonpersistent strains after 1 and 2 h contact times. In contrast, after 72 h contact time, the differences between adherence of persistent and nonpersistent strains were not as notable and three nonpersistent strains even displayed greater adherence than persistent strains. The authors concluded that the persistent strains they examined exhibited greater adherence at shorter contact times when compared to nonpersistent strains. They theorized that this difference may promote survival of persistent *L. monocytogenes* strains in food processing plants *(43)*.

Fagerlund, et al. *(19)* sequenced and analyzed five MLST sequence type-8 *L. monocytogenes* isolates from Norwegian salmon and poultry processing facilities. That study was narrow, as it only focused on one MLST sequence type and included only five isolates. The researchers concluded that WGS and SNP-based analysis of *L*. monocytogenes are well suited to evaluate persistence of *L*. monocytogenes isolates in food processing facilities. In addition, they compared the accessory genome of one of the ST-8 strains to two commonly used laboratory strains (EGD-e and 10403S) and found unique sections that contained of 50 ORFs, distributed in ten genomic loci. These loci contained two R-M systems (Type I and III) and encoded proteins with the following predicted functions: abortive resistance to bacteriophage infection, putative ATP-binding cassette, carbohydrate transport, and metallopeptidase.

Stasiewicz, et al. analyzed *L. monocytogenes* isolates from deli environments to identify genes significantly enriched among persistent or sporadic groups *(67)*. Due to the fact that the *L. monocytogenes* isolate sample size for that study was large, it has more power than many other studies that have been designed to identify genes associated with *L. monocytogenes* persistence, as many have a very small sample size. Gene enrichment analysis failed to uncover any genes that were enriched among isolates initially (putatively) identified as either persistent (n=92) or sporadic (n =29) or when comparing isolates with statistical evidence for persistence (n=38) to sporadic isolates of the same subtypes (n=16). They did find 10 genes that were significantly enriched among nonpersistent isolates; however, these genes were not present in a conserved region and were not annotated with functions associated with persistence or survival in a harsh environment. Stasiewicz, et al. concluded that there are no individual gene presence or absence patterns linked to the persistent strains analyzed in their study *(67)*.

Nowak, et al. evaluated persistent *L. monocytogenes* strains isolated from New Zealand mussel production facilities to find specific genes or genetic markers that may be linked to persistence or non-persistence *(54)*. They were unable to find any markers in either persistent or non-persistent isolate groups. In addition, they found that there was no overall clustering of persistent or sporadic isolates and that differences in prophages and plasmids were not associated with persistence. One limitation of this study is the small number of *L. monocytogenes* isolates evaluated (8 persistent and 8 sporadic).

References

1. Aase, B., G. Sundheim, S. Langsrud, and L. M. Rorvik. 2000. Occurrence of and a possible mechanism for resistance to a quaternary ammonium compound in *Listeria monocytogenes. Int. J. Food Microbiol.* 62:57-63.

2. Arndt, D., J. R. Grant, A. Marcu, T. Sajed, A. Pon, Y. Liang, and D. S. Wishart. 2016. Phaster: A better, faster version of the phast phage search tool. *Nucleic Acids Res.* 44:W16-W21.

3. Autio, T., R. Keto-Timonen, J. Lunden, J. Bjorkroth, and H. Korkeala. 2003. Characterisation of persistent and sporadic *Listeria monocytogenes* strains by pulsed-field gel electrophoresis (PFGE) and amplified fragment length polymorphism (AFLP). *Syst. Appl. Microbiol.* 26:539-545.

4. Berrang, M. E., R. J. Meinersmann, J. F. Frank, and S. R. Ladely. 2010. Colonization of a newly constructed commercial chicken further processing plant with *Listeria monocytogenes. J. Food Prot.* 73:286-291.

5. Berrang, M. E., R. J. Meinersmann, J. F. Frank, D. P. Smith, and L. L. Genzlinger. 2005. Distribution of *Listeria monocytogenes* subtypes within a poultry further processing plant. *J. Food Prot.* 68:980-985.

6. Borucki, M. K., J. D. Peppin, D. White, F. Loge, and D. R. Call. 2003. Variation in biofilm formation among strains of *Listeria monocytogenes. Appl. Environ. Microbiol.* 69:7336-7342.

7. Brosch, R., J. Chen, and J. B. Luchansky. 1994. Pulsed-field fingerprinting of listeriae: Identification of genomic divisions for *Listeria monocytogenes* and their correlation with serovar. *Appl. Environ. Microbiol.* 60:2584-92.

8. Carpentier, B., and O. Cerf. 2011. Review--persistence of *Listeria monocytogenes* in food industry equipment and premises. *Int. J. Food Microbiol.* 145:1-8.

9. Centers for Disease Control and Prevention (CDC). *Listeria* and whole genome sequencing: Whole genome sequencing improves the detection and investigation of foodborne outbreaks. Available at: https://www.cdc.gov/listeria/pdf/whole-genome-sequencing-and-listeria-508c.pdf. Accessed 30 August 2017.

10. Centers for Disease Control and Prevention (CDC). 2016. National enteric disease surveillance: The *Listeria* initiative. Available at: https://www.cdc.gov/listeria/pdf/listeriainitiativeoverview_508.pdf. Accessed 30 August 2017.

11. Centers for Disease Control and Prevention (CDC). 2017. Foodborne diseases active surveillance network (foodnet): Foodnet 2015 surveillance report (final data). Available at: https://www.cdc.gov/foodnet/reports/annual-reports-2015.html. Accessed 1 April 2018.

12. Centers for Disease Control and Prevention (CDC). 2017. *Listeria* (listeriosis): The *Listeria* initiative. Available at: https://www.cdc.gov/listeria/surveillance/listeria-initiative.html. Accessed 30 August 2017.

13. Centers for Disease Control and Prevention (CDC). 2017. *Listeria* (listeriosis): The *Listeria* whole genome sequencing project. Available at: https://www.cdc.gov/listeria/surveillance/whole-genome-sequencing.html. Accessed 1 September 2017.

14. Cruz, C. D., and G. C. Fletcher. 2011. Prevalence and biofilm-forming ability of *Listeria monocytogenes* in New Zealand mussel (*perna canaliculus*) processing plants. *Food Microbiol.* 28:1387-1393.

15. Davis, S., J. B. Pettengill, Y. Luo, J. Payne, A. Shpuntoff, H. Rand, and E. Strain. 2015. CFSAN SNP pipeline: An automated method for constructing SNP matrices from next-generation sequence data. *PeerJ Computer Science*. 1:e20.

16. Djordjevic, D., M. Wiedmann, and L. A. McLandsborough. 2002. Microtiter plate assay for assessment of *Listeria monocytogenes* biofilm formation. *Appl. Environ. Microbiol.* 68:2950-2958.

17. Earnshaw, A., and L. Lawrence. 1998. Sensitivity to commercial disinfectants, and the occurrence of plasmids within various *Listeria monocytogenes* genotypes isolated from poultry products and the poultry processing environment. *J. Appl. Microbiol.* 84:642-648.

18. Euzeby, J. P. 1997. List of bacterial names with standing in nomenclature: A folder available on the internet. *Int. J. Syst. Bacteriol.* 47:590-592.

19. Fagerlund, A., S. Langsrud, B. C. T. Schirmer, T. Moretro, and E. Heir. 2016. Genome analysis of *Listeria monocytogenes* sequence type 8 strains persisting in salmon and

poultry processing environments and comparison with related strains. *PLoS One*. 11:e0151117.

20. Ferreira, V., M. Wiedmann, P. Teixeira, and M. J. Stasiewicz. 2014. *Listeria monocytogenes* persistence in food-associated environments: Epidemiology, strain characteristics, and implications for public health. *J. Food Prot.* 77:150-170.

21. Graves, L. M., and B. Swaminathan. 2001. Pulsenet standardized protocol for subtyping *Listeria monocytogenes* by macrorestriction and pulsed-field gel electrophoresis. *Int. J. Food Microbiol.* 65:55-62.

22. Harvey, J., and A. Gilmour. 1994. Application of multilocus enzyme electrophoresis and restriction fragment length polymorphism analysis to the typing of *Listeria monocytogenes* strains isolated from raw milk, nondairy foods, and clinical and veterinary sources. *Appl. Environ. Microbiol.* 60:1547-1553.

23. Heir, E., B.-A. Lindstedt, O.-J. Røtterud, T. Vardund, G. Kapperud, and T. Nesbakken. 2004. Molecular epidemiology and disinfectant susceptibility of *Listeria monocytogenes* from meat processing plants and human infections. *Int. J. Food Microbiol.* 96:85-96.

24. Holah, J., J. Taylor, D. Dawson, and K. Hall. 2002. Biocide use in the food industry and the disinfectant resistance of persistent strains of *Listeria monocytogenes* and *Escherichia coli. J. Appl. Microbiol.* 92:111S-120S.

25. Jackson, B. R., C. Tarr, E. Strain, K. A. Jackson, A. Conrad, H. Carleton, L. S. Katz, S. Stroika, L. H. Gould, and R. K. Mody. 2016. Implementation of nationwide real-time whole-genome sequencing to enhance listeriosis outbreak detection and investigation. *Clin. Infect. Dis.* 63:380-386.

26. Jay, J. M. 2005. Foodborne listeriosis. *In* J.M. Jay, D.A. Golden, and M.J. Loessner (ed.), Modern food microbiology. 7th ed. Springer, New York

27. Jensen, A., M. H. Larsen, H. Ingmer, B. F. Vogel, and L. Gram. 2007. Sodium chloride enhances adherence and aggregation and strain variation influences invasiveness of *Listeria monocytogenes* strains. *J. Food Prot.* 70:592-599.

28. Jolley, K. A., and M. C. J. Maiden. 2010. Bigsdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*. 11:595.

29. Jordan, K., D. Leong, and A. Á. Ordóñez. 2015. *Listeria monocytogenes* in the food processing environment. Springer, Cham.

30. Katz L.S., Griswold T., Williams-Newkirk A.J., Wagner D., Petkau A., and et al. Lyve-set. Available at: https://github.com/lskatz/lyve-SET. Accessed 20 February 2018.

31. Katz, L. S., T. Griswold, A. J. Williams-Newkirk, D. Wagner, A. Petkau, C. Sieffert, G. Van Domselaar, X. Deng, and H. A. Carleton. 2017. A comparative analysis of the lyve-set phylogenomics pipeline for genomic epidemiology of foodborne pathogens. *Front. Microbiol.* 8:375.

32. Kerouanton, A., A. Brisabois, E. Denoyer, F. Dilasser, J. Grout, G. Salvat, and B. Picard. 1998. Comparison of five typing methods for the epidemiological study of *Listeria monocytogenes. Int. J. Food Microbiol.* 43:61-71.

33. Keto-Timonen, R., R. Tolvanen, J. Lunden, and H. Korkeala. 2007. An 8-year surveillance of the diversity and persistence of *Listeria monocytogenes* in a chilled food processing plant analyzed by amplified fragment length polymorphism. *J. Food Prot.* 70:1866-1873.

34. Knudsen, G. M., J. B. Nielsen, R. L. Marvig, Y. Ng, P. Worning, H. Westh, and L. Gram. 2017. Genome-wide-analyses of *Listeria monocytogenes* from food-processing plants reveal clonal diversity and date the emergence of persisting sequence types. *Environ. Microbiol. Rep.* 9:428-440.

35. Korenova, J., K. Oravcova, A. Veghova, R. Karpiskova, and T. Kuchta. 2016. Biofilm formation in various conditions is not a key factor of persistence potential of *Listeria monocytogenes* in food-processing environment. *J. Food Nutr. Res.* 55:189-193.

36. Kuenne, C., A. Billion, M. A. Mraheil, A. Strittmatter, R. Daniel, A. Goesmann, S. Barbuddhe, T. Hain, and T. Chakraborty. 2013. 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.

37. Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures. 2017. Prokaryotic nomenclature up-to-date: *Listeria*. Available at: https://www.dsmz.de/bacterial-diversity/prokaryotic-nomenclature-up-to-date/prokaryotic-nomenclature-up-to-date.html. Accessed 26 March 2018.

38. Leong, D., K. NicAogáin, L. Luque-Sastre, O. McManamon, K. Hunt, A. Alvarez-Ordóñez, J. Scollard, A. Schmalenberger, S. Fanning, C. O'Byrne, and K. Jordan. 2017. A 3-year multi-food study of the presence and persistence of *Listeria monocytogenes* in 54 small food businesses in Ireland. *Int. J. Food Microbiol.* 249:18-26.

39. List of prokaryotic names with standing in nomenclature (LSPN). 2018. List of prokaryotic names with standing in nomenclature (lspn). Available at: http://www.bacterio.net/-allnames.html. Accessed 26 March 2018.

40. Lunden, J., T. Autio, A. Markkula, S. Hellstrom, and H. Korkeala. 2003. Adaptive and cross-adaptive responses of persistent and non-persistent *Listeria monocytogenes* strains to disinfectants. *Int. J. Food Microbiol.* 82:265-272.

41. Lunden, J., R. Tolvanen, and H. Korkeala. 2008. Acid and heat tolerance of persistent and nonpersistent *Listeria monocytogenes* food plant strains. *Lett. Appl. Microbiol.* 46:276-280.

42. Lunden, J. M., T. J. Autio, A. M. Sjoberg, and H. J. Korkeala. 2003. Persistent and nonpersistent *Listeria monocytogenes* contamination in meat and poultry processing plants. *J. Food Prot.* 66:2062-2069.

43. Lunden, J. M., M. K. Miettinen, T. J. Autio, and H. J. Korkeala. 2000. Persistent *Listeria monocytogenes* strains show enhanced adherence to food contact surface after short contact times. *J. Food Prot.* 63:1204-1207.

44. Maiden, M. C. J., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt. 1998. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences*. 95:3140-3145.

45. Malley, T. J. V., M. J. Stasiewicz, Y. T. Grohn, S. Roof, S. Warchocki, K. Nightingale, and M. Wiedmann. 2013. Implementation of statistical tools to support identification and management of persistent *Listeria monocytogenes* contamination in smoked fish processing plants. *J. Food Prot.* 76:796-811.

46. McLauchlin, J., and C. E. D. Rees. 2015. *Listeria. In* F.R. W. B. Whitman, P. Kämpfer, M. Trujillo, J. Chun, P. DeVos, B. Hedlund and S. Dedysh (ed.), Bergey's manual of systematics of archaea and bacteriaed.

47. Montville, T. J., K. R. Matthews, and K. E. Kniel. 2012. Chapter 13: *Listeria monocytogenes.* p. 189-204. *In* T.J. Montville, K.R. Matthews, and K.E. Kniel (ed.), Food microbiology: An introduction 3rd ed. ASM Press, Washington; USA.

48. Møretrø, T., and S. Langsrud. 2004. *Listeria monocytogenes*: Biofilm formation and persistence in food-processing environments. *Biofilms*. 1:107-121.

49. Morganti, M., E. Scaltriti, P. Cozzolino, L. Bolzoni, G. Casadei, M. Pierantoni, E. Foni, and S. Pongolini. 2016. Processing-dependent and clonal contamination patterns of *Listeria monocytogenes* in the cured ham food chain revealed by genetic analysis. *Appl. Environ. Microbiol.* 82:822-831.

50. Moura, A., A. Criscuolo, H. Pouseele, M. M. Maury, A. Leclercq, C. Tarr, J. T. Björkman, T. Dallman, A. Reimer, and V. Enouf. 2016. Whole genome-based population biology and epidemiological surveillance of *Listeria monocytogenes. Nature Microbiology*. 2:16185.

51. Murugesan, L., Z. Kucerova, S. J. Knabel, and L. F. LaBorde. 2015. Predominance and distribution of a persistent *Listeria monocytogenes* clone in a commercial fresh mushroom processing environment. *J. Food Prot.* 78:1988-1998.

52. National Center for Biotechnology Information (NCBI). Taxonomy browser: *Listeria*. Available at: https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=1637. Accessed 26 March 2018.

53. Norwood, D. E., and A. Gilmour. 1999. Adherence of *Listeria monocytogenes* strains to stainless steel coupons. *J. Appl. Microbiol.* 86:576-582.

54. Nowak, J., C. D. Cruz, M. Tempelaars, T. Abee, A. H. M. van Vliet, G. C. Fletcher, D. Hedderley, J. Palmer, and S. Flint. 2017. Persistent *Listeria monocytogenes* strains isolated from mussel production facilities form more biofilm but are not linked to specific genetic markers. *Int. J. Food Microbiol.* 256:45-53.

55. Palma, F., F. Pasquali, A. Lucchi, A. De Cesare, and G. Manfreda. 2017. Whole genome sequencing for typing and characterisation of *Listeria monocytogenes* isolated in a rabbit meat processing plant. *Ital. J. Food Saf.* 6:6879.

56. Parte, A. C. 2014. Lpsn-list of prokaryotic names with standing in nomenclature. *Nucleic Acids Res.* 42:D613-D616.

57. Pasteur Institute. Bigsdb-lm: *Listeria* locus/sequence definitions database. Available at: http://bigsdb.pasteur.fr/perl/bigsdb/bigsdb.pl?db=pubmlst_listeria_seqdef_public. Accessed 26 March 2018.

58. Ragon, M., T. Wirth, F. Hollandt, R. Lavenir, M. Lecuit, A. Le Monnier, and S. Brisse. 2008. A new perspective on *Listeria monocytogenes* evolution. *PLoS Path.* 4:e1000146.

59. Ray, B., and A. Bhunia. 2014. Fundamental food microbiology CRC Press, Boca Raton, FL.

60. Rorvik, L. M., B. Aase, T. Alvestad, and D. A. Caugant. 2003. Molecular epidemiological survey of *Listeria monocytogenes* in broilers and poultry products. *J. Appl. Microbiol.* 94:633-640.

61. Rørvik, L. M., D. A. Caugant, and M. Yndestad. 1995. Contamination pattern of *Listeria monocytogenes* and other *Listeria* spp. In a salmon slaughterhouse and smoked salmon processing plant. *Int. J. Food Microbiol.* 25:19-27.

62. Ryser, E. T., and R. L. Buchanan. 2013. *Listeria monocytogenes.* p. 503-545. *In* M.P. Doyle, and R.L. Buchanan (ed.), Food microbiology: Fundamentals and frontiers 4th ed. ASM Press, Washington, D.C.

63. Salcedo, C., L. Arreaza, B. Alcala, L. de la Fuente, and J. A. Vazquez. 2003. Development of a multilocus sequence typing method for analysis of *Listeria monocytogenes* clones. *J. Clin. Microbiol.* 41:757-762.

64. Seeliger, H. P. R., and K. Höhne. 1979. Chapter ii serotyping of *Listeria monocytogenes* and related species. p. 31-49. *In* T. Bergan, and J.R. Norris (ed.), Methods in microbiology, vol. 13.ed. Academic Press.

65. Senczek, D., R. Stephan, and F. Untermann. 2000. Pulsed-field gel electrophoresis (PFGE) typing of *Listeria* strains isolated from a meat processing plant over a 2-year period. *Int. J. Food Microbiol.* 62:155-159.

66. Stasiewicz, M. J., H. C. den Bakker, and M. Wiedmann. 2015. Genomics tools in microbial food safety. *Curr. Opin. Food Sci.* 4:105-110.

67. Stasiewicz, M. J., H. F. Oliver, M. Wiedmann, and H. C. den Bakker. 2015. Whole-genome sequencing allows for improved identification of persistent *Listeria*

monocytogenes in food-associated environments. *Appl. Environ. Microbiol.* 81:6024-6037.

68. Strain, E., Y. Luo, J. Pettengill, H. A. Rand, S. Davis, J. Payne, A. Shpuntoff, J. D. Baugher, and Y. Wang. 2017. CFSAN SNP pipeline. Available at: http://snp-pipeline.readthedocs.io/en/latest/index.html. Accessed 2 April 2018.

69. Sumner, S. S., T. M. Sandros, M. C. Harmon, V. N. Scott, and D. T. Bernard. 1991. Heat resistance of *Salmonella* Typhimurium and *Listeria monocytogenes* in sucrose solutions of various water activities. *J. Food Sci.* 56:1741-1743.

70. Tompkin, R. B. 2002. Control of *Listeria monocytogenes* in the food-processing environment. *J. Food Prot.* 65:709-725.

71. Vangay, P., J. Steingrimsson, M. Wiedmann, and M. J. Stasiewicz. 2014. Classification of *Listeria monocytogenes* persistence in retail delicatessen environments using expert elicitation and machine learning. *Risk Anal.* 34:1830-1845.

72. Vasseur, C., L. Baverel, M. Hebraud, and J. Labadie. 1999. Effect of osmotic, alkaline, acid or thermal stresses on the growth and inhibition of *Listeria monocytogenes. J. Appl. Microbiol.* 86:469-476.

73. Vázquez-Boland, J. A., M. Kuhn, P. Berche, T. Chakraborty, G. Domínguez-Bernal, W. Goebel, B. González-Zorn, J. Wehland, and J. Kreft. 2001. *Listeria* pathogenesis and molecular virulence determinants. *Clin. Microbiol. Rev.* 14:584-640.

74. Verghese, B., M. Lok, J. Wen, V. Alessandria, Y. Chen, S. Kathariou, and S. Knabel. 2011. Comk prophage junction fragments as markers for *Listeria monocytogenes* genotypes unique to individual meat and poultry processing plants and a model for rapid niche-specific adaptation, biofilm formation, and persistence. *Appl. Environ. Microbiol.* 77:3279-3292.

75. Wang, J. J., A. J. Ray, S. R. Hammons, and H. F. Oliver. 2015. Persistent and transient *Listeria monocytogenes* strains from retail deli environments vary in their ability to adhere and form biofilms and rarely have inla premature stop codons. *Foodborne Pathog. Dis.* 12:151-158.

76. Ward, T. J., L. Gorski, M. K. Borucki, R. E. Mandrell, J. Hutchins, and K. Pupedis. 2004. Intraspecific phylogeny and lineage group identification based on the *prfa* virulence gene cluster of *Listeria monocytogenes. J. Bacteriol.* 186:4994–5002.

77. Wiedmann, M., J. L. Bruce, C. Keating, A. E. Johnson, P. L. McDonough, and C. A. Batt. 1997. Ribotypes and virulence gene polymorphisms suggest three distinct *Listeria monocytogenes* lineages with differences in pathogenic potential. *Infect. Immun.* 65:2707-2716.

78. Williams, S. K., S. Roof, E. A. Boyle, D. Burson, H. Thippareddi, I. Georgnaras, J. N. Sofos, M. Wiedmann, and K. Nightingale. 2011. Molecular ecology of *Listeria monocytogenes* and other *Listeria* species in small and very small ready-to-eat meat processing plants. *J. Food Prot.* 74:63-77.

79. Wulff, G., L. Gram, P. Ahrens, and B. F. Vogel. 2006. One group of genetically similar *Listeria monocytogenes* strains frequently dominates and persists in several fish slaughter-and smokehouses. *Appl. Environ. Microbiol.* 72:4313-4322.

80. Zhang, W., B. M. Jayarao, and S. J. Knabel. 2004. Multi-virulence-locus sequence typing of *Listeria monocytogenes. Appl. Environ. Microbiol.* 70:913-920.

81. Zhou, Y., Y. Liang, K. H. Lynch, J. J. Dennis, and D. S. Wishart. 2011. Phast: A fast phage search tool. *Nucleic Acids Res.* 39:W347-W352.

CHAPTER 3

GENOME-WIDE ANALYSIS AND CHARACTERIZATION OF *LISTERIA MONOCYTOGENES* ISOLATES FOUND IN TWO POULTRY FURTHER PROCESSING PLANTS[1]

[1]Lauren K. Hudson, Shaokang Zhang, Mark E. Berrang, Richard J. Meinersmann, Xiangyu Deng, and Mark A. Harrison. To be submitted to *Journal of Food Protection*.

Abstract

Genomic analysis of *Listeria monocytogenes* isolates from food processing facilities can yield insights into characteristics such as phylogeny and evolution, virulence, and antibiotic-, stress-, and sanitizer-resistance. These insights may lead to better ways to target and inactivate *L. monocytogenes* in processing environments. The objective of this study was to use whole genome sequencing (WGS) and subsequent genomic analyses to characterize *L. monocytogenes* isolates recovered in two previous longitudinal studies of *L. monocytogenes* in poultry further processing plants. In the present study, 156 genomes were extracted, sequenced on the Illumina MiSeq platform, *de novo* assembled, and annotated. Isolates were divided into lineages, MLST-typed, serotyped, and screened for acquired antibiotic-, stress-, metal-, and sanitizer-resistance genes. Fifty-six isolates belonged to lineage I and 99 belonged to lineage II. Eighteen unique MLST sequence types (ST) were found, the majority of which were identified as ST321 (n=41), ST5 (n=31), ST155 (n=27), and ST6 (n=20). All of the genomes contained *fosX*, *lmo0441*, *lmo0919*, *norB*, and *sul* antibiotic-resistance (ABR) genes, 14.1% (n=22) contained *aacA4* genes, and 3.2% (n=5) contained *tetM* ABR genes. Of the genomes, 82.7% (n=129) had 5 total ABR genes , 16% (n=25) had 6 total, and 1.3% (n=2) had 7 total. Of isolates screened, 82.7% (n=129) possessed the following genes related to stress-resistance: *lmo0444*, *lmo0445*, *lmo0446*, *lmo0447*, and *lmo0448*, all contained in a stress survival islet (SSI-1). The *bcrABC* genes related to benzalkonium chloride resistance were in 73.1% (n=114) of the isolates. This data provides insight about *L. monocytogenes* found in poultry further processing plants, including potentially high rates of ABR, stress-resistance, and sanitizer-resistance genes located in the genomes.

Listeria monocytogenes is a major concern in the food industry and is often found in poultry processing plants and on broilers and poultry products *(2, 6-10, 13, 24, 39)*. It has been the cause of poultry product recalls *(33, 46, 47)*, and listeriosis cases and outbreaks have been linked to poultry products *(11, 12, 17, 27, 33, 35)*.

L. monocytogenes may enter these types of processing plants through many routes, including via soil on items like workers' clothes or shoes, transport equipment, contaminated machinery, animals/pests that either excrete the bacterium or have contaminated external surfaces, raw plant material, raw meat foods, or possibly even asymptomatic human carriers *(40)*. Once inside these facilities, *L. monocytogenes* may contaminate food products and/or establish a niche and colonize in the food processing environment. If allowed to remain in the environment, *L. monocytogenes* may attach to surfaces and/or form biofilms, which make them more resistant to removal through cleaning and sanitation. Niches that commonly harbor *L. monocytogenes* are moist and contain nutrients, such as floor drains, condensate, standing/stagnant water, floors, and areas of processing equipment with organic residue *(24, 38, 40)*. Small parts, unreachable areas, and some types of surfaces may not be effectively cleaned and sanitized, and can subsequently serve as harborage sites of *L. monocytogenes (24, 38)*. It has been suggested that repeated isolation of *L. monocytogenes* strains may indicate persistence in the processing plant (i.e., the strain has created a niche in the plant) or repeated introduction (i.e., the strain may be frequently re-introduced into the plant from sources such as raw material or the surrounding environment).

There are six main virulence genes that are essential for the intracellular pathogen lifecycle of *L. monocytogenes*: *prfA*, *plcA*, *hly*, *mpl*, *actA*, and *plcB (34, 48, 49)*. These can all be

found in a virulence gene island, known as the *prfA* virulence gene cluster (pVGC) or *Listeria* pathogenicity island 1 (LIPI-1) *(48, 49)*. These genes encode a transcriptional regulator (*prfA*), phospholipases (*plcA* and *plcB*) and a hemolysin (*hly*) essential for lysis of host cell phagosomes, a metalloprotease (*mpl*) involved in extracellular activation of *plcB*, and a surface protein (*actA*) needed for actin-based motility and cell-to-cell spread *(48, 49)*.

L. monocytogenes is susceptible to most antibiotics, with the exception of cephalosporins and fluoroquinolones *(5, 14, 43-45, 50)*. Ampicillin alone or in combination with gentamicin (an aminoglycoside) is the standard (first-line) treatment for listeriosis *(5, 14, 26, 43, 44, 50)*. However, alternate (or second-line) treatments may be used in the case of allergies or for some disease states, which usually include trimethoprim/sulfamethoxazole, erythromycin, vancomycin, and the fluoroquinolones *(5, 14, 26, 43, 50)*. Fortunately, acquired (secondary) antibiotic resistance is not often seen in clinical *L. monocytogenes* isolates *(44)*. Nonetheless, it is important to monitor antibiotic resistance in *L. monocytogenes* in food products and food production environments to be able to track the prevalence and spread of antibiotic resistance genes in order to be able to effectively treat *L. monocytogenes* infections.

The objective of this study was to sequence the genomes of and further characterize *L. monocytogenes* isolates from two poultry further processing plants to gain insights into the types and genetic diversity and particularly types (MLST types, serotypes, etc.) of this pathogen that enter or colonize in these types of food processing environments and that may potentially contaminate food products. Understanding the strains of *L. monocytogenes* that enter and may establish niches in poultry processing plants may give us more awareness into the magnitude of this problem, contamination processes, and potential mitigation strategies. In addition, the *L. monocytogenes* genomes were also analyzed for the presence of antibiotic, stress, and

metal/sanitizer resistance-associated genes. Knowing the prevalence of these genes in *L. monocytogenes* isolates from processing environments gives insights into the status of antibiotic resistance in this this pathogen and the potential for these pathogens to be more resistant to current sanitizing agents or environmental control processes.

Materials and Methods

Isolate selection. *Listeria monocytogenes* isolates used in this project were collected from either a poultry further processing plant (Plant A) over a 12 month period by Berrang, et al. *(8)* or from a newly constructed poultry further processing plant (Plant B) over a 21 month period by Berrang, et al. *(7)*. In the first study, the 161 isolates were separated into 14 discrete subtypes by sequencing the *actA* gene. Four subtypes were found to be putatively persistent in the processing environment (floor drains). In the second study, 660 isolates were separated into 16 discrete subtypes by the same method. Of these, three subtypes were found to be putatively persistent in the processing environment (floor drains). A subset of isolates representing each subtype defined as persistent or transient in the original studies was selected for analysis. At least one isolate of each *actA* subtype isolated at each sampling point was selected for sequencing and analysis. A total of 169 isolates were selected. Details for all isolates selected for sequencing and further analysis (Table 3.1).

DNA extraction, isolation, and library preparation. Prior to the study, isolates were maintained at -76°C in cryogenic storage vials (Prolab Diagnostics™ Microbank™ Bacterial and Fungal Preservation System; Fisher Scientific, Wilmington, DE). Isolates were inoculated onto modifed Oxford agar (MOX) plates (Acumedia Oxford *Listeria* Agar Base [Neogen, Lansing, MI] plus moxalactam supplement [Sigma, St. Louis, MO]), incubated at 35°C for 24 h before transferring an isolated, typical colony to tryptic soy broth (TSB) which was incubated at 37°C

for 24 h. For each isolate, total DNA (chromosomal and plasmid) extraction and isolation was performed using the MO BIO Ultraclean Microbial DNA kit (Mo Bio Laboratories, Solana Beach, CA) according to manufacturer's instructions. A NanoDrop ND-2000 spectrophotometer (ThermoFisher Scientific, Wilmington, DE) was used to determine initial DNA concentration (by measuring the optical density at 260 nm) and purity. DNA was eluted into 10 mM Tris-HCL, pH 8.5 and stored at -20°C until library preparation.

The amounts of DNA in samples were quantified with a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA) and Qubit dsDNA BR (broad range) Assay Kit (Invitrogen, Eugene, OR) according to manufacturer's instructions. DNA concentrations for each sample were normalized to 0.2 ng/µl by adding water. DNA libraries were prepared for sequencing using the Nextera XT DNA sample preparation kit and the Nextera XT index kit with 96 indices (Illumina, Inc., San Diego, CA) according to manufacturer's instructions. DNA samples for PacBio sequencing were prepared based on the Genomic Tip-100 (#10243) gravity column (Qiagen, Germantown, MD).

Genome sequencing. Whole-genome sequencing was conducted using a Illumina MiSeq platform (Illumina, Inc., San Diego, CA). Four reference genomes (representative isolates from the persistent and transient subtypes) were sequenced using the Pacific Biosciences RS sequencing platform.

Genome assembly. FastQC (version v0.10.1) was used to view the quality of raw sequencing reads *(3)*. SPAdes Genome Assembler (version 3.7.0) was used for *de novo* assembly of the genomes *(4, 31, 32)*. The quality of the draft genome assemblies was evaluated using QUAST (version 2.3) *(1, 15)*. Assemblies with an N50 value higher than 100,000 were used for further analysis (160 genomes met this qualification).

Genome annotation. Each draft genome with a qualified N50 value was annotated using Prokka (version 1.11) with default parameters *(42)*.

Phylogenetic analysis and SNP calling with Lyve-SET. A phylogenetic tree was created using the hqSNPs (at least 75% consensus and 10x coverage thresholds) detected by Lyve-SET (version 1.1.4f) *(20, 21)*. One complete genome of *L. monocytogenes* (accession# CP006598) was used as reference genome. The phage regions were removed from analysis by the default setting of Lyve-set. During this step any genomes that were not believed to be *L. monocytogenes* were removed, i.e., they were identified as non-*monocytogenes Listeria* spp. (156 genomes remained).

WGS-based Subtyping. ETE3 (version 3.0.0b35) *(16)* was used to parse the phylogenetic tree and its nodes for WGS-based subtyping. A subtype was defined as a monophyletic group of closely related isolates as measured by the maximum pairwise SNP distance (≤30) among members of the group *(18)*. All calculations are based on the SNP matrix created using pairwise SNP distance matrix of Lyve-SET.

WGS-based isolate characterization. Isolates were *in silico* divided into lineages, MLST-typed, and PCR-serotyped using appropriate schemes provided by the Pasteur Institute *(36)*. Isolates were screened for acquired antibiotic-, stress-, metal-, and sanitizer-resistance genes using ABRicate *(41)*. Antibiotic resistance genes screened for included: *aacA4*, *aphA*, *cat_CHL*, *dfrD*, *ermB*, *fosX*, *lmo0441*, *lmo0919*, *norB*, *qnrB*, *str*, *sul*, *tetM*, and *tetS*. Antibiotic-resistance genes were also identified by using the ResFinder database. Stress-resistance genes screened for included: *lmo0444*, *lmo0445*, *lmo0446*, *lmo0447*, *lmo0448*, *lin0464*, and *lin0465*. Metal- and sanitizer-resistance genes screened for included: *bcrA*, *bcrB*, *bcrC*, *ermE*, *ermC*, *qacA*, *qacC*, and *Tn6188_qac*.

Nucleotide accession numbers. Raw sequence data and *de novo* assembled contigs have been deposited to the appropriate GenBank database (Sequence Read Archive [SRA] and Whole Genome Shotgun) under BioProject ID PRJNA450812. Individual genome sequencing metrics and GenBank accessions are listed in Table S1 in the supplemental material.

Results

Phylogeny and lineage. Based on the phylogenetic tree created by SNPs detected by Lyve-SET (Figure 3.1), the sampled *L. monocytogenes* population was divided into 43 subtypes (Table 3.1). Of the isolates analyzed, 56 belong to lineage I and 99 belong to lineage II. One isolate had the result of 'none' (Table 3.1).

MLST Sequence Types. Among the isolates analyzed, 18 unique MLST sequence types (ST) were found, including ST-1 (n=1), ST-5 (n=31), ST-6 (n=20), ST-7 (n=11), ST-9 (n=9), ST-155 (n=27), ST-199 (n=5), ST-224 (n=1), ST-290 (n=1), ST-321 (n=41), ST-323 (n=1), ST-371 (n=1), ST-372 (n=1), ST-378 (n=1), ST-551 (n=2), ST-631 (n=1), ST-920 (n=1), and ST-1006 (n=1) (Table 3.1).

***In silico* PCR-serotypes.** The majority of isolated belonged to serotype 2a (n=90), followed by 2b (n=33), 4b (n=23), 2c (n=9), and one was unable to be serotyped (Table 3.1).

Virulence genes. All the genomes analyzed contained the six main virulence genes (*prfA*, *plcA*, *hly*, *mpl*, *actA*, and *plcB*).

Antibiotic Resistance Genes. All the genomes screened (n=156) harbored *fosX*, *lmo0441*, *lmo0919*, *norB*, and *sul* ABR genes (Table 3.2). Of the genomes, 14.1% (n=22) contained *aacA4* and 3.2% (n=5) contained *tetM* ABR genes. Over 80 percent (82.7%; n=129) of isolate genomes had 5 total ABR genes identified, 16% (n=25) had 6 total ABR genes identified, and 1.3% (n=2) had 7 total ABR genes identified.

Stress Resistance Genes. The stress-resistance genes *lmo0444*, *lmo0445*, *lmo0446*, *lmo0447*, and *lmo0448* were found in 82.7% of isolates (Table 3.2). No other stress-resistance genes were identified in any isolates.

Sanitizer and Metal Resistance Genes. The sanitizer and metal-resistance genes *bcrA*, *bcrB*, and *bcrC* were present in 73.1% (n=114) of isolates (Table 3.2). No other sanitizer or metal-resistance genes were identified in any isolates.

Discussion

The results obtained in this study provide useful information about *L. monocytogenes* that may be present or persist in the environment of a poultry processing plant and could potentially contaminate poultry products and impact public health.

Currently, the Pasteur MLST database for *L. monocytogenes* contains 1,373 unique MLST types and 3,388 different isolates. Of the 228 of the isolates in the database described as being from a "production environment source," the most common MLST sequence types are ST-426 (18.4%; n=42), ST-177 (9.6%; n=22), ST-3 (7.9%; n=18), ST-1 (6.6%; n=15), ST-121 (6.6%; n=15), ST-2 (6.1%; n=14), ST-9 (4.8%; n=11), ST-7 (3.1%; n=7). Many of the most prevalent STs evaluated in the present study are not among the most common food production ST in the Pasteur database. Knudsen, et al. found ST-7, ST-8, and ST-121 MLST sequence types to be persistent in Danish food processing plants *(22)*, which contrasts with the results of the current study, as ST-7 was the only one of the three that were isolated from these poultry processing plants.

In contrast, the top isolates from human sources (n=1,140) are ST-1 (14.7%; n=168), ST-3 (8.3%; n=95), ST-2 (8.2%; n=94), ST-9 (4.5%; n=51), ST-4 (3.1%; n=35), ST-6 (2.9%; n=33), ST-155 (2.9%; n=33), ST-7 (2.3%; n=26), and ST-5 (2.1%; n=24). In the current study, of the

common clinical ST, the following were found: ST-1 (n=1), ST-9 (n=9), ST-6 (n=20), ST-155 (n=27), ST-7 (n=11), and ST-5 (n=31). The high prevalence of these clinically relevant *L. monocytogenes* STs in a food production plant may be of concern.

L. monocytogenes strains can be divided into 13 serovars *(19, 23, 29, 38, 40, 48)*: 1/2a, 3a, 1/2c, 3c, 4b, 4d, 4e, 1/2b, 3b, 7, 4a, 4ab and 4c. All of the serovars have the potential to cause disease, but most (>90%) isolates from humans with listeriosis are 1/2a, 1/2b, and 4b *(29, 40)*, with 1/2a and 1/2b being the most commonly isolated from human listeriosis patients in Europe and 4b most common in the United States and Canada *(38)*. Serovar 4b strains cause an estimated 33-50% sporadic listeriosis cases worldwide *(40)*. In contrast, 1/2a and 1/2c are the most common serovars isolated from foods *(40)*. All of the isolates in the current study fall into the main serotypes responsible for listeriosis (1/2a, 1/2b, and 4b). The fact that all isolates contained the six main virulence genes is expected, as most known virulence genes identified in *L. monocytogenes* are part of the core genome *(28)*. This suggests that they all are potentially pathogenic.

Troxler, et al. conducted a study to determine the natural susceptibility of various *Listeria* spp. to 71 antimicrobial agents *(45)*. They found that *L. monocytogenes* showed resistance to fosfomycin *(45)*, which agrees with the findings of the current study (100% of isolates contained the *fosX* gene that confers fosfomycin resistance). They also found that *L. monocytogenes* is naturally sensitive to trimethoprim *(45)*, which also agrees with the findings of the current study (the *dfrD* gene that confers trimethoprim resistance was not found in any isolates). The absence of *ermB*, *dfrD*, and *qnrB* genes in all isolates in the current study is a positive finding, since the antibiotics these genes confer resistance to (macrolides, trimethoprim, and fluoroquinolone, respectively) are often second-line treatments for listeriosis *(5, 14, 26, 43, 44, 50)*. However,

14.1% of isolates harbored *aacA4*, an aminoglycoside resistance-associated gene, which may be of concern as gentamicin is often used in conjunction with ampicillin as a first-line treatment for listeriosis *(5, 14, 26, 43, 44, 50)*.

Lyon, et al. (54) evaluated the antimicrobial susceptibility of isolates from the Berrang el al. *(8)* study and found isolates resistant to ceftriaxone (153 isolates [97%] intermediate or resistant), oxacillin with 2% NaCl (142 isolates [90%] resistant), ciprofloxacin (59 isolates [37%] intermediate or resistant), tetracycline (5 isolates [3%] resistant), clindamycin (43 isolates [27%] intermediate), linezolid (3 isolates [2%] intermediate), and trimethoprim/sulfamethoxazole (1 isolate [<1%] intermediate). Lyon, et al. concluded that antimicrobial resistance was not highly prevalent in *L. monocytogenes* isolates from that poultry further processing plant *(25)*. The prevalence of tetracycline resistance reported is in agreement with the prevalence of tetracycline-related resistance genes found in the current study (3.2% [n=5] isolates contained the *tetM* gene).

Benzalkonium chloride (BC) is a quaternary ammonium compound (QAC), which is commonly used in food processing plants. Due to the fact that the majority of isolates screened in the current study harbored *bcrA*, *bcrB*, and *bcrC* genes, which are associated with BC tolerance in *L. monocytogenes (51)*, it is important to ensure that this disinfectant is not to be relied on to eradicate *L. monocytogenes* in food processing environments. These results are in accordance with a study conducted by Mullapudi, et al., who found a large prevalence of resistance to BC in *L. monocytogenes* isolated from turkey processing plants *(30)*. In addition, the plasmid harboring the *bcrABC* cassette has also been associated with conferring cadmium (Cd) tolerance in *L. monocytogenes* and has potential to be transferred to non-pathogenic listeriae or other bacterial pathogens, such as *E. coli (30, 51)*. In addition, exposure to sublethal

concentrations of BC has also been associated with creating selective pressure for *L. monocytogenes* to show increased tolerance to some antibiotics and toxic chemicals *(37)*.

The results of the current study provide useful information about the characteristics of *L. monocytogenes* isolated from poultry processing facilities and poultry products. Clinically relevant MLST subtypes were identified and genes associated with resistance to antibiotics used to treat listeriosis were detected indicate that such isolates could be a threat to public health. In addition, information gained about antibiotic-, sanitizer-, and stress-resistance genes contained in these isolates may be helpful in making decisions about drug treatments for listeriosis patients, sanitizers and other antimicrobial treatment protocols used in food processing environments, and to gauge the potential for transfer of these resistance genes to other *L. monocytogenes* strains or to other foodborne pathogens.

Acknowledgements

Georgia Agricultural Experiment Stations, USDA's National Needs Fellowship program, and University of Georgia's Center for Food Safety provided support. Thanks to Steven Knapp, Eric Adams, Linda Genzlinger, and Renato Orsi for technical assistance.

References

1. Anonymous. Quast. Available at: http://quast.sourceforge.net/quast. Accessed 20 February 2018.

2. Autio, T., J. Lunden, M. Fredriksson-Ahomaa, J. Bjorkroth, A. M. Sjoberg, and H. Korkeala. 2002. Similar *Listeria monocytogenes* pulsotypes detected in several foods originating from different sources. *Int. J. Food Microbiol.* 77:83-90.

3. Babraham Bioinformatics. Fastqc. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 20 February 2018.

4. Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, and P. A. Pevzner. 2012. Spades: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19:455-477.

5. Bartlett, J. G. 2016. *Listeria monocytogenes*. Available at: https://www.hopkinsguides.com/hopkins/view/Johns_Hopkins_ABX_Guide/540318/all/Listeria_Monocytogenes. Accessed 22 February 2018.

6. Berrang, M. E., C. E. Lyon, D. P. Smith, and J. K. Northcutt. 2000. Incidence of *Listeria monocytogenes* on pre-scald and post-chill chicken. *J. Appl. Poult. Res.* 9:546-550.

7. Berrang, M. E., R. J. Meinersmann, J. F. Frank, and S. R. Ladely. 2010. Colonization of a newly constructed commercial chicken further processing plant with *Listeria monocytogenes*. *J. Food Prot.* 73:286-291.

8. Berrang, M. E., R. J. Meinersmann, J. F. Frank, D. P. Smith, and L. L. Genzlinger. 2005. Distribution of *Listeria monocytogenes* subtypes within a poultry further processing plant. *J. Food Prot.* 68:980-985.

9. Berrang, M. E., R. J. Meinersmann, J. K. Northcutt, and D. P. Smith. 2002. Molecular characterization of *Listeria monocytogenes* isolated from a poultry further processing facility and from fully cooked product. *Poult. Sci.* 81:130-130.

10. Berrang, M. E., R. J. Meinersmann, J. K. Northcutt, and D. P. Smith. 2002. Molecular characterization of *Listeria monocytogenes* isolated from a poultry further processing facility and from fully cooked product. *J. Food Prot.* 65:1574-1579.

11. Cartwright, E. J., K. A. Jackson, S. D. Johnson, L. M. Graves, B. J. Silk, and B. E. Mahon. 2013. Listeriosis outbreaks and associated food vehicles, United States, 1998–2008. *Emerging Infect. Dis.* 19:1-9.

12. Centers for Disease Control and Prevention (CDC). 2002. Public health dispatch: Outbreak of listeriosis --- northeastern United States, 2002. p. 950-951. *In*, MMWR, vol. 51.

13. Chasseignaux, E., M. T. Toquin, C. Ragimbeau, G. Salvat, P. Colin, and G. Ermel. 2001. Molecular epidemiology of *Listeria monocytogenes*, isolates collected from the environment, raw meat and raw products in two poultry- and pork-processing plants. *J. Appl. Microbiol.* 91:888-899.

14. de Vasconcelos Byrne, V., E. Hofer, D. C. Vallim, and R. C. de Castro Almeida. 2016. Occurrence and antimicrobial resistance patterns of *Listeria monocytogenes* isolated from vegetables. *Braz. J. Microbiol.* 47:438-443.

15. Gurevich, A., V. Saveliev, N. Vyahhi, and G. Tesler. 2013. Quast: Quality assessment tool for genome assemblies. *Bioinformatics*. 29:1072-1075.

16. Huerta-Cepas, J., F. Serra, and P. Bork. 2016. Ete 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* 33:1635-1638.

17. Interagency Food Safety Analytics Collaboration. 2017. Foodborne illness source attribution estimates for 2013 for *Salmonella, Escherichia coli* O157, *Listeria monocytogenes*, and *Campylobacter* using multi-year outbreak surveillance data, United States. *In*.

18. Jackson, B. R., C. Tarr, E. Strain, K. A. Jackson, A. Conrad, H. Carleton, L. S. Katz, S. Stroika, L. H. Gould, and R. K. Mody. 2016. Implementation of nationwide real-time whole-genome sequencing to enhance listeriosis outbreak detection and investigation. *Clin. Infect. Dis.* 63:380-386.

19. Jay, J. M. 2005. Foodborne listeriosis. *In* J.M. Jay, D.A. Golden, and M.J. Loessner (ed.), Modern food microbiology. 7th ed. Springer, New York

20. Katz L.S., Griswold T., Williams-Newkirk A.J., Wagner D., Petkau A., and et al. Lyve-set. Available at: https://github.com/lskatz/lyve-SET. Accessed 20 February 2018.

21. Katz, L. S., T. Griswold, A. J. Williams-Newkirk, D. Wagner, A. Petkau, C. Sieffert, G. Van Domselaar, X. Deng, and H. A. Carleton. 2017. A comparative analysis of the lyve-set phylogenomics pipeline for genomic epidemiology of foodborne pathogens. *Front. Microbiol.* 8:375.

22. Knudsen, G. M., J. B. Nielsen, R. L. Marvig, Y. Ng, P. Worning, H. Westh, and L. Gram. 2017. Genome-wide-analyses of *Listeria monocytogenes* from food-processing plants reveal clonal diversity and date the emergence of persisting sequence types. *Environ. Microbiol. Rep.* 9:428-440.

23. Kuenne, C., A. Billion, M. A. Mraheil, A. Strittmatter, R. Daniel, A. Goesmann, S. Barbuddhe, T. Hain, and T. Chakraborty. 2013. 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.

24. Lunden, J. M., T. J. Autio, A. M. Sjoberg, and H. J. Korkeala. 2003. Persistent and nonpersistent *Listeria monocytogenes* contamination in meat and poultry processing plants. *J. Food Prot.* 66:2062-2069.

25. Lyon, S. A., M. E. Berrang, P. J. Fedorka-Cray, D. L. Fletcher, and R. J. Meinersmann. 2008. Antimicrobial resistance of *Listeria monocytogenes* isolated from a poultry further processing plant. *Foodborne Pathog. Dis.* 5:253-259.

26. Madjunkov, M., S. Chaudhry, and S. Ito. 2017. Listeriosis during pregnancy. *Arch. Gynecol. Obstet.* 296:143-152.

27. Marcus, R., S. Hurd, L. Mank, P. Mshar, Q. Phan, K. Jackson, K. Watarida, Y. Salfinger, S. Kim, M. L. Ishida, and B. Kissler. 2009. Chicken salad as the source of a case of *Listeria monocytogenes* infection in connecticut. *J. Food Prot.* 72:2602-6.

28. Maury, M. M., Y.-H. Tsai, C. Charlier, M. Touchon, V. Chenal-Francisque, A. Leclercq, A. Criscuolo, C. Gaultier, S. Roussel, A. Brisabois, O. Disson, E. P. C. Rocha, S. Brisse, and M. Lecuit. 2016. Uncovering *Listeria monocytogenes* hypervirulence by harnessing its biodiversity. *Nat. Genet.* 48:308-313.

29. Montville, T. J., K. R. Matthews, and K. E. Kniel. 2012. Chapter 13: *Listeria monocytogenes.* p. 189-204. *In* T.J. Montville, K.R. Matthews, and K.E. Kniel (ed.), Food microbiology: An introduction 3rd ed. ASM Press, Washington; USA.

30. Mullapudi, S., R. M. Siletzky, and S. Kathariou. 2008. Heavy-metal and benzalkonium chloride resistance of *Listeria monocytogenes* isolates from the environment of turkey-processing plants. *Appl. Environ. Microbiol.* 74:1464-1468.

31. Nurk, S., A. Bankevich, D. Antipov, A. Gurevich, A. Korobeynikov, A. Lapidus, A. Prjibelsky, A. Pyshkin, A. Sirotkin, Y. Sirotkin, R. Stepanauskas, J. McLean, R. Lasken, S. Clingenpeel, T. Woyke, G. Tesler, M. Alekseyev, and P. Pevzner. 2013. Assembling genomes and mini-metagenomes from highly chimeric reads. p. 158-170. *In* M. Deng, et al. (ed.), Research in computational molecular biology, vol. 7821.ed. Springer Berlin Heidelberg.

32. Nurk, S., A. Bankevich, D. Antipov, A. A. Gurevich, A. Korobeynikov, A. Lapidus, A. D. Prjibelski, A. Pyshkin, A. Sirotkin, Y. Sirotkin, R. Stepanauskas, S. R. Clingenpeel, T. Woyke, J. S. McLean, R. Lasken, G. Tesler, M. A. Alekseyev, and P. A. Pevzner. 2013. Assembling single-cell genomes and mini-metagenomes from chimeric mda products. *J. Comput. Biol.* 20:714-737.

33. Olsen, S. J., M. Patrick, S. B. Hunter, V. Reddy, L. Kornstein, W. R. MacKenzie, K. Lane, S. Bidol, G. A. Stoltman, D. M. Frye, I. Lee, S. Hurd, T. F. Jones, T. N. LaPorte, W. Dewitt, L. Graves, M. Wiedmann, D. J. Schoonmaker-Bopp, A. J. Huang, C. Vincent, A. Bugenhagen, J. Corby, E. R. Carloni, M. E. Holcomb, R. F. Woron, S. M. Zansky, G. Dowdle, F. Smith, S. Ahrabi-Fard, A. R. Ong, N. Tucker, N. A. Hynes, and P. Mead. 2005. Multistate outbreak of *Listeria monocytogenes* infection linked to delicatessen turkey meat. *Clin. Infect. Dis.* 40:962-967.

34. Orsi, R. H., S. B. Maron, K. K. Nightingale, M. Jerome, H. Tabor, and M. Wiedmann. 2008. Lineage specific recombination and positive selection in coding and intragenic regions contributed to evolution of the main *Listeria monocytogenes* virulence gene cluster. *Infect. Genet. Evol.* 8:566-576.

35. Painter, J. A., R. M. Hoekstra, T. Ayers, R. V. Tauxe, C. R. Braden, F. J. Angulo, and P. M. Griffin. 2013. Attribution of foodborne illnesses, hospitalizations, and deaths to food commodities by using outbreak data, United States, 1998-2008. *Emerging Infect. Dis.* 19:407-415.

36. Pasteur Institute. Bigsdb-lm: *Listeria* locus/sequence definitions database. Available at: http://bigsdb.pasteur.fr/perl/bigsdb/bigsdb.pl?db=pubmlst_listeria_seqdef_public. Accessed 26 March 2018.

37. Rakic-Martinez, M., D. A. Drevets, V. Dutta, V. Katic, and S. Kathariou. 2011. *Listeria monocytogenes* strains selected on ciprofloxacin or the disinfectant benzalkonium chloride exhibit reduced susceptibility to ciprofloxacin, gentamicin, benzalkonium chloride, and other toxic compounds. *Appl. Environ. Microbiol.* 77:8714-8721.

38. Ray, B., and A. Bhunia. 2014. Fundamental food microbiology CRC Press, Boca Raton, FL.

39. Rorvik, L. M., B. Aase, T. Alvestad, and D. A. Caugant. 2003. Molecular epidemiological survey of *Listeria monocytogenes* in broilers and poultry products. *J. Appl. Microbiol.* 94:633-640.

40. Ryser, E. T., and R. L. Buchanan. 2013. *Listeria monocytogenes.* p. 503-545. *In* M.P. Doyle, and R.L. Buchanan (ed.), Food microbiology: Fundamentals and frontiers 4th ed. ASM Press, Washington, D.C.

41. Seemann, T. Abricate. Available at: https://github.com/tseemann/abricate. Accessed 20 February 2018.

42. Seemann, T. 2014. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*. 30:2068-2069.

43. Temple, M. E., and M. C. Nahata. 2000. Treatment of listeriosis. *The Annals of pharmacotherapy*. 34:656-61.

44. Thønnings, S., J. D. Knudsen, H. C. Schønheyder, M. Søgaard, M. Arpi, K. O. Gradel, and C. Østergaard. 2016. Antibiotic treatment and mortality in patients with *Listeria monocytogenes* meningitis or bacteraemia. *Clin. Microbiol. Infect.* 22:725-730.

45. Troxler, R., A. von Graevenitz, G. Funke, B. Wiedemann, and I. Stock. 2000. Natural antibiotic susceptibility of *Listeria* species: *L. grayi*, *L. innocua*, *L. ivanovii*, *L. monocytogenes*, *L. seeligeri* and *L. welshimeri* strains. *Clin. Microbiol. Infect.* 6:525-535.

46. United States Department of Agriculture. 2002. Recall notification report. Available at: https://www.fsis.usda.gov/wps/portal/frame-

redirect?url=https://www.fsis.usda.gov/OA/recalls/rnrfiles/rnr090-2002.htm. Accessed 4 March 2018.

47. United States Department of Agriculture. 2018. Recall case archive. Available at: https://www.fsis.usda.gov/wps/portal/fsis/topics/recalls-and-public-health-alerts/recall-case-archive. Accessed 4 March 2018.

48. Vázquez-Boland, J. A., M. Kuhn, P. Berche, T. Chakraborty, G. Domínguez-Bernal, W. Goebel, B. González-Zorn, J. Wehland, and J. Kreft. 2001. *Listeria* pathogenesis and molecular virulence determinants. *Clin. Microbiol. Rev.* 14:584-640.

49. Ward, T. J., L. Gorski, M. K. Borucki, R. E. Mandrell, J. Hutchins, and K. Pupedis. 2004. Intraspecific phylogeny and lineage group identification based on the *prfa* virulence gene cluster of *Listeria monocytogenes. J. Bacteriol.* 186:4994–5002.

50. Wieczorek, K., and J. Osek. 2017. Prevalence, genetic diversity and antimicrobial resistance of *Listeria monocytogenes* isolated from fresh and smoked fish in poland. *Food Microbiol.* 64:164-171.

51. Xu, D., Q. Nie, W. Wang, L. Shi, and H. Yan. 2016. Characterization of a transferable *bcrabc* and *cadac* genes-harboring plasmid in *Listeria monocytogenes* strain isolated from food products of animal origin. *Int. J. Food Microbiol.* 217:117-122.

Table 3.1. *Listeria monocytogenes* isolate metadata and characterization results of poultry processing plant derived isolates. Includes isolate ID, processing plant, lineage, MLST sequence type, serotype, *actA*-based subtype from original studies, WGS-based subtype from current study, sampling trip, pre- or post-operation sampling, location of isolation, sample type, other sampling information, and sampling time.

Isolate ID	Processing Plant[a]	Line-age	MLST Sequence Type	Sero-type	*actA*-Based Subtype[b]	WGS-Based Subtype[c]	Samp-ling Trip	Pre or Post[d]	Location	Sample Type[e]	Other Sampling Info.	Sample Time
232	A	I	6	4b	A	43	1	Post	Cook side	Drain 4a	Pipe	
233	A	I	5	2b	B	32	1	Pre		Raw product		Morning, after break
235	A	I	6	4b	A	1	2	Pre	Raw side	Drain 1a	Cover	
237	A	I	6	4b	A	42	2	Post	Raw side	Drain 1a	Cover	
239	A	II	321	2a	C	7	2	Post	Cook side	Drain 5a	Cover	
241	A	I	6	4b	A	1	2	Pre	Raw side	Drain 1a	Pipe	
243	A	I	6	4b	A	1	2	Post	Raw side	Drain 1a	Pipe	
246	A	I	6	4b	A	2	2	Post	Raw side	Drain 2a	Pipe	
247	A	I	5	2b	D	31	2	Post	Cook side	Drain 5a	Pipe	
249	A	II	551	2a	E	4	2	Pre		Raw product		Morning, after break
251	A	I	5	2b	D	12	2	Post		Raw product		Afternoon, after break
253	A	II	155	2a	F	16	3	Pre	Raw side	Drain 1a	Cover	
255	A	I	5	2b	D	12	3	Pre	Raw side	Drain 2a	Cover	

Isolate ID	Processing Plant[a]	Line-age	MLST Sequence Type	Sero-type	*actA*-Based Subtype[b]	WGS-Based Subtype[c]	Samp-ling Trip	Pre or Post[d]	Location	Sample Type[e]	Other Sampling Info.	Sample Time
257	A	II	9	2c	G	11	3	Post	Raw side	Drain 1a	Cover	
259	A	I	5	2b	D	34	3	Post	Cook side	Drain 3a	Cover	
261	A	II	7	2a	H	15	3	Pre	Raw side	Drain 1a	Pipe	
262	A	I	5	2b	D	33	3	Pre	Raw side	Drain 1a	Pipe	
263	A	I	5	2b	D	12	3	Pre	Raw side	Drain 2a	Pipe	
265	A	I	5	2b	D	35	3	Post	Raw side	Drain 2a	Pipe	
267	A	II	1006	2a	I	23	3	Post		Raw product		Afternoon, before break
277	A	I	6	4b	A	3	4	Pre	Raw side	Drain 1a	Cover	
279	A	II	321	2a	C	7	4	Pre	Raw side	Drain 2a	Cover	
281	A	II	321	2a	C	7	4	Post	Cook side	Drain 3a	Cover	
282	A	II	155	2a	F	16	4	Post	Cook side	Drain 3a	Cover	
283	A	I	6	4b	A	3	4	Pre	Raw side	Drain 1a	Pipe	
285	A	II	9	2c	G	9	4	Post	Raw side	Drain 1a	Pipe	
287	A	II	321	2a	C	7	4	Post	Raw side	Drain 2a	Pipe	
288	A	II	7	2a	H	27	4	Post	Raw side	Drain 2a	Pipe	
293	A	I	6	4b	A	3	5	Pre	Raw side	Drain 1a	Cover	

Isolate ID	Processing Plant[a]	Line-age	MLST Sequence Type	Sero-type	*actA*-Based Subtype[b]	WGS-Based Subtype[c]	Samp-ling Trip	Pre or Post[d]	Location	Sample Type[e]	Other Sampling Info.	Sample Time
295	A	I	6	4b	UT[f]	3	5	Post	Raw side	Drain 1a	Cover	
296	A	I	6	4b	UT[f]	3	5	Post	Raw side	Drain 1a	Cover	
298	A	I	5	2b	D	10	5	Pre	Raw side	Drain 1a	Pipe	
299	A	I	5	2b	D	12	5	Pre	Raw side	Drain 2a	Pipe	
301	A	I	5	2b	D	12	5	Post	Raw side	Drain 1a	Pipe	
303	A	I	5	2b	D	12	5	Post	Cook side	Drain 5a	Pipe	
305	A	I	5	2b	D	36	5	Post		Raw product		Afternoon, after break
307	A	II	321	2a	C	7	5	Pre		Other samples	Hose	Morning
308	A	II	321	2a	C	7	5	Pre		Other samples	Cooler frame	Morning
311	A	I	290	4b	J	38	5	Pre		Other samples	Door handle	Afternoon
325	A	II	7	2a	H	14	6	Post	Raw side	Drain 1a	Cover	
327	A	II	9	2c	G	11	6	Pre	Raw side	Drain 2a	Pipe	
329	A	II	551	2a	E	4	6	Post	Raw side	Drain 2a	Pipe	
330	A	I	5	2b	D	12	6	Post	Raw side	Drain 2a	Pipe	
331	A	II	155	2a	F	13	6	Post		Other samples	Drain	Afternoon
333	A	I	5	2b	D	12	6	Pre	Raw side	Drain 1a	Cover	

Isolate ID	Processing Plant[a]	Line-age	MLST Sequence Type	Sero-type	*actA*-Based Subtype[b]	WGS-Based Subtype[c]	Samp-ling Trip	Pre or Post[d]	Location	Sample Type[e]	Other Sampling Info.	Sample Time
337	A	I	6	4b	A	3	7	Pre	Raw side	Drain 1a	Cover	
339	A	II	9	2c	G	9	7	Post	Raw side	Drain 1a	Cover	
340	A	I	6	4b	A	41	7	Post	Raw side	Drain 1a	Cover	
341	A	II	321	2a	C	7	7	Post	Raw side	Drain 2a	Cover	
343	A	II	321	2a	C	7	7	Post	Raw side	Drain 6a	Cover	
345	A	I	5	2b	D	12	7	Post	Raw side	Drain 1a	Pipe	
346	A	II	155	2a	F	13	7	Post	Raw side	Drain 1a	Pipe	
347	A	I	6	4b	A	3	7	Pre	Raw side	Drain 6a	Cover	
348	A	I	6	4b	D	3	7	Pre	Raw side	Drain 6a	Cover	
350	A	II	199	2a	L	8	7	Post	Raw side	Drain 2a	Pipe	
351	A	II	9	2c	G	9	7	Post		Raw product		Afternoon, after break
357	A	I	5	2b	D	10	7	Post	Raw side	Drain 1a	Cover	
363	A	II	7	2a	H	14	7	Post	Raw side	Drain 1a	Pipe	
370	A	UT[f]	920	UT[f]	M	22	7	Post	Raw product			Afternoon, before break
371	A	II	7	2a	H	26	8	Post	Raw side	Drain 1a	Cover	
373	A	II	199	2a	L	8	8	Post	Raw side	Drain 2a	Cover	

Isolate ID	Processing Plant[a]	Line-age	MLST Sequence Type	Sero-type	*actA*-Based Subtype[b]	WGS-Based Subtype[c]	Samp-ling Trip	Pre or Post[d]	Location	Sample Type[e]	Other Sampling Info.	Sample Time
375	A	II	155	2a	F	16	8	Post	Raw side	Drain 6a	Cover	
377	A	I	5	2b	D	37	8	Post	Raw side	Drain 1a	Pipe	
378	A	I	6	4b	A	3	8	Post	Raw side	Drain 1a	Pipe	
379	A	II	155	2a	F	13	8	Post	Raw side	Drain 2a	Pipe	
380	A	II	199	2a	L	8	8	Post	Raw side	Drain 2a	Pipe	
381	A	II	321	2a	C	7	8	Pre		Raw product		Morning, before break
382	A	II	199	2a	L	8	8	Pre		Raw product		Morning, before break
383	A	I	6	4b	A	3	8	Pre		Raw product		Morning, after break
386	A	II	155	2a	F	13	9	Pre	Raw side	Drain 1a	Cover	
388	A	I	6	4b	A	2	9	Post	Raw side	Drain 1a	Cover	
390	A	II	371	2a	N	21	9	Post	Raw side	Drain 2a	Cover	
391	A	II	9	2c	G	28	9	Post	Raw side	Drain 6a	Cover	
392	A	II	155	2a	F	16	9	Post	Raw side	Drain 6a	Cover	
393	A	II	155	2a	F	13	9	Pre	Raw side	Drain 1a	Pipe	
395	A	II	321	2a	C	7	9	Pre	Raw side	Drain 2a	Pipe	
396	A	I	224	2b	K	30	9	Pre	Raw side	Drain 2a	Pipe	

Isolate ID	Processing Plant[a]	Line-age	MLST Sequence Type	Sero-type	*actA*-Based Subtype[b]	WGS-Based Subtype[c]	Samp-ling Trip	Pre or Post[d]	Location	Sample Type[e]	Other Sampling Info.	Sample Time
398	A	I	6	4b	A	2	9	Post	Raw side	Drain 1a	Pipe	
401	A	II	155	2a	F	16	9	Post	Raw side	Drain 6a	Pipe	
403	A	II	378	2a	C	24	9	Pre		Raw product		Morning, after break
405	A	II	7	2a	H	18	9	Post		Other samples	Puddle	Afternoon
406	A	II	7	2a	F	18	9	Pre		Other samples	Oven drain	Morning
411	A	I	5	2b	D	5	9	Post		Raw product		Afternoon, after break
412	A	I	5	2b	UT[f]	5	9	Post		Raw product		Afternoon, after break
431	B	II	155	2a	B1	16	1	Proc		Drain 1b		
432	B	II	155	2a	B3	16	1	Proc		Drain 1b		
433	B	II	321	2a	C3	7	1	Proc		Drain 2b		
437	B	II	7	2a	C7	15	3	Proc		Drain 3b		
439	B	I	631	4b	A4	20	3			Raw product		
441	B	II	321	2a	C1	7	4	Proc		Drain 4b		
442	B	II	321	2a	C6	7	4	Proc		Drain 4b		
443	B	II	321	2a	C3	7	4	Proc		Drain 4b		
447	B	II	321	2a	C3	7	4	Clean		Drain 4b		

Isolate ID	Processing Plant[a]	Line-age	MLST Sequence Type	Sero-type	*actA*-Based Subtype[b]	WGS-Based Subtype[c]	Samp-ling Trip	Pre or Post[d]	Location	Sample Type[e]	Other Sampling Info.	Sample Time
456	B	II	155	2a	B3	17	6	Clean		Drain 6b		
458	B	II	321	2a	C5	7	6	Proc		Drain 4b		
459	B	II	321	2a	C3	7	6	Proc		Drain 4b		
460	B	II	9	2c	C3	11	6			Raw product		
461	B	II	9	2c	B4	11	6			Raw product		
462	B	II	9	2c	B2	11	6			Raw product		
464	B	I	5	2b	A1	6	6			Raw product		
469	B	II	199	2a	C8	8	7	Proc		Drain 2b		
471	B	II	155	2a	B3	17	7	Clean		Drain 6b		
472	B	II	155	2a	B3	17	7	Clean		Drain 6b		
474	B	II	321	2a	C3	7	8	Clean		Drain 2b		
475	B	II	321	2a	C1	7	8	Clean		Drain 2b		
480	B	II	321	2a	C3	7	8	Clean		Drain 4b		
481	B	I	1	4b	A5	39	8			Raw product		
494	B	II	321	2a	C3	7	9	Proc		Drain 4b		
497	B	II	321	2a	C3	7	9	Clean		Drain 1b		

Isolate ID	Processing Plant[a]	Line-age	MLST Sequence Type	Sero-type	*actA*-Based Subtype[b]	WGS-Based Subtype[c]	Samp-ling Trip	Pre or Post[d]	Location	Sample Type[e]	Other Sampling Info.	Sample Time
498	B	II	7	2a	C7	19	9	Clean		Drain 1b		
502	B	I	323	2b	A2	29	9			Raw product		
508	B	I	5	2b	A1	6	9			Raw product		
512	B	II	321	2a	C3	7	10	Proc		Drain 4b		
517	B	II	321	2a	C3	7	10	Clean		Drain 1b		
522	B	I	5	2b	A1	5	10			Raw product		
529	B	II	155	2a	B3	16	11	Clean		Drain 4b		
532	B	II	372	2a	B3	25	11	Proc		Drain 6b		
533	B	II	321	2a	C3	7	11	Proc		Drain 8b		
535	B	I	5	2b	A1	5	11			Raw product		
537	B	II	155	2a	B1	17	12	Proc		Drain 6b		
538	B	II	155	2a	B3	17	12	Proc		Drain 6b		
541	B	II	321	2a	C3	7	12	Clean		Drain 1b		
544	B	II	155	2a	B3	17	12	Proc		Drain 8b		
546	B	II	321	2a	C1	7	12	Proc		Drain 4b		
547	B	II	321	2a	C3	7	12	Proc		Drain 4b		

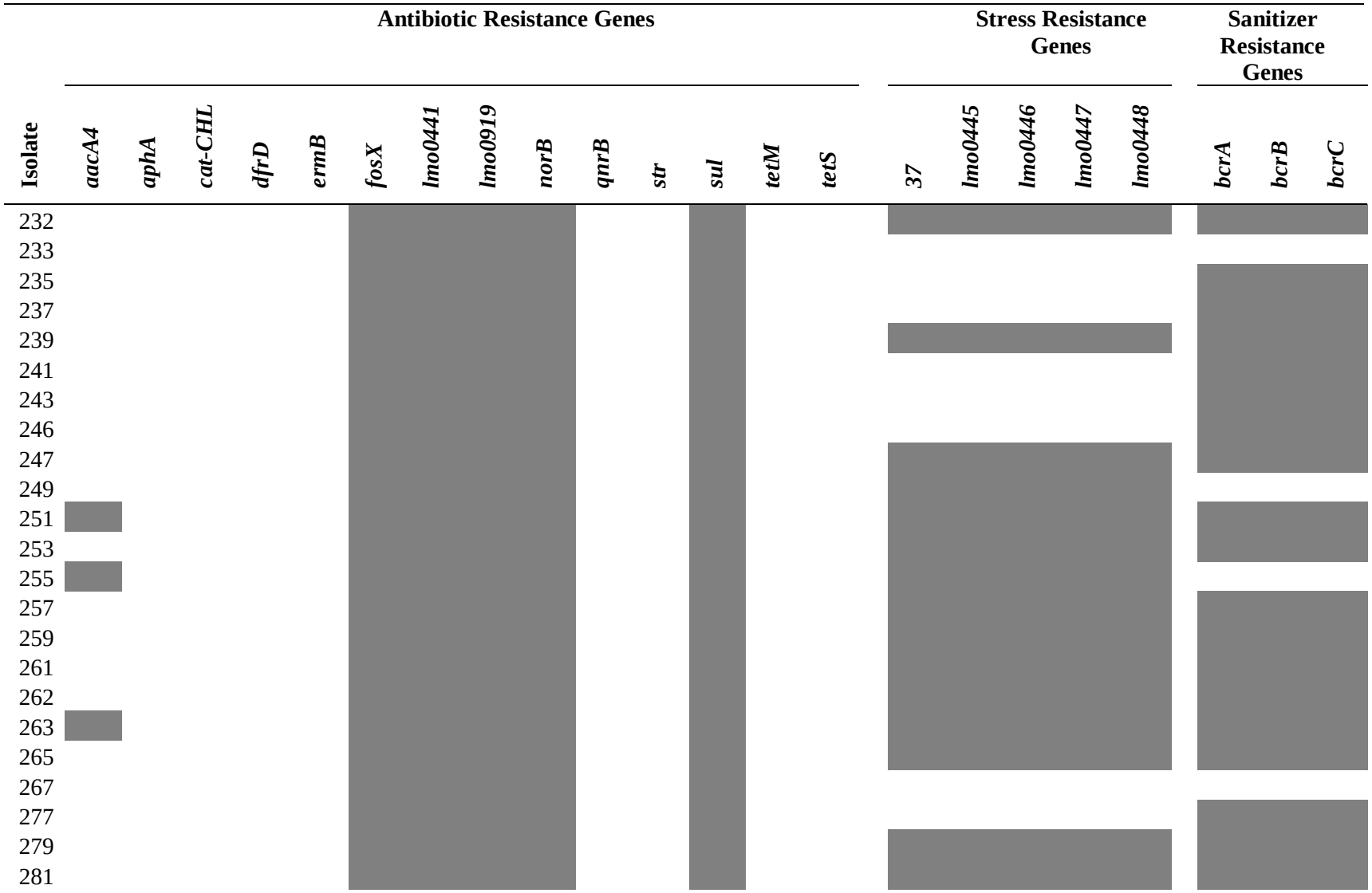
Isolate ID	Processing Plant[a]	Line-age	MLST Sequence Type	Sero-type	*actA*-Based Subtype[b]	WGS-Based Subtype[c]	Samp-ling Trip	Pre or Post[d]	Location	Sample Type[e]	Other Sampling Info.	Sample Time
549	B	II	155	2a	B3	16	12	Clean		Drain 4b		
550	B	II	155	2a	C3	16	12	Clean		Drain 4b		
552	B	I	5	2b	A1	6	12			Raw product		
555	B	I	5	2b	C8	6	13	Proc		Drain 1b		
557	B	II	155	2a	B3	17	13	Proc		Drain 8b		
559	B	I	5	2b	C3	5	14	Proc		Drain 9b		
561	B	II	155	2a	B3	17	14	Clean		Drain 6b		
565	B	II	321	2a	C4	7	16	Proc		Drain 4b		
566	B	II	155	2a	B1	17	16	Proc		Drain 6b		
567	B	II	321	2a	C1	7	16	Clean		Drain 1b		
569	B	I	5	2b	A1	5	16			Raw product		
570	B	II	7	2a	C7	19	17	Proc		Drain 1b		
572	B	II	321	2a	C3	7	17	Proc		Drain 4b		
574	B	II	321	2a	C3	7	17	Proc		Drain 8b		
575	B	II	321	2a	C2	7	17	Proc		Drain 8b		
576	B	II	321	2a	C3	7	17	Clean		Drain 3b		

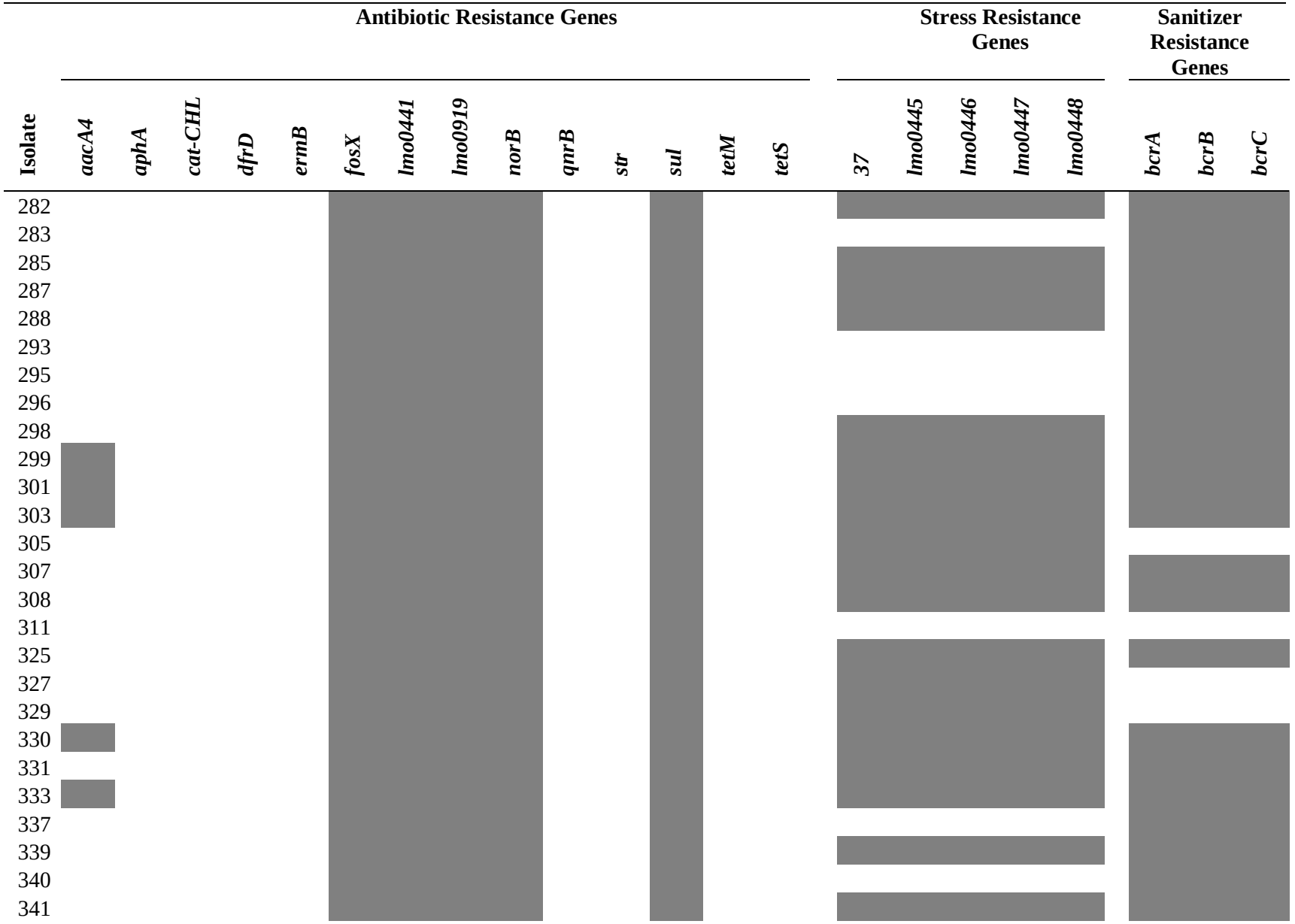
Isolate ID	Processing Plant[a]	Line-age	MLST Sequence Type	Sero-type	*actA*-Based Subtype[b]	WGS-Based Subtype[c]	Samp-ling Trip	Pre or Post[d]	Location	Sample Type[e]	Other Sampling Info.	Sample Time
578	B	I	5	2b	A1	6	17	Clean		Drain 7b		
580	B	II	7	2a	C7	19	18	Proc		Drain 1b		
581	B	I	5	2b	A1	5	18			Raw product		
582	B	II	321	2a	C3	7	19	Proc		Drain 4b		
583	B	II	321	2a	C3	7	19	Clean		Drain 4b		
584	B	I	5	2b	A1	5	19			Raw product		
585	B	II	321	2a	C3	7	20	Proc		Drain 2b		
586	B	II	321	2a	C3	7	20	Proc		Drain 4b		
587	B	II	155	2a	B3	17	20	Clean		Drain 6b		
588	B	II	321	2a	C3	7	21	Proc		Drain 1b		
589	B	II	321	2a	C3	7	21	Proc		Drain 4b		
590	B	II	155	2a	UT	17	21	Proc		Drain 6b		
591	B	II	321	2a	C3	7	21	Clean		Drain 1b		
592	B	II	155	2a	B3	17	21	Clean		Drain 6b		
593	B	I	6	4b	A3	40	21			Raw product		

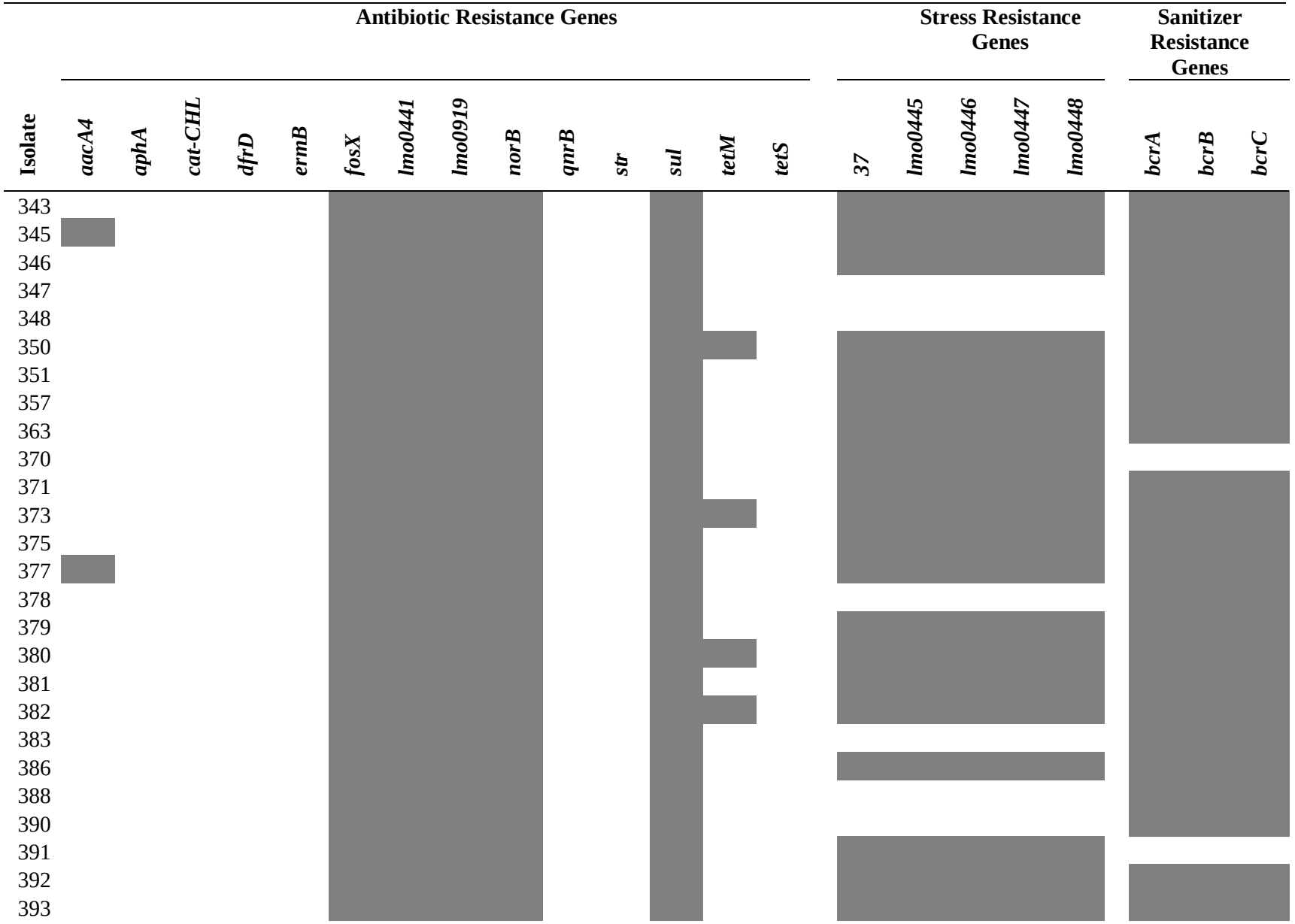
[a] A, processing plant from Berrang, et al. *(8)*; B, processing plant from Berrang, et al. *(7)*
[b] Determined by DNA sequence analysis of *actA*; type designations from plant A are not the same as from plant B

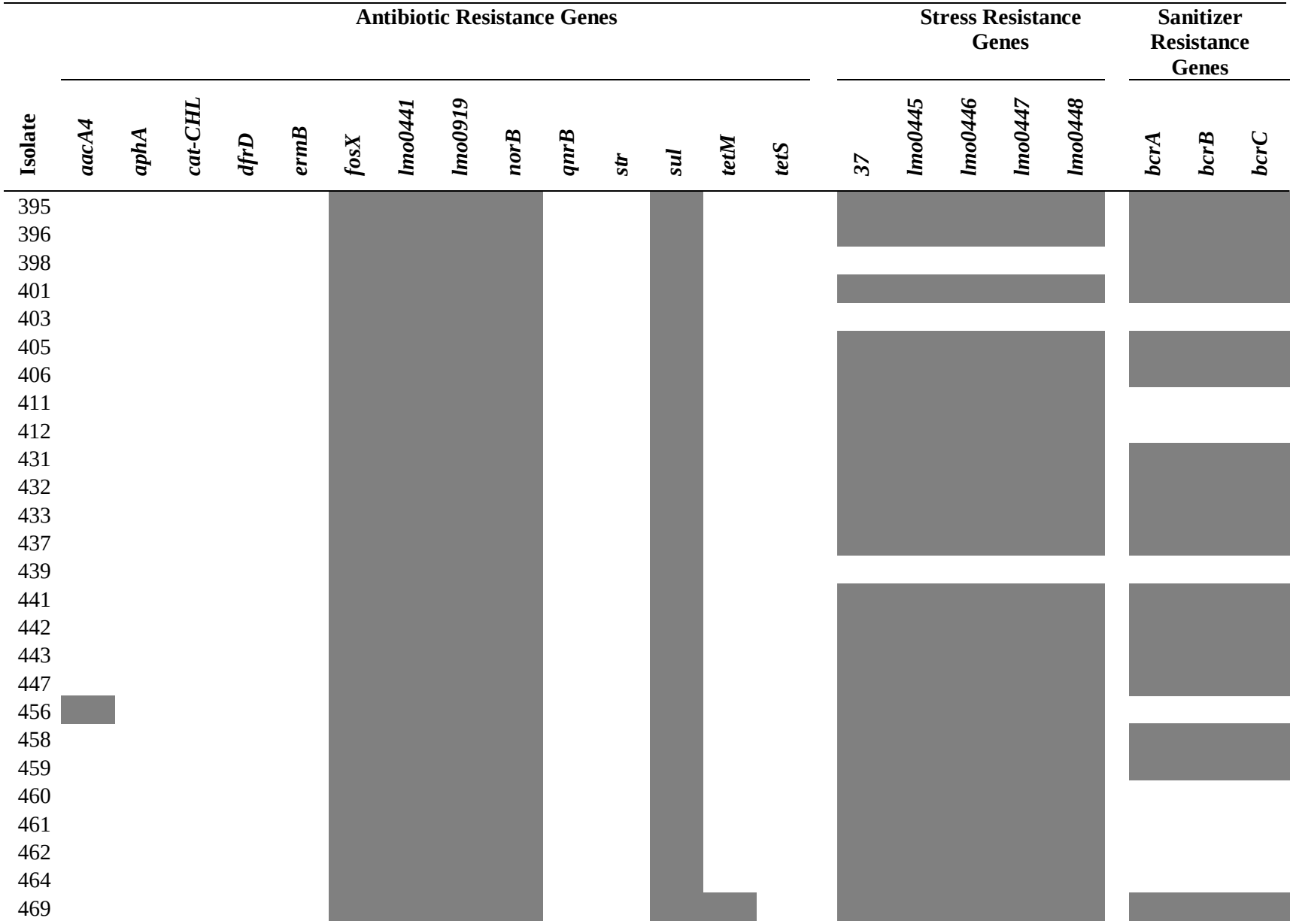
[c] Determined by hqSNP analysis using Lyve-SET
[d] Proc, processing; Clean, cleanup shift
[e] For processing plant A, drains 1, 2, and 6 were located on the raw side of the plant, drains 3, 4, and 5 were located on the cooked side of the plant. For processing plant B, drains 1 through 4 were located on the raw product side of the plant; drains 5 through 9 were located on the cooked product side of the plant
[f] Isolate untypable by methods used

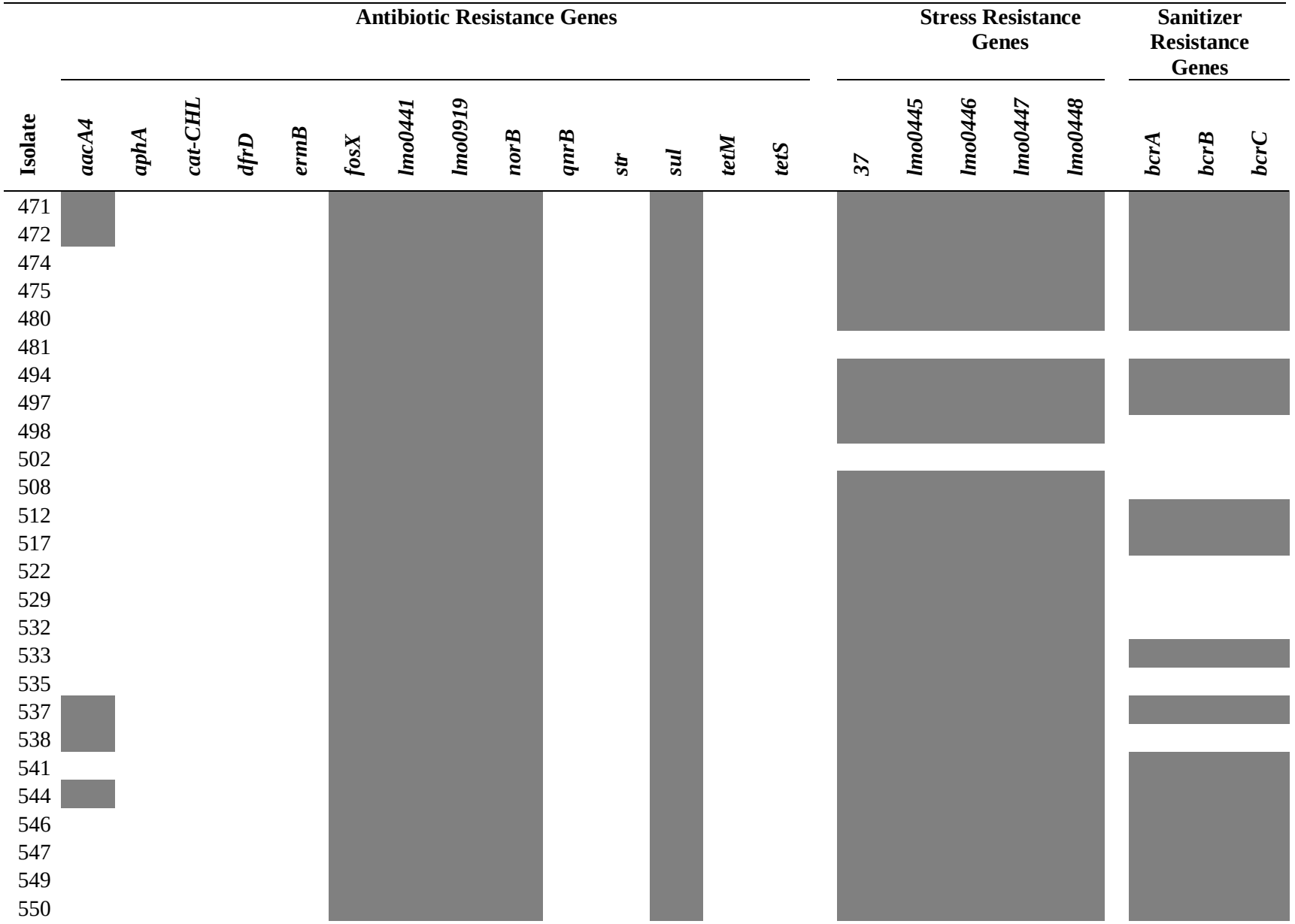
Table 3.2. Antibiotic, stress, and sanitizer/metal resistance genes located in genomes of *L. monocytogenes* isolates from poultry processing plants genomes.

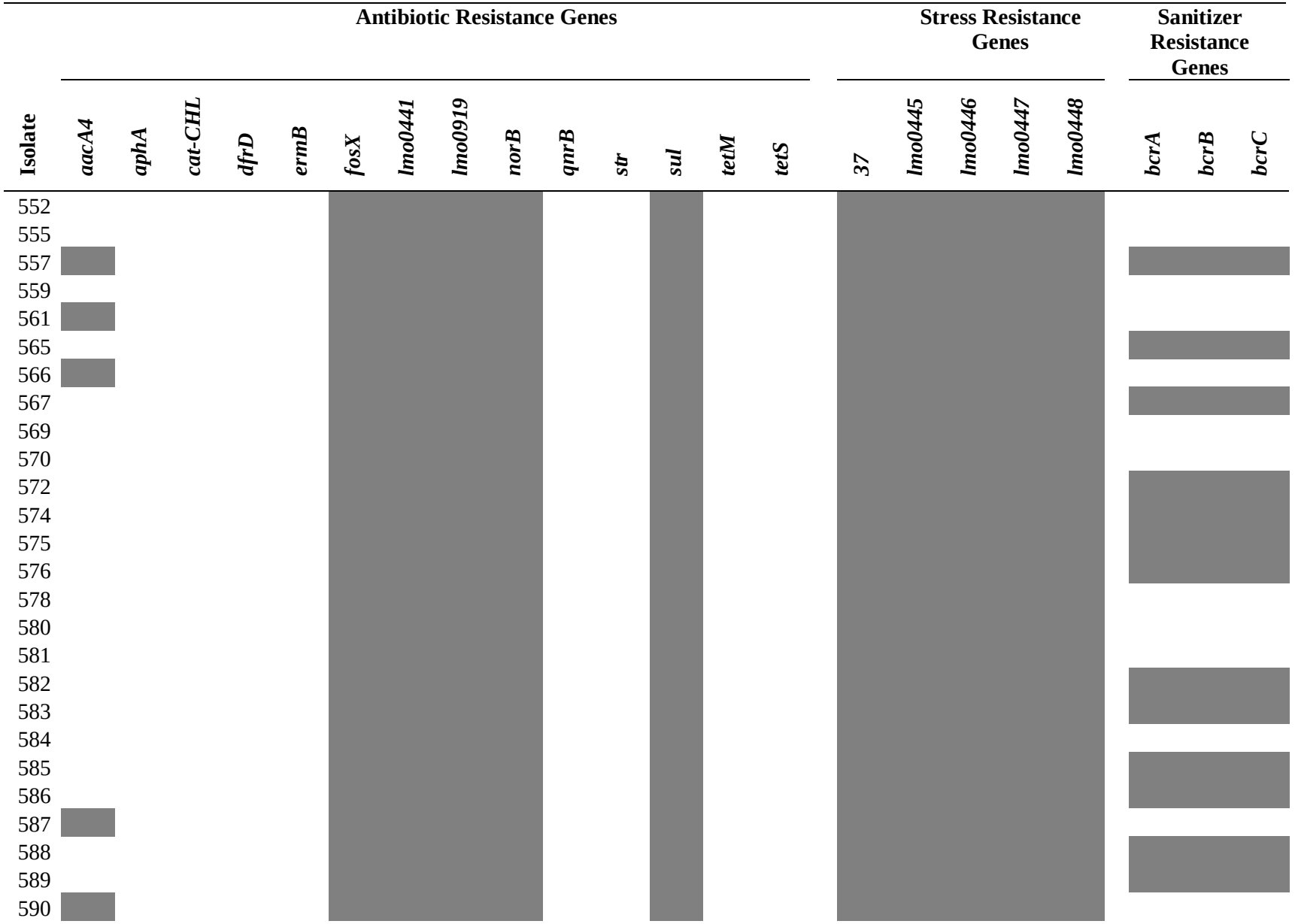


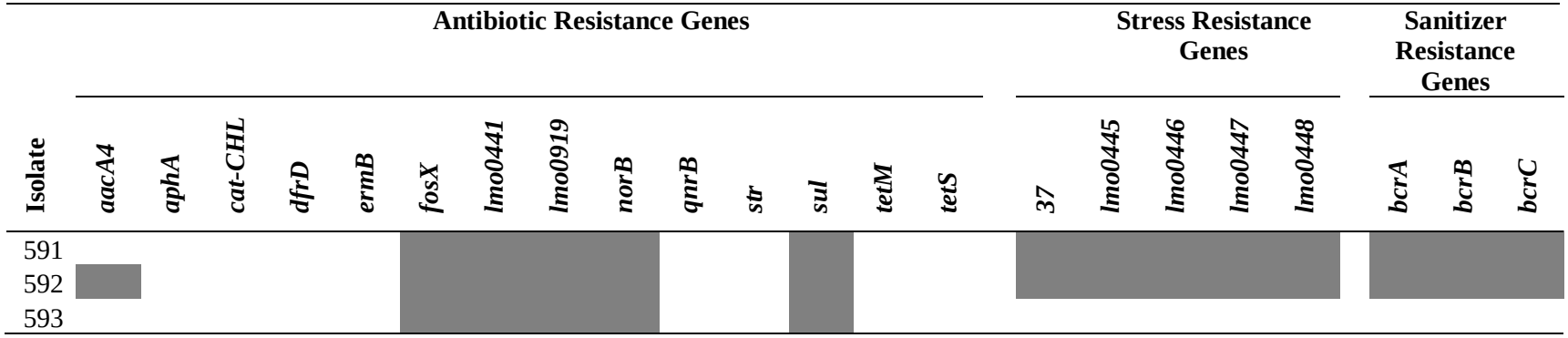

Antibiotic Resistance Genes
Stress Resistance Genes
Sanitizer Resistance Genes
Isolate
aacA4
aphA
cat-CHL
dfrD
ermB
fosX
lmo0441
lmo0919
norB
qnrB
str
sul
tetM
tetS
37
lmo0445
lmo0446
lmo0447
lmo0448
bcrA
bcrB
bcrC
282
283
285
287
288
293
295
296
298
299
301
303
305
307
308
311
325
327
329
330
331
333
337
339
340
341


Antibiotic Resistance Genes
Stress Resistance Genes
Sanitizer Resistance Genes
Isolate
aacA4
aphA
cat-CHL
dfrD
ermB
fosX
lmo0441
lmo0919
norB
qnrB
str
sul
tetM
tetS
37
lmo0445
lmo0446
lmo0447
lmo0448
bcrA
bcrB
bcrC
343
345
346
347
348
350
351
357
363
370
371
373
375
377
378
379
380
381
382
383
386
388
390
391
392
393


Antibiotic Resistance Genes
Stress Resistance Genes
Sanitizer Resistance Genes
Isolate
aacA4
aphA
cat-CHL
dfrD
ermB
fosX
lmo0441
lmo0919
norB
qnrB
str
sul
tetM
tetS
37
lmo0445
lmo0446
lmo0447
lmo0448
bcrA
bcrB
bcrC
395
396
398
401
403
405
406
411
412
431
432
433
437
439
441
442
443
447
456
458
459
460
461
462
464
469


Antibiotic Resistance Genes
Stress Resistance Genes
Sanitizer Resistance Genes
Isolate
aacA4
aphA
cat-CHL
dfrD
ermB
fosX
lmo0441
lmo0919
norB
qnrB
str
sul
tetM
tetS
37
lmo0445
lmo0446
lmo0447
lmo0448
bcrA
bcrB
bcrC
471
472
474
475
480
481
494
497
498
502
508
512
517
522
529
532
533
535
537
538
541
544
546
547
549
550


Antibiotic Resistance Genes
Stress Resistance Genes
Sanitizer Resistance Genes
Isolate
aacA4
aphA
cat-CHL
dfrD
ermB
fosX
lmo0441
lmo0919
norB
qnrB
str
sul
tetM
tetS
37
lmo0445
lmo0446
lmo0447
lmo0448
bcrA
bcrB
bcrC
552
555
557
559
561
565
566
567
569
570
572
574
575
576
578
580
581
582
583
584
585
586
587
588
589
590


Antibiotic Resistance Genes
Stress Resistance Genes
Sanitizer Resistance Genes
Isolate
aacA4
aphA
cat-CHL
dfrD
ermB
fosX
lmo0441
lmo0919
norB
qnrB
str
sul
tetM
tetS
37
lmo0445
lmo0446
lmo0447
lmo0448
bcrA
bcrB
bcrC
591
592
593

A. Lineage I isolates

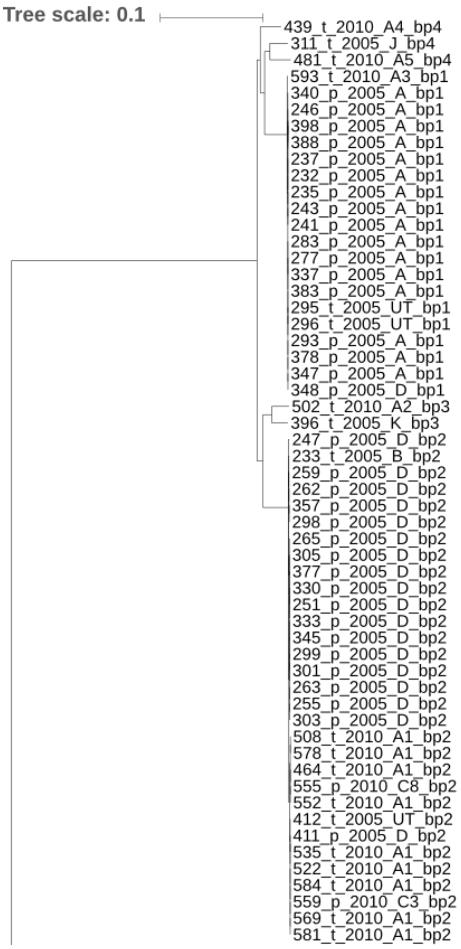


Figure 3.1.A. Phylogenetic tree of *L. monocytogenes* isolates analyzed in this study based on hqSNP analysis using Lyve-SET. The tree is separated into two subfigures to aid in visibility: A contains Lineage I isolates and B contains Lineage II isolates. Leaf names consist of "isolate number_persistent or transient_year of study_*actA* subtype_SNP subtype."

B. Lineage II isolates

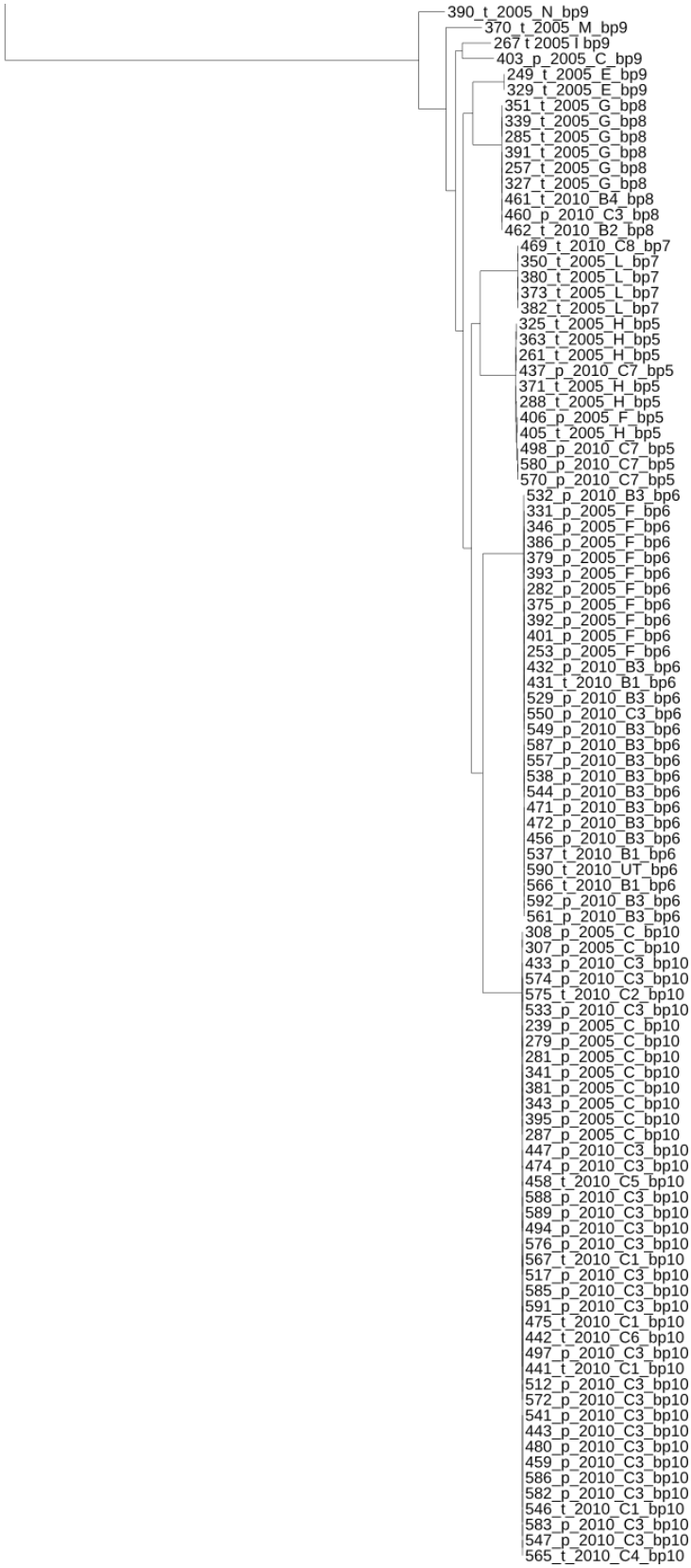


Figure 3.1.B. Phylogenetic tree of *L. monocytogenes* isolates analyzed in this study based on hqSNP analysis using Lyve-SET. The tree is separated into two subfigures to aid in visibility: A contains Lineage I isolates and B contains Lineage II isolates. Leaf names consist of "isolate number_persistent or transient_year of study_*actA* subtype_SNP subtype."

CHAPTER 4

PAN-GENOME-WIDE ASSOCIATION STUDY TO DETERMINE IF GENE CONTENT OF *LISTERIA MONOCYTOGENES* ISOLATES CORRESPONDS TO REPEATED ISOLATION IN POULTRY PROCESSING PLANTS[2]

[2]Lauren K. Hudson, Shaokang Zhang, Mark E. Berrang, Richard J. Meinersmann, Xiangyu Deng, and Mark A. Harrison. To be submitted to *Journal of Food Protection*.

Abstract

Some *Listeria monocytogenes* strains are able to persist in food processing facilities, which can create opportunities to cause food contamination. Much is still unknown about genetic markers or factors that may relate to persistence. In this study, we used whole genome sequencing (WGS) data and subsequent analysis of *L. monocytogenes* isolates from two previous longitudinal studies of poultry processing plants to identify genes significantly positively or negatively associated with repeated isolation in those facilities. We found a total of 352 genes or gene clusters significantly associated with repeated isolation (naive P-value < 0.05 and a Bonferroni-corrected P-value < 0.05), with 180 positively and 172 negatively associated. Some notable genes/gene clusters were annotated as internalin precursors (n=28; *inlJ*, *inlA*, and *inlB*), CRISPR-associated (n=6), and various other genes (*iap*, *cadA*, etc.) Understanding the relationship of these genes to *L. monocytogenes* persistence in food processing environments or as markers to differentiate persistent and transient strains may help to better plan and improve methods used to clean and sanitize food processing environments.

Controlling *Listeria monocytogenes* on ready-to-eat meat and poultry products and in food processing facilities is a challenge. *L. monocytogenes* can be introduced into food processing facilities through many routes, including via soil on items like workers' clothes or shoes, transport equipment, contaminated machinery, animals/pests that either excrete the bacterium or have contaminated external surfaces, raw plant material, raw meat or poultry, or possibly even asymptomatic human carriers *(43)*. Surveys have found that some *L. monocytogenes* subtypes are more persistent in processing facilities than others, but the reason is unknown *(3, 4)*. Persistent strains may vary genetically from transient strains giving them an advantage that possibly aids in their attachment to food contact and non-food contact surfaces and their ability to form biofilms in food processing facilities. Information gained through genetic and phenotypic studies could be used to determine if more effective cleaning and sanitizing agents targeting the more persistent types of *L. monocytogenes* could be developed. In addition, there are currently no methods available to distinguish between persistent and transient strains based on whole-genome sequencing (WGS) data. If specific genes or genetic elements could be clearly linked to persistence or repeated isolation of *L. monocytogenes*, this could be used as a tool to identify putatively persistent strains isolated from food processing facilities.

Stasiewicz, et al. conducted a study in which whole-genome sequencing was used to study 188 *L. monocytogenes* isolates collected in retail delicatessens *(46)*. The goals of that study were to subtype isolates using phylogenetic methods, use sequencing data to differentiate persistent and transient strains, and determine if any genetic causes of persistence exist. They found whole-genome sequencing methods were adequate for subtyping *L. monocytogenes*, but they were unable to find any individual genes that were enriched more in persistent isolates than transient isolates.

The objective of the current study was to identify genetic factors that may be involved in or interact with the regulation of genes that influence the ability of *L. monocytogenes* to thrive in poultry processing environments, on food contact surfaces, and/or on poultry products. We hypothesized that there are differences in genetic factors between persistent and transient *L. monocytogenes* isolates found in poultry processing plants. Whole-genome sequencing and a pan-genome-wide association study (pan-GWAS) was used to determine genetic factors of interest that might contribute to or be correlated with persistence.

Materials and Methods

Previously (Chapter 3), 156 *L. monocytogenes* isolates from two longitudinal sampling studies of poultry further processing plants *(3, 4)* were selected for analysis and sequenced; the whole genome sequencing (WGS) raw reads were trimmed, assembled into contigs, and annotated. The phylogeny of these isolates was determined by hqSNP analysis and divided into subtypes. Subtypes were determined to be either repeatedly isolated or sporadically isolated, based on frequency of isolation in the original sampling studies (if a particular subtype was found more than once on separate sampling times in the same processing plant, it was classified as repeatedly isolated). The isolation frequency phenotypes and annotations were used in the current study for a pan-genome-wide association study (pan-GWAS).

Pan-genome-wide association study. A pan-GWAS was performed to compare gene content among the sequenced strains using Roary (version 1.006924; with Prokka annotation output file previously described used as input file) *(36)* and statistical analysis was done using Scoary (version 1.6.11) *(6)* to identify genes or markers associated with repeated isolation of a subtype within the genomes. Genes predominantly present or absent among repeatedly isolated subtypes were identified. Genes were considered significantly associated with an isolate being

repeatedly isolated if they had a naive P-value < 0.05 and a Bonferroni-corrected P-value < 0.05. From the pan-GWAS, the positively and negatively associated genes were classified by having an Odds Ratio of > 1 and < 1, respectively.

Results

A total of 5,622 genes (pangenome) were identified from all the analyzed *L. monocytogenes* genomes. The core genome (found in at least 99% of genomes) contained 2,423 genes. The accessory genome of this population was composed of 42 soft core genes (found in 95-99% of genomes), 1,015 shell genes (found in 15-95% of genomes), and 2,142 cloud genes (found in less than 15% of genomes).

We found that 1,033 individual gene clusters in the genomes of the isolates studied were significantly (naive p<0.05) either positively (n=405) or negatively (n=540) associated with repeated isolation in these two poultry further processing plants. The genes can be seen in Table 4.1, listed by statistical significance (Bonferroni-corrected P-value) with repeated isolation. Of these 1,033 genes, 286 unique annotations were available, based on the annotations output from Prokka, and 569 of the annotations were "hypothetical protein."

Three-hundred-fifty-two gene clusters met the condition of having a naive P-value < 0.05 and a Bonferroni-corrected P-value < 0.05. Of these, 180 were positively associated with repeated isolation (including gene clusters with an odds ratio of infinity) and 172 were negatively associated with repeated isolation, and 118 unique annotations were available (Table 4.1), based on the annotations output from Prokka, and 180 of the annotations were "hypothetical protein." A large portion of the gene clusters were annotated as internalin precursors, internalin-J precursor (n=17), internalin-A precursor (n=9), and internalin-B precursor (n=2), with similar numbers being positively (n=13) and negatively (n=15) associated with repeated isolation. Six

CRISPR-associated genes were also in the list of genes meeting p-value requirements, with most of them (n=5) being positively associated with repeated isolation.

Discussion

It is thought that some *L. monocytogenes* isolates have the ability to persist in food processing facilities better than other isolates *(1, 3, 4, 8, 9, 11, 13, 15, 16, 18, 20-22, 24-28, 30-32, 37, 41, 42, 45, 46, 52, 53, 55, 56)*. Conclusions from some studies have been somewhat limited due to small sample size. Speculation about what genetic differences may exist between strains that might be more persistence and those that are not has varied *(13, 17, 19, 38, 39, 54)*. The current study was focused on the genetic composition of 156 *L. monocytogenes* isolates collected from two different further processed poultry facilities. We hoped to determine if there were commonalities in the genomes of isolates that could be used to predict whether or not they might be more or less persistent in a food processing facility.

Theoretically, some strains of *L. monocytogenes* may be able to survive and persist in food processing facilities due to relevant phenotypic traits, such as better attachment to surfaces and biofilm-forming ability, increased resistance to stresses like heat, drying and desiccation, and pH, enhanced resistance to sanitizing and disinfecting agents, etc. *(13, 18, 22, 57)*. Studies have previously identified genes associated with some of these abilities. Piercey, et al. found that interruptions in genes encoding cell wall biosynthesis, motility, metabolism, stress response, and cell surface associated proteins resulted in increased biofilm formation and discovered that interruptions of 9 genes not previously associated with biofilm formation in *L. monocytogenes* (*lmo2572*, *lmo2488* [*uvrA*], *lmo1224*, *lmo0434* [*inlB*], *lmo0263* [*inlH*], *lmo0543*, *lmo0057* [*EsaA*], *lmo2563*, and *lmo0453*) resulted in increased biofilm formation *(39)*. Of these, only *inlB* was found to be significantly associated with repeated isolation (along with *inlA* and *inlJ*), with

two groups being positively associated and one negatively associated. Internalin proteins are cell wall components that contain leucine-rich repeat (LLR) regions and play a role in protein binding functions *(2, 44, 47, 48, 50, 51)*.

CRISPR (clustered regulatory interspaced short palindromic repeats) systems allow some bacteria to protect themselves from bacteriophage activity *(10)*, and associated genes were among those significantly associated. *clpP*, another gene that was significantly enriched in repeatedly isolated strains, encodes endopeptidase Clp ATP-binding chain C and is involved in stress survival (including stresses like low pH, iron limitation, elevated temperature and salt stress), intracellular growth, and growth at high temperature *(14, 33, 40)*. *CadA*, which we found to be positively associated with the repeatedly isolated phenotype, encodes a cadmium-transporting ATPase and is associated with cadmium tolerance *(58)*.

Previous studies have shown that some other cell membrane-associated proteins have been associated with the ability of *L. monocytogenes* to survive in the environment, including ones with domains involved in cell-wall anchoring or protein-protein interactions, like LPXTG, GW, P60, LysM, lipo-box, and LLR *(5, 23)*. *iap*, which encodes a putative endopeptidase P60 precursor involved in cell wall catabolic processes *(49, 59)*, was one of the genes among the top repeated isolation associated genes identified in the current study. A gene encoding a LysM domain protein (Prokka group 4121) also appeared in the significantly associated genes in the current study. The LysM domain is involved in binding peptidoglycan *(7, 29, 34)*.

Other studies have not shown any genes significantly associated with the persistence or transience of *L. monocytogenes* isolates *(11, 12, 35, 46)*. One potential reason for this is that many previous studies have only analyzed small sample sizes of *L. monocytogenes* isolates. Previous findings indicate that persistence phenotypes may not be caused by single mutations,

phages, or other genetic differences, but may be unique to distinct *L. monocytogenes* lineages and/or due to more complex genetic elements, phenotypic stimuli, or gene regulation systems *(35)*.

Fagerlund, et al. *(11)* sequenced and analyzed five MLST sequence type-8 *L. monocytogenes* isolates from Norwegian salmon and poultry processing facilities. That study was narrow, as it only focused on one MLST sequence type, which was previously shown to be persistent in food processing facilities, and only included five isolates. Fagerlund, et al. concluded that WGS and SNP-based analysis of *L. monocytogenes* are well suited to evaluate persistence of *L. monocytogenes* isolates in food processing facilities. In addition, they compared the accessory genome of one of the ST-8 strains to two commonly used laboratory strains (EGD-e and 10403S) and found unique sections that contained of 50 ORFs, distributed in ten genomic loci. These loci contained two R-M systems (Type I and III) and encoded proteins with the following predicted functions: abortive resistance to bacteriophage infection, putative ATP-binding cassette, carbohydrate transport, and metallopeptidase.

Nowak, et al. evaluated persistent *L. monocytogenes* strains isolated from New Zealand mussel production facilities to find specific genes or genetic markers that may be linked to persistence or non-persistence *(35)*. They were unable to find any markers in either the persistent or non-persistent isolate groups. In addition, they found that there was no overall clustering of persistent or sporadic isolates and that differences in prophages and plasmids were not associated with persistence. As with other studies, a limitation of this study is the small number of *L. monocytogenes* isolates evaluated (8 persistent and 8 sporadic).

Stasiewicz, et al. analyzed *L. monocytogenes* isolates from deli environments to identify genes significantly enriched among persistent or sporadic groups *(46)*. Due to the fact that the *L.*

monocytogenes isolate sample size for this study was large, it has more power than many other studies with the goal of identifying genes associated with *L. monocytogenes* persistence, as many have a very small sample size. However, gene enrichment analysis failed to find any genes that were enriched among isolates initially (putatively) identified as either persistent (n=92) or sporadic (n =29) or when comparing isolates with statistical evidence for persistence (n=38) to sporadic isolates of the same subtypes (n=16). They did find 10 genes that were significantly enriched among non-persistent isolates; however, these genes were not present in a conserved region and were not annotated with functions associated with persistence or survival in a harsh environment. They concluded that their findings suggest there are no individual gene presence or absence patterns linked to the persistent strains analyzed in their study *(46)*.

Future research should involve creation of gene knockout mutants and the study of phenotypic features of persistent isolates to determine biological significance of particular genetic factors. For example, Nowak, et al. found that persistent *L. monocytogenes* isolates from mussel production facilities exhibited significantly higher biofilm formation after 48 h using a cell enumeration measurement method and near significant difference using the crystal violet assay (both at 30°C) *(35)*. In addition, there may be genetic regulatory factors involved in persistence (as opposed to just the presence/absence of genes). Analyzing larger and more diverse *L. monocytogenes* isolate data sets may also aid in pinpointing specific genetic markers associated with persistent strains.

Acknowledgements

This material is based upon work supported by Georgia Agricultural Experiment Stations, the University of Georgia Center for Food Safety, USDA's National Needs Fellowship program,

and the University of Georgia Graduate School. Thanks to Steven Knapp, Eric Adams, and Linda Genzlinger for technical assistance.

References

1. Autio, T., R. Keto-Timonen, J. Lunden, J. Bjorkroth, and H. Korkeala. 2003. Characterisation of persistent and sporadic *Listeria monocytogenes* strains by pulsed-field gel electrophoresis (PFGE) and amplified fragment length polymorphism (AFLP). *Syst. Appl. Microbiol.* 26:539-545.

2. Bergmann, B., D. Raffelsbauer, M. Kuhn, M. Goetz, S. Hom, and W. Goebel. 2002. InlA- but not InlB-mediated internalization of *Listeria monocytogenes* by non-phagocytic mammalian cells needs the support of other internalins. *Mol. Microbiol.* 43:557-570.

3. Berrang, M. E., R. J. Meinersmann, J. F. Frank, and S. R. Ladely. 2010. Colonization of a newly constructed commercial chicken further processing plant with *Listeria monocytogenes*. *J. Food Prot.* 73:286-291.

4. Berrang, M. E., R. J. Meinersmann, J. F. Frank, D. P. Smith, and L. L. Genzlinger. 2005. Distribution of *Listeria monocytogenes* subtypes within a poultry further processing plant. *J. Food Prot.* 68:980-985.

5. Bierne, H., and P. Cossart. 2007. *Listeria monocytogenes* surface proteins: From genome predictions to function. *Microbiol. Mol. Biol. Rev.* 71:377-397.

6. Brynildsrud, O., J. Bohlin, L. Scheffer, and V. Eldholm. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with scoary. *Genome Biol.* 17:238.

7. Buist, G., A. Steen, J. Kok, and O. P. Kuipers. 2008. LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol. Microbiol.* 68:838-47.

8. Carpentier, B., and O. Cerf. 2011. Review--persistence of *Listeria monocytogenes* in food industry equipment and premises. *Int. J. Food Microbiol.* 145:1-8.

9. Cruz, C. D., and G. C. Fletcher. 2011. Prevalence and biofilm-forming ability of *Listeria monocytogenes* in New Zealand mussel (*perna canaliculus*) processing plants. *Food Microbiol.* 28:1387-1393.

10. Di, H. L., L. Ye, H. Yan, H. C. Meng, S. Yamasak, and L. Shi. 2014. Comparative analysis of CRISPR loci in different *Listeria monocytogenes* lineages. *Biochem. Biophys. Res. Commun.* 454:399-403.

11. Fagerlund, A., S. Langsrud, B. C. T. Schirmer, T. Moretro, and E. Heir. 2016. Genome analysis of *Listeria monocytogenes* sequence type 8 strains persisting in salmon and poultry processing environments and comparison with related strains. *PLoS One*. 11:e0151117.

12. Fang, C., T. Cao, Y. Shan, Y. Xia, Y. P. Xin, C. Y. Cheng, H. H. Song, J. Bowman, X. L. Li, X. Y. Zhou, and W. H. Fang. 2016. Comparative genomic analysis reveals that the 20k and 38k prophages in *Listeria monocytogenes* serovar 4a strains lm850658 and m7 contribute to genetic diversity but not to virulence. *J. Microbiol. Biotechnol.* 26:197-206.

13. Ferreira, V., M. Wiedmann, P. Teixeira, and M. J. Stasiewicz. 2014. *Listeria monocytogenes* persistence in food-associated environments: Epidemiology, strain characteristics, and implications for public health. *J. Food Prot.* 77:150-170.

14. Gahan, C. G. M., and C. Hill. 1999. The relationship between acid stress responses and virulence in *Salmonella* Typhimurium and *Listeria monocytogenes. Int. J. Food Microbiol.* 50:93-100.

15. Harvey, J., and A. Gilmour. 1994. Application of multilocus enzyme electrophoresis and restriction fragment length polymorphism analysis to the typing of *Listeria monocytogenes* strains isolated from raw milk, nondairy foods, and clinical and veterinary sources. *Appl. Environ. Microbiol.* 60:1547-1553.

16. Heir, E., B.-A. Lindstedt, O.-J. Røtterud, T. Vardund, G. Kapperud, and T. Nesbakken. 2004. Molecular epidemiology and disinfectant susceptibility of *Listeria monocytogenes* from meat processing plants and human infections. *Int. J. Food Microbiol.* 96:85-96.

17. Hingston, P. A., M. J. Piercey, and L. Truelstrup Hansen. 2015. Genes associated with desiccation and osmotic stress in *Listeria monocytogenes* as revealed by insertional mutagenesis. *Appl. Environ. Microbiol.* 81:5350-5362.

18. Holah, J., J. Taylor, D. Dawson, and K. Hall. 2002. Biocide use in the food industry and the disinfectant resistance of persistent strains of *Listeria monocytogenes* and *Escherichia coli. J. Appl. Microbiol.* 92:111S-120S.

19. Holch, A., K. Webb, O. Lukjancenko, D. Ussery, B. M. Rosenthal, and L. Gram. 2013. Genome sequencing identifies two nearly unchanged strains of persistent *Listeria monocytogenes* isolated at two different fish processing plants sampled 6 years apart. *Appl. Environ. Microbiol.* 79:2944-2951.

20. Keto-Timonen, R., R. Tolvanen, J. Lunden, and H. Korkeala. 2007. An 8-year surveillance of the diversity and persistence of *Listeria monocytogenes* in a chilled food processing plant analyzed by amplified fragment length polymorphism. *J. Food Prot.* 70:1866-1873.

21. Knudsen, G. M., J. B. Nielsen, R. L. Marvig, Y. Ng, P. Worning, H. Westh, and L. Gram. 2017. Genome-wide-analyses of *Listeria monocytogenes* from food-processing plants reveal clonal diversity and date the emergence of persisting sequence types. *Environ. Microbiol. Rep.* 9:428-440.

22. Korenova, J., K. Oravcova, A. Veghova, R. Karpiskova, and T. Kuchta. 2016. Biofilm formation in various conditions is not a key factor of persistence potential of *Listeria monocytogenes* in food-processing environment. *J. Food Nutr. Res.* 55:189-193.

23. Kuenne, C., A. Billion, M. A. Mraheil, A. Strittmatter, R. Daniel, A. Goesmann, S. Barbuddhe, T. Hain, and T. Chakraborty. 2013. 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.

24. Leong, D., K. NicAogáin, L. Luque-Sastre, O. McManamon, K. Hunt, A. Alvarez-Ordóñez, J. Scollard, A. Schmalenberger, S. Fanning, C. O'Byrne, and K. Jordan. 2017. A 3-year multi-food study of the presence and persistence of *Listeria monocytogenes* in 54 small food businesses in Ireland. *Int. J. Food Microbiol.* 249:18-26.

25. Lunden, J., T. Autio, A. Markkula, S. Hellstrom, and H. Korkeala. 2003. Adaptive and cross-adaptive responses of persistent and non-persistent *Listeria monocytogenes* strains to disinfectants. *Int. J. Food Microbiol.* 82:265-272.

26. Lunden, J., R. Tolvanen, and H. Korkeala. 2008. Acid and heat tolerance of persistent and nonpersistent *Listeria monocytogenes* food plant strains. *Lett. Appl. Microbiol.* 46:276-280.

27. Lunden, J. M., T. J. Autio, A. M. Sjoberg, and H. J. Korkeala. 2003. Persistent and nonpersistent *Listeria monocytogenes* contamination in meat and poultry processing plants. *J. Food Prot.* 66:2062-2069.

28. Lunden, J. M., M. K. Miettinen, T. J. Autio, and H. J. Korkeala. 2000. Persistent *Listeria monocytogenes* strains show enhanced adherence to food contact surface after short contact times. *J. Food Prot.* 63:1204-1207.

29. Marchler-Bauer, A., Y. Bo, L. Han, J. He, C. J. Lanczycki, S. Lu, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng, L. Y. Geer, and S. H. Bryant. 2017. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 45:D200-D203.

30. Møretrø, T., and S. Langsrud. 2004. *Listeria monocytogenes*: Biofilm formation and persistence in food-processing environments. *Biofilms*. 1:107-121.

31. Morganti, M., E. Scaltriti, P. Cozzolino, L. Bolzoni, G. Casadei, M. Pierantoni, E. Foni, and S. Pongolini. 2016. Processing-dependent and clonal contamination patterns of *Listeria monocytogenes* in the cured ham food chain revealed by genetic analysis. *Appl. Environ. Microbiol.* 82:822-831.

32. Murugesan, L., Z. Kucerova, S. J. Knabel, and L. F. LaBorde. 2015. Predominance and distribution of a persistent *Listeria monocytogenes* clone in a commercial fresh mushroom processing environment. *J. Food Prot.* 78:1988-1998.

33. Nair, S., I. Derré, T. Msadek, O. Gaillot, and P. Berche. 2000. Ctsr controls class iii heat shock gene expression in the human pathogen *Listeria monocytogenes. Mol. Microbiol.* 35:800-811.

34. NCBI. 2018. Conserved protein domain family - LysM. Available at: https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=cd00118. Accessed 23 April 2018.

35. Nowak, J., C. D. Cruz, M. Tempelaars, T. Abee, A. H. M. van Vliet, G. C. Fletcher, D. Hedderley, J. Palmer, and S. Flint. 2017. Persistent *Listeria monocytogenes* strains isolated from mussel production facilities form more biofilm but are not linked to specific genetic markers. *Int. J. Food Microbiol.* 256:45-53.

36. Page, A. J., C. A. Cummins, M. Hunt, V. K. Wong, S. Reuter, M. T. Holden, M. Fookes, D. Falush, J. A. Keane, and J. Parkhill. 2015. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics*. 31:3691-3693.

37. Palma, F., F. Pasquali, A. Lucchi, A. De Cesare, and G. Manfreda. 2017. Whole genome sequencing for typing and characterisation of *Listeria monocytogenes* isolated in a rabbit meat processing plant. *Ital. J. Food Saf.* 6:6879.

38. Piercey, M. J., T. C. Ells, A. J. Macintosh, and L. Truelstrup Hansen. 2017. Variations in biofilm formation, desiccation resistance and benzalkonium chloride susceptibility among *Listeria monocytogenes* strains isolated in canada. *Int. J. Food Microbiol.* 257:254-261.

39. Piercey, M. J., P. A. Hingston, and L. Truelstrup Hansen. 2016. Genes involved in *Listeria monocytogenes* biofilm formation at a simulated food processing plant temperature of 15°C. *Int. J. Food Microbiol.* 223:63-74.

40. Ripio, M. T., J. A. Vázquez-Boland, Y. Vega, S. Nair, and P. Berche. 1998. Evidence for expressional crosstalk between the central virulence regulator prfa and the stress response mediator clpc in *Listeria monocytogenes. FEMS Microbiol. Lett.* 158:45-50.

41. Rorvik, L. M., B. Aase, T. Alvestad, and D. A. Caugant. 2003. Molecular epidemiological survey of *Listeria monocytogenes* in broilers and poultry products. *J. Appl. Microbiol.* 94:633-640.

42. Rørvik, L. M., D. A. Caugant, and M. Yndestad. 1995. Contamination pattern of *Listeria monocytogenes* and other *Listeria* spp. In a salmon slaughterhouse and smoked salmon processing plant. *Int. J. Food Microbiol.* 25:19-27.

43. Ryser, E. T., and R. L. Buchanan. 2013. *Listeria monocytogenes.* p. 503-545. *In* M.P. Doyle, and R.L. Buchanan (ed.), Food microbiology: Fundamentals and frontiers 4th ed. ASM Press, Washington, D.C.

44. Sabet, C., M. Lecuit, D. Cabanes, P. Cossart, and H. Bierne. 2005. LPXTG protein InlJ, a newly identified internalin involved in *Listeria monocytogenes* virulence. *Infect. Immun.* 73:6912-6922.

45. Senczek, D., R. Stephan, and F. Untermann. 2000. Pulsed-field gel electrophoresis (PFGE) typing of *Listeria* strains isolated from a meat processing plant over a 2-year period. *Int. J. Food Microbiol.* 62:155-159.

46. Stasiewicz, M. J., H. F. Oliver, M. Wiedmann, and H. C. den Bakker. 2015. Whole-genome sequencing allows for improved identification of persistent *Listeria monocytogenes* in food-associated environments. *Appl. Environ. Microbiol.* 81:6024-6037.

47. The UniProt Consortium. 2017. Uniprot: The universal protein knowledgebase. *Nucleic Acids Res.* 45:D158-D169.

48. The UniProt Consortium. 2018. UniProtKB - G2K3G6 (INLA_LISM4). Available at: http://www.uniprot.org/uniprot/G2K3G6. Accessed 23 April 2018.

49. The UniProt Consortium. 2018. UniProtKB - P21171 (P60_LISMO). Available at: http://www.uniprot.org/uniprot/P21171. Accessed 23 April 2018.

50. The UniProt Consortium. 2018. UniProtKB - P25147 (INLB_LISMO). Available at: http://www.uniprot.org/uniprot/P25147. Accessed 23 April 2018.

51. The UniProt Consortium. 2018. UniProtKB - Q8Y3L4 (INLJ_LISMO). Available at: http://www.uniprot.org/uniprot/Q8Y3L4. Accessed 23 April 2018.

52. Vangay, P., J. Steingrimsson, M. Wiedmann, and M. J. Stasiewicz. 2014. Classification of *Listeria monocytogenes* persistence in retail delicatessen environments using expert elicitation and machine learning. *Risk Anal.* 34:1830-1845.

53. Verghese, B., M. Lok, J. Wen, V. Alessandria, Y. Chen, S. Kathariou, and S. Knabel. 2011. Comk prophage junction fragments as markers for *Listeria monocytogenes* genotypes unique to individual meat and poultry processing plants and a model for rapid niche-specific adaptation, biofilm formation, and persistence. *Appl. Environ. Microbiol.* 77:3279-3292.

54. Vongkamjan, K., S. Benjakul, H. T. Kim Vu, and V. Vuddhakul. 2017. Longitudinal monitoring of *Listeria monocytogenes* and *Listeria* phages in seafood processing environments in Thailand. *Food Microbiol.* 66:11-19.

55. Wang, J. J., A. J. Ray, S. R. Hammons, and H. F. Oliver. 2015. Persistent and transient *Listeria monocytogenes* strains from retail deli environments vary in their ability to adhere and form biofilms and rarely have inla premature stop codons. *Foodborne Pathog. Dis.* 12:151-158.

56. Williams, S. K., S. Roof, E. A. Boyle, D. Burson, H. Thippareddi, I. Georgnaras, J. N. Sofos, M. Wiedmann, and K. Nightingale. 2011. Molecular ecology of *Listeria monocytogenes* and other *Listeria* species in small and very small ready-to-eat meat processing plants. *J. Food Prot.* 74:63-77.

57. Wulff, G., L. Gram, P. Ahrens, and B. F. Vogel. 2006. One group of genetically similar *Listeria monocytogenes* strains frequently dominates and persists in several fish slaughter-and smokehouses. *Appl. Environ. Microbiol.* 72:4313-4322.

58. Xu, D., Q. Nie, W. Wang, L. Shi, and H. Yan. 2016. Characterization of a transferable *bcrabc* and *cadac* genes-harboring plasmid in *Listeria monocytogenes* strain isolated from food products of animal origin. *Int. J. Food Microbiol.* 217:117-122.

59. Yoshikawa, Y., Y. Ochiai, M. Mochizuki, O. Fujita, T. Takano, R. Hondo, and F. Ueda. 2016. Genetic subtyping of *Listeria monocytogenes* via multiple-locus sequence typing using *iap*, *sigB* and *actA. J. Vet. Med. Sci.* 78:1831-1839.

Table 4.1. Pan-genome-wide analysis of genes associated with repeated isolation of *L. monocytogenes* in poultry further processing plants results (genes and gene clusters listed in order of strength of association with repeated isolation phenotype). Includes gene names (or Prokka cluster ID), non-unique gene names (if available), annotation (from Prokka), number of isolates each gene/group is present and absent in (for both repeatedly isolated and non-repeatedly isolated groups), sensitivity, specificity, odds ratio, naive p-value, Bonferroni-corrected p-value, Benjamini-Hochberg-corrected p-value, maximum pairwise comparisons, maximum supporting pairs, maximum opposing pairs, best pairwise comparison p-value, and worst pairwise comparison p-value.

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2427		hypothetical protein	83	3	34	36	70.94	92.31	29.294	1.75E-12	5.75E-09	5.75E-09	4	4	1	0.125	0.625
group 218		hypothetical protein	36	34	81	5	30.77	12.82	0.065	4.68E-10	1.54E-06	7.69E-07	4	3	3	0.625	0.625
recD		RecBCD enzyme subunit RecD	62	1	55	38	52.99	97.44	42.836	2.23E-09	7.34E-06	2.45E-06	4	3	1	0.625	0.625
cas2		CRISPR-associated endoribonuclease Cas2	73	4	44	35	62.39	89.74	14.517	6.84E-09	2.25E-05	2.50E-06	3	2	3	0.250	1.000
cas3		CRISPR-associated nuclease/helicase Cas3	73	4	44	35	62.39	89.74	14.517	6.84E-09	2.25E-05	2.50E-06	3	2	3	0.250	1.000
cas1		CRISPR-associated endonuclease Cas1	73	4	44	35	62.39	89.74	14.517	6.84E-09	2.25E-05	2.50E-06	3	2	3	0.250	1.000
group 3124		hypothetical protein	73	4	44	35	62.39	89.74	14.517	6.84E-09	2.25E-05	2.50E-06	3	2	3	0.250	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 913		CRISPR associated protein Cas6	73	4	44	35	62.39	89.74	14.517	6.84E-09	2.25E-05	2.50E-06	3	2	3	0.250	1.000
group 3125		CRISPR-associated protein (Cas_Cas5)	73	4	44	35	62.39	89.74	14.517	6.84E-09	2.25E-05	2.50E-06	3	2	3	0.250	1.000
macB 2		Macrolide export ATP-binding/permease protein MacB	98	13	19	26	83.76	66.67	10.316	8.73E-09	2.87E-05	2.87E-06	6	6	1	0.031	0.219
group 3062		hypothetical protein	87	9	30	30	74.36	76.92	9.667	2.38E-08	7.82E-05	7.11E-06	5	4	2	0.375	1.000
group 3063		hypothetical protein	86	9	31	30	73.50	76.92	9.247	3.63E-08	1.19E-04	9.94E-06	3	3	1	0.250	1.000
group 914		hypothetical protein	65	3	52	36	55.56	92.31	15.000	5.50E-08	1.81E-04	1.39E-05	4	3	2	0.625	1.000
group 895		hypothetical protein	88	10	29	29	75.21	74.36	8.800	6.51E-08	2.14E-04	1.43E-05	3	3	1	0.250	1.000
group 1068	inlJ_2	Internalin-J precursor	29	29	88	10	24.79	25.64	0.114	6.51E-08	2.14E-04	1.43E-05	3	1	3	0.250	1.000
group 2436		hypothetical protein	68	4	49	35	58.12	89.74	12.143	8.04E-08	2.64E-04	1.65E-05	2	2	2	0.500	0.500
group 3134		hypothetical protein	87	10	30	29	74.36	74.36	8.410	9.94E-08	3.27E-04	1.82E-05	4	3	2	0.625	1.000
group 3479		hypothetical protein	59	2	58	37	50.43	94.87	18.819	9.96E-08	3.28E-04	1.82E-05	4	4	2	0.125	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3061		hypothetical protein	46	0	71	39	39.32	100.00	Inf[a]	1.17E-07	3.85E-04	2.03E-05	1	1	0	1.000	1.000
group 293		Replication initiation and membrane attachment	70	5	47	34	59.83	87.18	10.128	2.36E-07	7.75E-04	2.37E-05	7	5	4	0.453	1.000
group 451		Leucine Rich repeats (2 copies)	47	34	70	5	40.17	12.82	0.099	2.36E-07	7.75E-04	2.37E-05	4	4	1	0.125	0.625
group 3472		hypothetical protein	57	2	60	37	48.72	94.87	17.575	2.55E-07	8.39E-04	2.37E-05	4	3	2	0.625	1.000
mrr		Mrr restriction system protein	57	2	60	37	48.72	94.87	17.575	2.55E-07	8.39E-04	2.37E-05	4	3	2	0.625	1.000
group 2118	inlA_9	Internalin-A precursor	29	28	88	11	24.79	28.21	0.129	2.56E-07	8.42E-04	2.37E-05	2	1	2	0.500	1.000
inlA 6		Internalin-A precursor	88	11	29	28	75.21	71.79	7.724	2.56E-07	8.42E-04	2.37E-05	2	2	1	0.500	1.000
group 2443		hypothetical protein	44	0	73	39	37.61	100.00	inf	2.76E-07	9.06E-04	2.37E-05	3	3	0	0.250	0.250
group 2444		hypothetical protein	44	0	73	39	37.61	100.00	inf	2.76E-07	9.06E-04	2.37E-05	2	2	0	0.500	0.500
topB 2	lmo2756	DNA topoisomerase 3	44	0	73	39	37.61	100.00	inf	2.76E-07	9.06E-04	2.37E-05	3	3	0	0.250	0.250
yobL		Bacillus transposase protein	66	38	51	1	56.41	2.56	0.034	3.11E-07	1.02E-03	2.37E-05	3	1	3	0.250	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 356	pspA	Phosphoserine phosphatase 1	18	23	99	16	15.38	41.03	0.126	3.72E-07	1.22E-03	2.37E-05	5	3	4	0.375	1.000
group 182		hypothetical protein	74	39	43	0	63.25	0.00	0.000	5.35E-07	1.76E-03	2.37E-05	2	0	2	0.500	0.500
group 185		hypothetical protein	74	39	43	0	63.25	0.00	0.000	5.35E-07	1.76E-03	2.37E-05	2	0	2	0.500	0.500
group 3797		Bacteriophage Gp15 protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	4	4	0	0.125	0.125
group 916		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3174		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
iap 4		putative endopeptidase p60 precursor	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 2446		chromosome segregation protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3166		Type IV secretion system protein VirB11	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 2448		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3153		AAA-like domain protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3150		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3155		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3172		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3167		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3156		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3173		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3171		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3169		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
traG		Conjugal transfer protein TraG	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3168		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3151		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3175		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3164		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3170		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3165		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3176		CAAX amino terminal protease self- immunity	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3152		hypothetical protein	43	0	74	39	36.75	100.00	inf	5.35E-07	1.76E-03	2.37E-05	2	2	0	0.500	0.500
group 3163	cadC	Cadmium resistance transcriptional regulatory protein CadC	64	4	53	35	54.70	89.74	10.566	5.78E-07	1.90E-03	2.37E-05	4	4	2	0.125	1.000
clpY		ATP-dependent protease ATPase subunit ClpY	89	12	28	27	76.07	69.23	7.152	6.24E-07	2.05E-03	2.37E-05	2	2	1	0.500	1.000
group 2401	bglF_1	PTS system beta-glucoside-specific EIIBCA component	89	12	28	27	76.07	69.23	7.152	6.24E-07	2.05E-03	2.37E-05	2	2	1	0.500	1.000
group 601		hypothetical protein	68	5	49	34	58.12	87.18	9.437	6.54E-07	2.15E-03	2.37E-05	3	2	3	0.250	1.000
group 906		hypothetical protein	87	11	30	28	74.36	71.79	7.382	6.60E-07	2.17E-03	2.37E-05	3	2	2	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2400		hypothetical protein	87	11	30	28	74.36	71.79	7.382	6.60E-07	2.17E-03	2.37E-05	2	2	2	0.500	0.500
group 3162	hin_1	DNA-invertase hin	63	4	54	35	53.85	89.74	10.208	7.35E-07	2.42E-03	2.37E-05	4	4	2	0.125	1.000
group 3161		Tn3 transposase DDE domain protein	63	4	54	35	53.85	89.74	10.208	7.35E-07	2.42E-03	2.37E-05	4	4	2	0.125	1.000
group 1471		hypothetical protein	54	2	63	37	46.15	94.87	15.857	7.39E-07	2.43E-03	2.37E-05	3	2	2	1.000	1.000
group 453	inlB_2	Leucine Rich repeats (2 copies)	71	6	46	33	60.68	84.62	8.489	7.52E-07	2.47E-03	2.37E-05	5	2	5	0.063	1.000
group 2807	nifS	Putative cysteine desulfurase NifS	31	28	86	11	26.50	28.21	0.142	8.52E-07	2.80E-03	2.37E-05	2	1	2	0.500	1.000
nifS		Putative cysteine desulfurase NifS	86	11	31	28	73.50	71.79	7.062	8.52E-07	2.80E-03	2.37E-05	2	2	1	0.500	1.000
group 1415		hypothetical protein	83	10	34	29	70.94	74.36	7.079	1.12E-06	3.69E-03	2.37E-05	2	2	2	0.500	0.500
pknA		Serine/threonine-protein kinase PknA	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000
group 391		hypothetical protein	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000
group 3110		hypothetical protein	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 394		hypothetical protein	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000
group 2449		hypothetical protein	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000
group 3177		hypothetical protein	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000
group 3108		hypothetical protein	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000
group 392		hypothetical protein	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000
aprN		Subtilisin NAT precursor	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000
group 2421		hypothetical protein	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000
group 3104		hypothetical protein	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000
group 917		hypothetical protein	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000
group 3178		hypothetical protein	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000
group 3179		hypothetical protein	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3123		CRISPR-associated protein (Cas_CXXC_CXXC)	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000
group 3113		hypothetical protein	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000
group 3106		GAD-like domain protein	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000
group 3107		hypothetical protein	41	0	76	39	35.04	100.00	inf	1.20E-06	3.93E-03	2.37E-05	1	1	0	1.000	1.000
group 3809		hypothetical protein	40	0	77	39	34.19	100.00	inf	1.43E-06	4.70E-03	2.37E-05	4	4	0	0.125	0.125
group 3798		Bacterial Ig-like domain (group 2)	40	0	77	39	34.19	100.00	inf	1.43E-06	4.70E-03	2.37E-05	3	3	0	0.250	0.250
group 597		CRISPR associated protein Cas6	40	0	77	39	34.19	100.00	inf	1.43E-06	4.70E-03	2.37E-05	1	1	0	1.000	1.000
inlA 8		Internalin-A precursor	101	18	16	21	86.32	53.85	7.365	1.47E-06	4.85E-03	2.37E-05	3	3	1	0.250	1.000
group 852		hypothetical protein	16	21	101	18	13.68	46.15	0.136	1.47E-06	4.85E-03	2.37E-05	3	1	3	0.250	1.000
group 2357		hypothetical protein	16	21	101	18	13.68	46.15	0.136	1.47E-06	4.85E-03	2.37E-05	3	1	3	0.250	1.000
group 3055		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
rpe 3		Ribulose-phosphate 3-epimerase	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
ulaB 1		Ascorbate-specific phosphotransferase enzyme IIB component	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
czcR		Transcriptional activator protein CzcR	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3149		Ureidoglycolate lyase	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3111		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2417		DNA-binding transcriptional repressor FabR	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
inlA 4		Internalin-A precursor	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 891		MacB-like periplasmic core domain protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3053		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3120		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3499		Acetyltransferase (GNAT) family protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
ogt 1		Methylated-DNA--protein-cysteine methyltransferase, constitutive	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3084		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
comEA		ComE operon protein 1	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 1387		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
licT 2		Transcription antiterminator LicT	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
sirC		Precorrin-2 dehydrogenase	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
cobU		Bifunctional adenosylcobalamin biosynthesis protein CobU	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
yvbK		putative N-acetyltransferase YvbK	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
kdgA		KHG/KDPG aldolase	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
dasR		HTH-type transcriptional repressor DasR	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
bglF 2		PTS system beta-glucoside-specific EIIBCA component	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2413		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2419		SMI1 / KNR4 family protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3090	ulaC_1	Ascorbate-specific phosphotransferase enzyme IIA component	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3115		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
cbiE		putative cobalt-precorrin-6Y C(5)-methyltransferase	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 1423		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
sdrD		Serine-aspartate repeat-containing protein D precursor	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
ftsW 3		Lipid II flippase FtsW	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3099		Acetamidase/Formamidase family protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 1416		lineage-specific thermal regulator protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3088		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 898		topology modulation protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
inlA 11		Internalin-A precursor	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2431		Acetyltransferase (GNAT) family protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3101		ADP-ribosylglycohydrolase	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
dapE		putative succinyl-diaminopimelate desuccinylase	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 598	inlJ_3	Internalin-J precursor	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 1413		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3097		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 896		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 909		Enterocin A Immunity	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3056		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
inlJ 6		Internalin-J precursor	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
rbsR 2		Ribose operon repressor	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
xylF		2-hydroxymuconate semialdehyde hydrolase	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
ulaA 1		Ascorbate-specific permease IIC component UlaA	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3100		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
thiO		Glycine oxidase	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 1391		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
inlJ 3		Internalin-J precursor	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3060		putative ABC transporter ATP-binding protein/MT1014	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3112		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3128		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3048		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3098	inlJ_2	Internalin-J precursor	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
sigV		RNA polymerase sigma factor SigV	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3052	rpiB_1	Ribose-5-phosphate isomerase B	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 399		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
bglH 2		Aryl-phospho-beta-D-glucosidase BglH	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3085		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 397	inlJ_9	Internalin-J precursor	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
crnA 1		Creatinine amidohydrolase	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 3064		hypothetical protein	88	12	29	27	75.21	69.23	6.828	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 649		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2773		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2612		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 777		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	3	2	2	1.000	1.000
group 2787		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2774		DNA-binding transcriptional repressor AcrR	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2792		DNA-binding transcriptional regulator FrlR	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
murQ		N-acetylmuramic acid 6-phosphate etherase	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
yknZ 1		putative ABC transporter permease YknZ	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
inlJ 16		Internalin-J precursor	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 1710		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2796		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 1709	cobU	Bifunctional adenosylcobalamin biosynthesis protein CobU	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
inlJ 4		Internalin-J precursor	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
inlJ 5		Internalin-J precursor	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2674		topology modulation protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2751		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 1671		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2116	inlA_8	Internalin-A precursor	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2750		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2778	xylF	2-hydroxymuconate semialdehyde hydrolase	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2607		PTS system EIIBC component	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 1059		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	3	2	2	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 367	thiO	Glycine oxidase	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2608		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 1362		Pesticidal crystal protein cry22Aa	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 994	clpY	ATP-dependent protease ATPase subunit ClpY	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2725	inlB_1	Internalin B precursor	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 1585	comEA	ComE operon protein 1	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2677	sirC	Precorrin-2 dehydrogenase	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2140		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2758		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 1713		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2707	ogt_1	Methylated-DNA--protein-cysteine methyltransferase, constitutive	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2757	inlJ_10	Internalin-J precursor	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2720	inlJ_3	Internalin-J precursor	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 1003		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2804		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
inlA 7		Internalin-A precursor	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2721		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
ybbH 3		putative HTH-type transcriptional regulator YbbH	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2793		macrolide transporter subunit MacA	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 1842		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2759		Acetamidase/Formamidase family protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2784	dapE	putative succinyl-diaminopimelate desuccinylase	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2771		Acetyltransferase (GNAT) family protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2611		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 1843		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2794	macB_3	Macrolide export ATP-binding/permease protein MacB	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2779		Bacterial regulatory proteins, tetR family	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2770		Ureidoglycolate lyase	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2737		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
inlJ 15		Internalin-J precursor	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2789		Acetyltransferase (GNAT) family protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
inlJ 12		Internalin-J precursor	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2675	cbiE	putative cobalt-precorrin-6Y C(5)-methyltransferase	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2780		bifunctional 3-demethylubiquinone-9 3-methyltransferase/ 2-octaprenyl-6-hydroxy phenol methylase	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2777	yvbK	putative N-acetyltransferase YvbK	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
group 2739		hypothetical protein	29	27	88	12	24.79	30.77	0.146	1.58E-06	5.20E-03	2.37E-05	1	1	1	1.000	1.000
pspA		Phosphoserine phosphatase 1	79	9	38	30	67.52	76.92	6.930	1.69E-06	5.54E-03	2.52E-05	4	4	2	0.125	1.000
group 3661		HNH endonuclease	0	9	117	30	0.00	76.92	0.000	1.78E-06	5.85E-03	2.63E-05	2	0	2	0.500	0.500
group 3660		hypothetical protein	0	9	117	30	0.00	76.92	0.000	1.78E-06	5.85E-03	2.63E-05	2	0	2	0.500	0.500
group 602		Integrase core domain protein	61	4	56	35	52.14	89.74	9.531	1.79E-06	5.88E-03	2.63E-05	5	4	3	0.375	1.000
group 1412		topology modulation protein	87	12	30	27	74.36	69.23	6.525	2.03E-06	6.69E-03	2.96E-05	2	1	2	0.500	1.000
cssS		Sensor histidine kinase CssS	87	12	30	27	74.36	69.23	6.525	2.03E-06	6.69E-03	2.96E-05	1	1	1	1.000	1.000
group 980		topology modulation protein	30	27	87	12	25.64	30.77	0.153	2.03E-06	6.69E-03	2.96E-05	2	2	1	0.500	1.000
group 203		hypothetical protein	5	14	112	25	4.27	64.10	0.080	2.22E-06	7.31E-03	3.22E-05	8	3	7	0.070	0.727

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 595		hypothetical protein	39	0	78	39	33.33	100.00	inf	2.62E-06	8.60E-03	3.74E-05	1	1	0	1.000	1.000
group 3122		hypothetical protein	39	0	78	39	33.33	100.00	inf	2.62E-06	8.60E-03	3.74E-05	1	1	0	1.000	1.000
group 3127		hypothetical protein	39	0	78	39	33.33	100.00	inf	2.62E-06	8.60E-03	3.74E-05	1	1	0	1.000	1.000
group 1235	fruA_5	PTS system fructose-specific EIIBC component	31	27	86	12	26.50	30.77	0.160	2.97E-06	9.77E-03	4.23E-05	3	3	1	0.250	1.000
inlB 2		Internalin B precursor	81	10	36	29	69.23	74.36	6.525	3.53E-06	1.16E-02	4.99E-05	3	3	2	0.250	1.000
group 2415	inlJ_1	Internalin-J precursor	81	10	36	29	69.23	74.36	6.525	3.53E-06	1.16E-02	4.99E-05	3	3	2	0.250	1.000
group 362	inlA_4	Internalin-A precursor	28	26	89	13	23.93	33.33	0.157	3.63E-06	1.19E-02	5.08E-05	3	2	2	1.000	1.000
group 331		hypothetical protein	28	26	89	13	23.93	33.33	0.157	3.63E-06	1.19E-02	5.08E-05	3	2	2	1.000	1.000
inlA 2		Internalin-A precursor	80	10	37	29	68.38	74.36	6.270	4.23E-06	1.39E-02	5.90E-05	3	3	2	0.250	1.000
group 2243		Leucine Rich repeats (2 copies)	29	26	88	13	24.79	33.33	0.165	4.64E-06	1.53E-02	6.28E-05	2	2	1	0.500	1.000
group 2211		Bacterial Ig-like domain (group 3)	29	26	88	13	24.79	33.33	0.165	4.64E-06	1.53E-02	6.28E-05	2	2	1	0.500	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 1027		hypothetical protein	29	26	88	13	24.79	33.33	0.165	4.64E-06	1.53E-02	6.28E-05	2	2	1	0.500	1.000
group 2785		hypothetical protein	29	26	88	13	24.79	33.33	0.165	4.64E-06	1.53E-02	6.28E-05	2	2	1	0.500	1.000
group 767	inlJ_13	Internalin-J precursor	29	26	88	13	24.79	33.33	0.165	4.64E-06	1.53E-02	6.28E-05	2	2	1	0.500	1.000
group 794		Acetyltransferase (GNAT) family protein	29	26	88	13	24.79	33.33	0.165	4.64E-06	1.53E-02	6.28E-05	2	2	1	0.500	1.000
group 1074	macB_1	Macrolide export ATP-binding/permease protein MacB	29	26	88	13	24.79	33.33	0.165	4.64E-06	1.53E-02	6.28E-05	2	2	1	0.500	1.000
group 1426		hypothetical protein	38	0	79	39	32.48	100.00	inf	5.61E-06	1.84E-02	7.41E-05	1	1	0	1.000	1.000
group 3805		Helix-turn-helix domain protein	38	0	79	39	32.48	100.00	inf	5.61E-06	1.84E-02	7.41E-05	3	3	0	0.250	0.250
tkt 4		Transketolase	38	0	79	39	32.48	100.00	inf	5.61E-06	1.84E-02	7.41E-05	1	1	0	1.000	1.000
group 1400		hypothetical protein	38	0	79	39	32.48	100.00	inf	5.61E-06	1.84E-02	7.41E-05	1	1	0	1.000	1.000
group 403		hypothetical protein	37	0	80	39	31.62	100.00	inf	5.61E-06	1.85E-02	7.41E-05	1	1	0	1.000	1.000
group 915		phosphoribulokinase	37	0	80	39	31.62	100.00	inf	5.61E-06	1.85E-02	7.41E-05	1	1	0	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 1153		hypothetical protein	38	29	79	10	32.48	25.64	0.166	5.77E-06	1.90E-02	7.59E-05	4	3	3	0.625	0.625
group 3586		hypothetical protein	6	14	111	25	5.13	64.10	0.097	6.16E-06	2.03E-02	8.07E-05	5	2	5	0.063	1.000
group 1386		hypothetical protein	85	12	32	27	72.65	69.23	5.977	6.39E-06	2.10E-02	8.34E-05	4	1	4	0.125	0.625
group 1417		Pesticidal crystal protein cry22Aa	49	2	68	37	41.88	94.87	13.331	7.60E-06	2.50E-02	8.38E-05	3	2	2	1.000	1.000
group 875		Phage protein Gp14	43	1	74	38	36.75	97.44	22.081	8.09E-06	2.66E-02	8.38E-05	5	4	1	0.375	0.375
clpC 2		putative ATP-dependent Clp protease ATP-binding subunit	58	4	59	35	49.57	89.74	8.602	8.22E-06	2.70E-02	8.38E-05	5	4	2	0.375	1.000
Int-Tn 1		Transposase from transposon Tn916	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4138		Terminase-like family protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4125		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4115		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4119		Phage late control gene D protein (GPD)	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 4161		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4169		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4160		Endodeoxyribonuclease RusA	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4175		Helix-turn-helix	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4144		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4173		Helix-turn-helix domain protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4172		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4128		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4153		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4116		Baseplate J-like protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4148		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 4162		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4156		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4140		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4151		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4117		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4112		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
immA 1		Metallopeptidase ImmA	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4106		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4111		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4137		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4141		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 4167		RecT family protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4105		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4165		Replication initiation and membrane attachment	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4110		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4139		Phage terminase small subunit	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
cwlC 1		Sporulation-specific N-acetylmuramoyl-L-alanine amidase	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4146		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4124		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4181		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4126		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4130		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 4168	smc_1	Chromosome partition protein Smc	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4143		putative methyltransferase	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4104		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4147		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4142		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4157		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
mepM		Murein DD-endopeptidase MepM	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4174		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4136		Phage Mu protein F like protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4158		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4108		Bacteriophage holin	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 4113		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4120		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4131		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4129		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4171		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4114		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4118		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4155		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 872		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	6	0	6	0.031	0.031
group 4134		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4133		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 4127		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4123		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4150		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4132		Encapsulating protein for peroxidase	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4145		Phage antirepressor protein KilAC domain protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4152		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
xre 2		HTH-type transcriptional regulator Xre	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4163		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4159		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4164	dnaC_2	DNA replication protein DnaC	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4109		membrane-bound lytic murein transglycosylase D	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 4149		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4135		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4178		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4121		LysM domain protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4170		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 4154		hypothetical protein	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
yycJ 1		Putative metallo-hydrolase YycJ	0	8	117	31	0.00	79.49	0.000	8.49E-06	2.79E-02	8.38E-05	1	0	1	1.000	1.000
group 311		hypothetical protein	13	18	104	21	11.11	53.85	0.146	9.23E-06	3.04E-02	9.09E-05	8	4	6	0.289	1.000
group 3157		hypothetical protein	57	4	60	35	48.72	89.74	8.313	9.54E-06	3.14E-02	9.28E-05	4	3	2	0.625	1.000
group 2445		hypothetical protein	57	4	60	35	48.72	89.74	8.313	9.54E-06	3.14E-02	9.28E-05	4	3	2	0.625	1.000
tnpR		Transposon Tn3 resolvase	57	4	60	35	48.72	89.74	8.313	9.54E-06	3.14E-02	9.28E-05	4	3	2	0.625	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3160		CobQ/CobB/MinD/ParA nucleotide binding domain protein	57	4	60	35	48.72	89.74	8.313	9.54E-06	3.14E-02	9.28E-05	4	3	2	0.625	1.000
cadA		putative cadmium-transporting ATPase	62	5	55	34	52.99	87.18	7.665	1.04E-05	3.41E-02	1.01E-04	6	5	3	0.219	1.000
group 69		hypothetical protein	35	0	82	39	29.91	100.00	inf	1.18E-05	3.90E-02	1.14E-04	3	3	0	0.250	0.250
group 4246		YqaJ-like viral recombinase domain protein	35	0	82	39	29.91	100.00	inf	1.18E-05	3.90E-02	1.14E-04	3	3	0	0.250	0.250
group 152		hypothetical protein	35	0	82	39	29.91	100.00	inf	1.18E-05	3.90E-02	1.14E-04	1	1	0	1.000	1.000
group 1329		hypothetical protein	75	9	42	30	64.10	76.92	5.952	1.26E-05	4.13E-02	1.20E-04	7	6	4	0.125	1.000
sbcC		Nuclease SbcCD subunit C	37	28	80	11	31.62	28.21	0.182	1.36E-05	4.48E-02	1.30E-04	2	2	2	0.500	0.500
group 674		hypothetical protein	31	26	86	13	26.50	33.33	0.180	1.42E-05	4.68E-02	1.36E-04	2	2	1	0.500	1.000
group 795		Acetyltransferase (GNAT) family protein	88	14	29	25	75.21	64.10	5.419	1.45E-05	4.76E-02	1.36E-04	2	1	2	0.500	1.000
group 360	inlA_10	Internalin-A precursor	29	25	88	14	24.79	35.90	0.185	1.45E-05	4.76E-02	1.36E-04	2	2	1	0.500	1.000
group 2723		hypothetical protein	29	25	88	14	24.79	35.90	0.185	1.45E-05	4.76E-02	1.36E-04	2	2	1	0.500	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2724		hypothetical protein	29	25	88	14	24.79	35.90	0.185	1.45E-05	4.76E-02	1.36E-04	2	2	1	0.500	1.000
xerC 3		Tyrosine recombinase XerC	1	9	116	30	0.85	76.92	0.029	1.45E-05	4.77E-02	1.36E-04	4	1	3	0.625	0.625
group 1446		hypothetical protein	1	9	116	30	0.85	76.92	0.029	1.45E-05	4.77E-02	1.36E-04	3	1	2	1.000	1.000
group 1407		hypothetical protein	77	10	40	29	65.81	74.36	5.583	1.46E-05	4.81E-02	1.37E-04	5	2	5	0.063	1.000
group 613		ASCH domain protein	2	10	115	29	1.71	74.36	0.050	1.61E-05	5.28E-02	1.50E-04	5	2	3	1.000	1.000
group 3143		Putative phosphatase MPN_427	41	1	76	38	35.04	97.44	20.500	1.71E-05	5.64E-02	1.55E-04	2	1	1	1.000	1.000
group 2424		hypothetical protein	41	1	76	38	35.04	97.44	20.500	1.71E-05	5.64E-02	1.55E-04	1	1	1	1.000	1.000
group 3148	ybbH_3	putative HTH-type transcriptional regulator YbbH	41	1	76	38	35.04	97.44	20.500	1.71E-05	5.64E-02	1.55E-04	2	1	1	1.000	1.000
group 3144		hypothetical protein	41	1	76	38	35.04	97.44	20.500	1.71E-05	5.64E-02	1.55E-04	2	1	1	1.000	1.000
group 3139		hypothetical protein	41	1	76	38	35.04	97.44	20.500	1.71E-05	5.64E-02	1.55E-04	2	1	1	1.000	1.000
licB 8		Lichenan-specific phosphotransferase enzyme IIB component	41	1	76	38	35.04	97.44	20.500	1.71E-05	5.64E-02	1.55E-04	2	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
licA 5		Lichenan-specific phosphotransferase enzyme IIA component	41	1	76	38	35.04	97.44	20.500	1.71E-05	5.64E-02	1.55E-04	2	1	1	1.000	1.000
nanE 2		Putative N-acetylmannosamine-6-phosphate 2-epimerase	41	1	76	38	35.04	97.44	20.500	1.71E-05	5.64E-02	1.55E-04	2	1	1	1.000	1.000
group 3138	bin3	Putative transposon Tn552 DNA-invertase bin3	41	1	76	38	35.04	97.44	20.500	1.71E-05	5.64E-02	1.55E-04	2	1	1	1.000	1.000
licC 7		Lichenan permease IIC component	41	1	76	38	35.04	97.44	20.500	1.71E-05	5.64E-02	1.55E-04	2	1	1	1.000	1.000
group 3140		Cupin domain protein	41	1	76	38	35.04	97.44	20.500	1.71E-05	5.64E-02	1.55E-04	2	1	1	1.000	1.000
group 2430		hypothetical protein	83	12	34	27	70.94	69.23	5.493	1.89E-05	6.21E-02	1.70E-04	4	3	3	0.625	0.625
group 455		hypothetical protein	27	24	90	15	23.08	38.46	0.188	2.14E-05	7.05E-02	1.93E-04	5	3	3	1.000	1.000
dhaK 1		PTS-dependent dihydroxyacetone kinase, dihydroxyacetone-binding subunit DhaK	82	12	35	27	70.09	69.23	5.271	2.27E-05	7.46E-02	2.03E-04	2	1	2	0.500	1.000
group 2418		hypothetical protein	34	0	83	39	29.06	100.00	inf	2.32E-05	7.63E-02	2.07E-04	1	1	0	1.000	1.000
group 1215		hypothetical protein	38	28	79	11	32.48	28.21	0.189	2.65E-05	8.71E-02	2.36E-04	6	5	3	0.219	1.000
group 2941		hypothetical protein	16	19	101	20	13.68	51.28	0.167	2.74E-05	9.00E-02	2.41E-04	3	1	3	0.250	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2961		hypothetical protein	16	19	101	20	13.68	51.28	0.167	2.74E-05	9.00E-02	2.41E-04	2	1	2	0.500	1.000
group 2938		hypothetical protein	16	19	101	20	13.68	51.28	0.167	2.74E-05	9.00E-02	2.41E-04	3	1	3	0.250	1.000
group 2940		hypothetical protein	16	19	101	20	13.68	51.28	0.167	2.74E-05	9.00E-02	2.41E-04	3	1	3	0.250	1.000
group 2939		Ribonuclease	16	19	101	20	13.68	51.28	0.167	2.74E-05	9.00E-02	2.41E-04	3	1	3	0.250	1.000
group 1439		hypothetical protein	50	3	67	36	42.74	92.31	8.955	2.87E-05	9.44E-02	2.52E-04	7	6	3	0.125	1.000
group 2327		Fic/DOC family protein	76	10	41	29	64.96	74.36	5.376	2.95E-05	9.72E-02	2.52E-04	5	5	3	0.063	1.000
group 840		hypothetical protein	76	10	41	29	64.96	74.36	5.376	2.95E-05	9.72E-02	2.52E-04	5	5	3	0.063	1.000
group 2325		hypothetical protein	76	10	41	29	64.96	74.36	5.376	2.95E-05	9.72E-02	2.52E-04	5	5	3	0.063	1.000
group 2909		hypothetical protein	76	10	41	29	64.96	74.36	5.376	2.95E-05	9.72E-02	2.52E-04	5	5	3	0.063	1.000
group 2326		hypothetical protein	76	10	41	29	64.96	74.36	5.376	2.95E-05	9.72E-02	2.52E-04	5	5	3	0.063	1.000
group 2324		hypothetical protein	76	10	41	29	64.96	74.36	5.376	2.95E-05	9.72E-02	2.52E-04	5	5	3	0.063	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2911		Helix-turn-helix	76	10	41	29	64.96	74.36	5.376	2.95E-05	9.72E-02	2.52E-04	5	5	3	0.063	1.000
group 2912		hypothetical protein	76	10	41	29	64.96	74.36	5.376	2.95E-05	9.72E-02	2.52E-04	5	5	3	0.063	1.000
group 2910		hypothetical protein	76	10	41	29	64.96	74.36	5.376	2.95E-05	9.72E-02	2.52E-04	5	5	3	0.063	1.000
group 1332		hypothetical protein	76	10	41	29	64.96	74.36	5.376	2.95E-05	9.72E-02	2.52E-04	5	5	3	0.063	1.000
group 1406		hypothetical protein	70	8	47	31	59.83	79.49	5.771	3.22E-05	1.06E-01	2.75E-04	4	3	3	0.625	0.625
group 876		Phage protein Gp14	5	12	112	27	4.27	69.23	0.100	3.29E-05	1.08E-01	2.80E-04	4	1	4	0.125	0.625
group 2438		hypothetical protein	40	1	77	38	34.19	97.44	19.740	3.45E-05	1.13E-01	2.92E-04	2	1	1	1.000	1.000
inlB 7		Internalin B precursor	40	1	77	38	34.19	97.44	19.740	3.45E-05	1.13E-01	2.92E-04	2	1	1	1.000	1.000
group 1427		putative transporter	39	1	78	38	33.33	97.44	19.000	3.55E-05	1.17E-01	3.00E-04	2	1	1	1.000	1.000
group 3159		hypothetical protein	54	4	63	35	46.15	89.74	7.500	3.97E-05	1.31E-01	3.34E-04	4	3	2	0.625	1.000
group 3026		Bacteriophage Gp15 protein	10	15	107	24	8.55	61.54	0.150	4.56E-05	1.50E-01	3.83E-04	6	3	6	0.031	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3800		hypothetical protein	32	0	85	39	27.35	100.00	inf	4.62E-05	1.52E-01	3.83E-04	2	2	0	0.500	0.500
group 3801		hypothetical protein	32	0	85	39	27.35	100.00	inf	4.62E-05	1.52E-01	3.83E-04	2	2	0	0.500	0.500
group 3802	ssb_2	Single-stranded DNA-binding protein ssb	32	0	85	39	27.35	100.00	inf	4.62E-05	1.52E-01	3.83E-04	2	2	0	0.500	0.500
group 3799		hypothetical protein	32	0	85	39	27.35	100.00	inf	4.62E-05	1.52E-01	3.83E-04	2	2	0	0.500	0.500
group 99		hypothetical protein	32	0	85	39	27.35	100.00	inf	4.62E-05	1.52E-01	3.83E-04	2	2	0	0.500	0.500
group 1170	dhaK_1	PTS-dependent dihydroxyacetone kinase, dihydroxyacetone-binding subunit DhaK	36	27	81	12	30.77	30.77	0.198	5.20E-05	1.71E-01	4.30E-04	2	2	1	0.500	1.000
tsf 2		Elongation factor Ts	57	5	60	34	48.72	87.18	6.460	5.37E-05	1.77E-01	4.39E-04	5	3	3	1.000	1.000
group 3041		hypothetical protein	57	5	60	34	48.72	87.18	6.460	5.37E-05	1.77E-01	4.39E-04	5	3	3	1.000	1.000
chpS		Antitoxin ChpS	57	5	60	34	48.72	87.18	6.460	5.37E-05	1.77E-01	4.39E-04	5	3	3	1.000	1.000
mazF		mRNA interferase MazF	57	5	60	34	48.72	87.18	6.460	5.37E-05	1.77E-01	4.39E-04	5	3	3	1.000	1.000
group 2396		hypothetical protein	57	5	60	34	48.72	87.18	6.460	5.37E-05	1.77E-01	4.39E-04	5	3	3	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3130		hypothetical protein	49	3	68	36	41.88	92.31	8.647	5.68E-05	1.87E-01	4.60E-04	2	2	2	0.500	0.500
group 3131		hypothetical protein	49	3	68	36	41.88	92.31	8.647	5.68E-05	1.87E-01	4.60E-04	2	2	2	0.500	0.500
group 3132		hypothetical protein	49	3	68	36	41.88	92.31	8.647	5.68E-05	1.87E-01	4.60E-04	2	2	2	0.500	0.500
group 1134		hypothetical protein	27	23	90	16	23.08	41.03	0.209	6.04E-05	1.99E-01	4.88E-04	5	2	4	0.375	1.000
group 883		Phage holin	1	8	116	31	0.85	79.49	0.033	6.22E-05	2.04E-01	4.97E-04	1	1	1	1.000	1.000
group 1144	csbC	hypothetical protein	1	8	116	31	0.85	79.49	0.033	6.22E-05	2.04E-01	4.97E-04	2	1	1	1.000	1.000
group 1381		Protein gp23 (Bacteriophage A118)	1	8	116	31	0.85	79.49	0.033	6.22E-05	2.04E-01	4.97E-04	1	1	1	1.000	1.000
group 369	yhdN_3	hypothetical protein	1	8	116	31	0.85	79.49	0.033	6.22E-05	2.04E-01	4.97E-04	2	1	1	1.000	1.000
group 2439		hypothetical protein	88	15	29	24	75.21	61.54	4.855	6.32E-05	2.08E-01	5.01E-04	2	1	2	0.500	1.000
skfE		SkfA peptide export ATP-binding protein SkfE	88	15	29	24	75.21	61.54	4.855	6.32E-05	2.08E-01	5.01E-04	2	1	2	0.500	1.000
ytrA 4		HTH-type transcriptional repressor YtrA	88	15	29	24	75.21	61.54	4.855	6.32E-05	2.08E-01	5.01E-04	2	1	2	0.500	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2740		hypothetical protein	29	24	88	15	24.79	38.46	0.206	6.32E-05	2.08E-01	5.01E-04	2	2	1	0.500	1.000
group 3251		hypothetical protein	2	9	115	30	1.71	76.92	0.058	6.51E-05	2.14E-01	5.12E-04	3	1	2	1.000	1.000
group 1458		hypothetical protein	2	9	115	30	1.71	76.92	0.058	6.51E-05	2.14E-01	5.12E-04	3	1	2	1.000	1.000
group 930		hypothetical protein	2	9	115	30	1.71	76.92	0.058	6.51E-05	2.14E-01	5.12E-04	4	2	2	1.000	1.000
gpmA 7		2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	38	1	79	38	32.48	97.44	18.278	6.88E-05	2.26E-01	5.40E-04	2	1	1	1.000	1.000
group 388		Minor capsid protein from bacteriophage	37	1	80	38	31.62	97.44	17.575	7.22E-05	2.37E-01	5.55E-04	5	4	1	0.375	0.375
group 1121		hypothetical protein	101	21	16	18	86.32	46.15	5.411	7.22E-05	2.37E-01	5.55E-04	3	3	1	0.250	1.000
group 1916		hypothetical protein	101	21	16	18	86.32	46.15	5.411	7.22E-05	2.37E-01	5.55E-04	3	3	1	0.250	1.000
group 2937		hypothetical protein	16	18	101	21	13.68	53.85	0.185	7.22E-05	2.37E-01	5.55E-04	2	1	2	0.500	1.000
group 2352		hypothetical protein	16	18	101	21	13.68	53.85	0.185	7.22E-05	2.37E-01	5.55E-04	2	1	2	0.500	1.000
group 1374		Endodeoxyribonuclease RusA	16	18	101	21	13.68	53.85	0.185	7.22E-05	2.37E-01	5.55E-04	8	4	6	0.289	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 1347		hypothetical protein	16	18	101	21	13.68	53.85	0.185	7.22E-05	2.37E-01	5.55E-04	4	1	4	0.125	0.625
group 36	ssb_2	Single-stranded DNA-binding protein ssb	16	18	101	21	13.68	53.85	0.185	7.22E-05	2.37E-01	5.55E-04	8	4	6	0.289	1.000
group 1122		hypothetical protein	16	18	101	21	13.68	53.85	0.185	7.22E-05	2.37E-01	5.55E-04	2	1	2	0.500	1.000
group 1331		hypothetical protein	76	11	41	28	64.96	71.79	4.718	7.92E-05	2.61E-01	6.06E-04	5	5	4	0.063	0.375
group 2756		hypothetical protein	75	37	42	2	64.10	5.13	0.097	7.96E-05	2.62E-01	6.06E-04	4	1	4	0.125	0.625
group 3236		hypothetical protein	42	2	75	37	35.90	94.87	10.360	7.96E-05	2.62E-01	6.06E-04	4	4	1	0.125	0.625
group 943		Phage minor structural protein GP20	42	2	75	37	35.90	94.87	10.360	7.96E-05	2.62E-01	6.06E-04	5	5	0	0.063	0.063
group 3585		hypothetical protein	6	12	111	27	5.13	69.23	0.122	8.13E-05	2.67E-01	6.17E-04	5	2	4	0.375	1.000
group 786		hypothetical protein	47	30	70	9	40.17	23.08	0.201	8.19E-05	2.69E-01	6.21E-04	4	3	3	0.625	0.625
group 1431		Integrase core domain protein	52	4	65	35	44.44	89.74	7.000	8.29E-05	2.73E-01	6.27E-04	5	2	3	1.000	1.000
group 199		hypothetical protein	11	15	106	24	9.40	61.54	0.166	8.95E-05	2.94E-01	6.74E-04	3	1	3	0.250	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 387		Minor capsid protein from bacteriophage	11	15	106	24	9.40	61.54	0.166	8.95E-05	2.94E-01	6.74E-04	7	4	6	0.125	1.000
group 1403		hypothetical protein	51	4	66	35	43.59	89.74	6.761	9.25E-05	3.04E-01	6.93E-04	4	3	3	0.625	0.625
group 1395		hypothetical protein	51	4	66	35	43.59	89.74	6.761	9.25E-05	3.04E-01	6.93E-04	7	7	1	0.016	0.125
group 3015		hypothetical protein	5	11	112	28	4.27	71.79	0.114	1.18E-04	3.87E-01	8.79E-04	4	1	4	0.125	0.625
group 1930		hypothetical protein	28	23	89	16	23.93	41.03	0.219	1.27E-04	4.16E-01	9.44E-04	4	2	3	0.625	1.000
group 1405		hypothetical protein	54	5	63	34	46.15	87.18	5.829	1.30E-04	4.28E-01	9.66E-04	3	2	2	1.000	1.000
group 919		hypothetical protein	54	5	63	34	46.15	87.18	5.829	1.30E-04	4.28E-01	9.66E-04	8	5	5	0.727	0.727
group 3483		hypothetical protein	36	1	81	38	30.77	97.44	16.889	1.35E-04	4.43E-01	9.99E-04	3	3	1	0.250	1.000
group 1418		hypothetical protein	41	2	76	37	35.04	94.87	9.980	1.40E-04	4.59E-01	1.03E-03	2	1	2	0.500	1.000
group 3121		hypothetical protein	41	2	76	37	35.04	94.87	9.980	1.40E-04	4.59E-01	1.03E-03	1	1	1	1.000	1.000
hin 2		DNA-invertase hin	95	19	22	20	81.20	51.28	4.545	1.53E-04	5.04E-01	1.12E-03	10	6	5	0.754	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
ebrB		Multidrug resistance protein EbrB	95	19	22	20	81.20	51.28	4.545	1.53E-04	5.04E-01	1.12E-03	10	6	5	0.754	1.000
slmA		Nucleoid occlusion factor SlmA	95	19	22	20	81.20	51.28	4.545	1.53E-04	5.04E-01	1.12E-03	10	6	5	0.754	1.000
qacC		Quaternary ammonium compound-resistance protein QacC	95	19	22	20	81.20	51.28	4.545	1.53E-04	5.04E-01	1.12E-03	10	6	5	0.754	1.000
lolD		Lipoprotein-releasing system ATP-binding protein LolD	42	28	75	11	35.90	28.21	0.220	1.56E-04	5.11E-01	1.13E-03	3	2	2	1.000	1.000
group 1975		hypothetical protein	42	28	75	11	35.90	28.21	0.220	1.56E-04	5.11E-01	1.13E-03	3	2	2	1.000	1.000
group 2230		hypothetical protein	29	23	88	16	24.79	41.03	0.229	1.58E-04	5.21E-01	1.15E-03	5	5	1	0.063	0.375
group 160		Methyltransferase domain protein	4	10	113	29	3.42	74.36	0.103	1.63E-04	5.36E-01	1.18E-03	6	3	3	1.000	1.000
group 3567		Telomeric repeat-binding factor 2	4	10	113	29	3.42	74.36	0.103	1.63E-04	5.36E-01	1.18E-03	4	2	3	0.625	1.000
group 465		hypothetical protein	31	24	86	15	26.50	38.46	0.225	1.73E-04	5.68E-01	1.25E-03	9	6	7	0.180	0.508
bacD		Alanine-anticapsin ligase BacD	28	0	89	39	23.93	100.00	inf	1.78E-04	5.84E-01	1.28E-03	1	1	0	1.000	1.000
group 29	cwlK_2	Peptidoglycan L-alanyl-D-glutamate endopeptidase CwlK precursor	0	6	117	33	0.00	84.62	0.000	1.80E-04	5.91E-01	1.29E-03	1	0	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2959		hypothetical protein	16	17	101	22	13.68	56.41	0.205	1.97E-04	6.49E-01	1.39E-03	2	1	2	0.500	1.000
group 2337		hypothetical protein	16	17	101	22	13.68	56.41	0.205	1.97E-04	6.49E-01	1.39E-03	3	1	3	0.250	1.000
group 2974		hypothetical protein	16	17	101	22	13.68	56.41	0.205	1.97E-04	6.49E-01	1.39E-03	3	1	3	0.250	1.000
group 2973		hypothetical protein	16	17	101	22	13.68	56.41	0.205	1.97E-04	6.49E-01	1.39E-03	3	1	3	0.250	1.000
group 2964		putative peptidase	16	17	101	22	13.68	56.41	0.205	1.97E-04	6.49E-01	1.39E-03	2	1	2	0.500	1.000
group 4064		Bacterial Ig-like domain (group 3)	16	17	101	22	13.68	56.41	0.205	1.97E-04	6.49E-01	1.39E-03	2	1	2	0.500	1.000
group 2949		hypothetical protein	16	17	101	22	13.68	56.41	0.205	1.97E-04	6.49E-01	1.39E-03	2	1	2	0.500	1.000
group 828		hypothetical protein	101	22	16	17	86.32	43.59	4.878	1.97E-04	6.49E-01	1.39E-03	3	3	1	0.250	1.000
group 1113		hypothetical protein	101	22	16	17	86.32	43.59	4.878	1.97E-04	6.49E-01	1.39E-03	2	2	1	0.500	1.000
group 46		hypothetical protein	32	24	85	15	27.35	38.46	0.235	2.03E-04	6.66E-01	1.42E-03	8	4	5	0.727	1.000
group 574		hypothetical protein	3	9	114	30	2.56	76.92	0.088	2.12E-04	6.98E-01	1.49E-03	3	1	2	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 201		hypothetical protein	53	5	64	34	45.30	87.18	5.631	2.26E-04	7.45E-01	1.58E-03	7	5	4	0.453	1.000
fer		Ferredoxin	103	23	14	16	88.03	41.03	5.118	2.37E-04	7.80E-01	1.66E-03	4	3	3	0.625	0.625
group 889		hypothetical protein	44	3	73	36	37.61	92.31	7.233	2.39E-04	7.87E-01	1.67E-03	6	5	1	0.219	0.219
dpnM		Modification methylase DpnIIA	1	7	116	32	0.85	82.05	0.039	2.57E-04	8.44E-01	1.79E-03	4	1	4	0.125	0.625
group 2423		hypothetical protein	40	2	77	37	34.19	94.87	9.610	2.72E-04	8.95E-01	1.88E-03	2	1	2	0.500	1.000
group 2422		hypothetical protein	40	2	77	37	34.19	94.87	9.610	2.72E-04	8.95E-01	1.88E-03	2	1	2	0.500	1.000
yjcD		Putative ATP-dependent DNA helicase YjcD	11	14	106	25	9.40	64.10	0.185	2.76E-04	9.07E-01	1.91E-03	3	2	3	0.250	1.000
group 3482		ASCH domain protein	33	1	84	38	28.21	97.44	14.929	2.82E-04	9.28E-01	1.95E-03	2	2	1	0.500	1.000
group 2402		hypothetical protein	56	6	61	33	47.86	84.62	5.049	2.87E-04	9.43E-01	1.97E-03	8	7	3	0.070	0.727
group 261		Integrase core domain protein	60	7	57	32	51.28	82.05	4.812	3.14E-04	1.00E+00	2.16E-03	6	6	2	0.031	0.688
group 611		hypothetical protein	48	4	69	35	41.03	89.74	6.087	3.29E-04	1.00E+00	2.26E-03	5	4	4	0.375	0.375

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 1420		Listeria-Bacteroides repeat domain (List_Bact_rpt)	26	0	91	39	22.22	100.00	inf	3.44E-04	1.00E+00	2.35E-03	1	1	0	1.000	1.000
group 4247		hypothetical protein	27	0	90	39	23.08	100.00	inf	3.45E-04	1.00E+00	2.35E-03	1	1	0	1.000	1.000
group 147		hypothetical protein	15	16	102	23	12.82	58.97	0.211	3.46E-04	1.00E+00	2.36E-03	4	3	3	0.625	0.625
group 386		Minor capsid protein	8	12	109	27	6.84	69.23	0.165	3.72E-04	1.00E+00	2.53E-03	5	2	5	0.063	1.000
group 3590		hypothetical protein	5	10	112	29	4.27	74.36	0.129	3.99E-04	1.00E+00	2.65E-03	3	1	3	0.250	1.000
group 3589		hypothetical protein	5	10	112	29	4.27	74.36	0.129	3.99E-04	1.00E+00	2.65E-03	3	1	3	0.250	1.000
group 3566		hypothetical protein	5	10	112	29	4.27	74.36	0.129	3.99E-04	1.00E+00	2.65E-03	4	2	2	1.000	1.000
group 1498	xerC_2	Tyrosine recombinase XerC	5	10	112	29	4.27	74.36	0.129	3.99E-04	1.00E+00	2.65E-03	2	1	2	0.500	1.000
group 276		hypothetical protein	5	10	112	29	4.27	74.36	0.129	3.99E-04	1.00E+00	2.65E-03	3	1	3	0.250	1.000
group 2572		hypothetical protein	5	10	112	29	4.27	74.36	0.129	3.99E-04	1.00E+00	2.65E-03	2	1	2	0.500	1.000
group 3564		hypothetical protein	5	10	112	29	4.27	74.36	0.129	3.99E-04	1.00E+00	2.65E-03	4	2	2	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3591		hypothetical protein	5	10	112	29	4.27	74.36	0.129	3.99E-04	1.00E+00	2.65E-03	3	1	3	0.250	1.000
group 4068		Phage terminase large subunit	5	10	112	29	4.27	74.36	0.129	3.99E-04	1.00E+00	2.65E-03	3	1	3	0.250	1.000
group 3565	immR_2	HTH-type transcriptional regulator ImmR	5	10	112	29	4.27	74.36	0.129	3.99E-04	1.00E+00	2.65E-03	4	2	2	1.000	1.000
uup		ABC transporter ATP-binding protein uup	13	15	104	24	11.11	61.54	0.200	4.22E-04	1.00E+00	2.80E-03	4	3	4	0.125	0.625
group 15		Listeria-Bacteroides repeat domain (List_Bact_rpt)	13	15	104	24	11.11	61.54	0.200	4.22E-04	1.00E+00	2.80E-03	5	3	5	0.063	1.000
group 2458		hypothetical protein	43	3	74	36	36.75	92.31	6.973	4.34E-04	1.00E+00	2.87E-03	8	7	2	0.070	0.289
group 590		hypothetical protein	51	5	66	34	43.59	87.18	5.255	4.43E-04	1.00E+00	2.93E-03	6	6	2	0.031	0.688
group 3189		HTH domain protein	16	16	101	23	13.68	58.97	0.228	5.34E-04	1.00E+00	3.47E-03	3	2	2	1.000	1.000
group 2965		TM2 domain protein	16	16	101	23	13.68	58.97	0.228	5.34E-04	1.00E+00	3.47E-03	2	1	2	0.500	1.000
group 2958		hypothetical protein	16	16	101	23	13.68	58.97	0.228	5.34E-04	1.00E+00	3.47E-03	2	1	2	0.500	1.000
group 2336		hypothetical protein	16	16	101	23	13.68	58.97	0.228	5.34E-04	1.00E+00	3.47E-03	2	1	2	0.500	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2960		hypothetical protein	16	16	101	23	13.68	58.97	0.228	5.34E-04	1.00E+00	3.47E-03	2	1	2	0.500	1.000
group 2926		hypothetical protein	16	16	101	23	13.68	58.97	0.228	5.34E-04	1.00E+00	3.47E-03	2	1	2	0.500	1.000
group 2966	lntA	Listeria nuclear targeted protein A precursor	16	16	101	23	13.68	58.97	0.228	5.34E-04	1.00E+00	3.47E-03	2	1	2	0.500	1.000
lntA		Listeria nuclear targeted protein A precursor	101	23	16	16	86.32	41.03	4.391	5.34E-04	1.00E+00	3.47E-03	2	2	1	0.500	1.000
group 2383		hypothetical protein	4	9	113	30	3.42	76.92	0.118	5.62E-04	1.00E+00	3.63E-03	3	2	2	1.000	1.000
group 297		hypothetical protein	4	9	113	30	3.42	76.92	0.118	5.62E-04	1.00E+00	3.63E-03	4	2	2	1.000	1.000
group 3792		Helix-turn-helix domain protein	4	9	113	30	3.42	76.92	0.118	5.62E-04	1.00E+00	3.63E-03	3	2	2	1.000	1.000
group 237		hypothetical protein	27	21	90	18	23.08	46.15	0.257	5.63E-04	1.00E+00	3.63E-03	7	4	6	0.125	1.000
cgkA		Kappa-carrageenase precursor	7	11	110	28	5.98	71.79	0.162	5.74E-04	1.00E+00	3.69E-03	3	2	3	0.250	1.000
group 1372		hypothetical protein	14	15	103	24	11.97	61.54	0.217	5.98E-04	1.00E+00	3.84E-03	2	1	2	0.500	1.000
group 3109		hypothetical protein	46	4	71	35	39.32	89.74	5.669	6.28E-04	1.00E+00	4.02E-03	2	1	2	0.500	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 1375		Terminase small subunit	46	4	71	35	39.32	89.74	5.669	6.28E-04	1.00E+00	4.02E-03	8	5	4	0.727	1.000
group 2745		hypothetical protein	18	17	99	22	15.38	56.41	0.235	6.39E-04	1.00E+00	4.04E-03	3	2	3	0.250	1.000
group 2744		hypothetical protein	18	17	99	22	15.38	56.41	0.235	6.39E-04	1.00E+00	4.04E-03	3	2	3	0.250	1.000
group 2746	eccCa1	ESX-1 secretion system protein EccCa1	18	17	99	22	15.38	56.41	0.235	6.39E-04	1.00E+00	4.04E-03	3	2	3	0.250	1.000
group 1137		hypothetical protein	18	17	99	22	15.38	56.41	0.235	6.39E-04	1.00E+00	4.04E-03	3	2	3	0.250	1.000
group 1105		putative sulfate transporter/MT1781	99	22	18	17	84.62	43.59	4.250	6.39E-04	1.00E+00	4.04E-03	3	3	2	0.250	1.000
group 40		hypothetical protein	29	22	88	17	24.79	43.59	0.255	6.41E-04	1.00E+00	4.04E-03	5	4	2	0.375	1.000
group 2749		hypothetical protein	29	22	88	17	24.79	43.59	0.255	6.41E-04	1.00E+00	4.04E-03	2	2	1	0.500	1.000
group 3450		hypothetical protein	24	0	93	39	20.51	100.00	inf	6.61E-04	1.00E+00	4.16E-03	2	2	0	0.500	0.500
group 115		hypothetical protein	61	8	56	31	52.14	79.49	4.221	6.95E-04	1.00E+00	4.37E-03	9	7	5	0.180	1.000
group 918		hypothetical protein	68	10	49	29	58.12	74.36	4.024	7.43E-04	1.00E+00	4.66E-03	4	2	3	0.625	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 4180		hypothetical protein	3	8	114	31	2.56	79.49	0.102	7.53E-04	1.00E+00	4.72E-03	1	1	1	1.000	1.000
group 3668		hypothetical protein	0	5	117	34	0.00	87.18	0.000	7.98E-04	1.00E+00	4.97E-03	1	0	1	1.000	1.000
group 3667		hypothetical protein	0	5	117	34	0.00	87.18	0.000	7.98E-04	1.00E+00	4.97E-03	1	0	1	1.000	1.000
group 888		hypothetical protein	0	5	117	34	0.00	87.18	0.000	7.98E-04	1.00E+00	4.97E-03	4	0	4	0.125	0.125
group 1440		hypothetical protein	41	3	76	36	35.04	92.31	6.474	8.13E-04	1.00E+00	5.02E-03	6	4	2	0.688	0.688
group 374		hypothetical protein	41	3	76	36	35.04	92.31	6.474	8.13E-04	1.00E+00	5.02E-03	2	1	2	0.500	1.000
group 186		hypothetical protein	41	3	76	36	35.04	92.31	6.474	8.13E-04	1.00E+00	5.02E-03	2	1	2	0.500	1.000
group 58		Bacillus transposase protein	41	3	76	36	35.04	92.31	6.474	8.13E-04	1.00E+00	5.02E-03	2	1	2	0.500	1.000
group 183		hypothetical protein	41	3	76	36	35.04	92.31	6.474	8.13E-04	1.00E+00	5.02E-03	2	1	2	0.500	1.000
group 3250		hypothetical protein	6	10	111	29	5.13	74.36	0.157	8.67E-04	1.00E+00	5.31E-03	4	2	2	1.000	1.000
group 2596		Terminase small subunit	6	10	111	29	5.13	74.36	0.157	8.67E-04	1.00E+00	5.31E-03	4	2	3	0.625	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3187		hypothetical protein	40	3	77	36	34.19	92.31	6.234	8.68E-04	1.00E+00	5.31E-03	6	4	3	0.688	1.000
group 2454		hypothetical protein	40	3	77	36	34.19	92.31	6.234	8.68E-04	1.00E+00	5.31E-03	6	4	3	0.688	1.000
group 435		hypothetical protein	19	17	98	22	16.24	56.41	0.251	8.70E-04	1.00E+00	5.32E-03	5	3	3	1.000	1.000
group 2588		hypothetical protein	30	1	87	38	25.64	97.44	13.103	9.20E-04	1.00E+00	5.61E-03	3	2	1	1.000	1.000
group 568		hypothetical protein	31	1	86	38	26.50	97.44	13.698	9.47E-04	1.00E+00	5.75E-03	5	5	1	0.063	0.375
group 1477		hypothetical protein	31	1	86	38	26.50	97.44	13.698	9.47E-04	1.00E+00	5.75E-03	3	3	1	0.250	1.000
group 871		hypothetical protein	1	6	116	33	0.85	84.62	0.047	1.02E-03	1.00E+00	6.09E-03	4	1	4	0.125	0.625
group 30	cwlK_2	Peptidoglycan L-alanyl-D-glutamate endopeptidase CwlK precursor	1	6	116	33	0.85	84.62	0.047	1.02E-03	1.00E+00	6.09E-03	4	1	3	0.625	0.625
group 540		hypothetical protein	26	20	91	19	22.22	48.72	0.271	1.02E-03	1.00E+00	6.09E-03	5	4	3	0.375	1.000
group 2155		hypothetical protein	13	14	104	25	11.11	64.10	0.223	1.02E-03	1.00E+00	6.09E-03	3	1	3	0.250	1.000
group 9		hypothetical protein	13	14	104	25	11.11	64.10	0.223	1.02E-03	1.00E+00	6.09E-03	4	3	4	0.125	0.625

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2967		chromosome segregation protein	104	25	13	14	88.89	35.90	4.480	1.02E-03	1.00E+00	6.09E-03	3	3	1	0.250	1.000
acpB		Capsule synthesis positive regulator AcpB	104	25	13	14	88.89	35.90	4.480	1.02E-03	1.00E+00	6.09E-03	3	3	1	0.250	1.000
group 2361		hypothetical protein	104	25	13	14	88.89	35.90	4.480	1.02E-03	1.00E+00	6.09E-03	3	3	1	0.250	1.000
gadB 1		Glutamate decarboxylase	104	25	13	14	88.89	35.90	4.480	1.02E-03	1.00E+00	6.09E-03	3	3	1	0.250	1.000
group 2969		Penicillin acylase precursor	104	25	13	14	88.89	35.90	4.480	1.02E-03	1.00E+00	6.09E-03	3	3	1	0.250	1.000
gadC 2		Glutamate/gamma-aminobutyrate antiporter	104	25	13	14	88.89	35.90	4.480	1.02E-03	1.00E+00	6.09E-03	3	3	1	0.250	1.000
group 28	cwlK_1	Peptidoglycan L-alanyl-D-glutamate endopeptidase CwlK precursor	8	11	109	28	6.84	71.79	0.187	1.11E-03	1.00E+00	6.63E-03	7	4	4	1.000	1.000
group 37		hypothetical protein	28	21	89	18	23.93	46.15	0.270	1.18E-03	1.00E+00	7.02E-03	10	7	8	0.109	0.344
group 2506		hypothetical protein	22	0	95	39	18.80	100.00	inf	1.26E-03	1.00E+00	7.40E-03	2	2	0	0.500	0.500
group 3449		hypothetical protein	22	0	95	39	18.80	100.00	inf	1.26E-03	1.00E+00	7.40E-03	2	2	0	0.500	0.500
group 196		hypothetical protein	5	9	112	30	4.27	76.92	0.149	1.28E-03	1.00E+00	7.40E-03	4	2	3	0.625	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 380		hypothetical protein	5	9	112	30	4.27	76.92	0.149	1.28E-03	1.00E+00	7.40E-03	2	1	2	0.500	1.000
group 944		Phage minor structural protein GP20	5	9	112	30	4.27	76.92	0.149	1.28E-03	1.00E+00	7.40E-03	2	1	2	0.500	1.000
group 92		Phage portal protein, SPP1 Gp6-like	5	9	112	30	4.27	76.92	0.149	1.28E-03	1.00E+00	7.40E-03	2	1	2	0.500	1.000
group 274		hypothetical protein	5	9	112	30	4.27	76.92	0.149	1.28E-03	1.00E+00	7.40E-03	2	1	2	0.500	1.000
group 94		Phage minor capsid protein 2	5	9	112	30	4.27	76.92	0.149	1.28E-03	1.00E+00	7.40E-03	2	1	2	0.500	1.000
group 925		hypothetical protein	5	9	112	30	4.27	76.92	0.149	1.28E-03	1.00E+00	7.40E-03	4	2	2	1.000	1.000
group 929		hypothetical protein	5	9	112	30	4.27	76.92	0.149	1.28E-03	1.00E+00	7.40E-03	2	1	2	0.500	1.000
group 3693		Helix-turn-helix domain protein	5	9	112	30	4.27	76.92	0.149	1.28E-03	1.00E+00	7.40E-03	2	1	2	0.500	1.000
group 379		hypothetical protein	5	9	112	30	4.27	76.92	0.149	1.28E-03	1.00E+00	7.40E-03	4	2	2	1.000	1.000
group 4075		hypothetical protein	5	9	112	30	4.27	76.92	0.149	1.28E-03	1.00E+00	7.40E-03	2	1	2	0.500	1.000
group 375		hypothetical protein	5	9	112	30	4.27	76.92	0.149	1.28E-03	1.00E+00	7.40E-03	3	1	3	0.250	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2543		Helix-turn-helix domain protein	5	9	112	30	4.27	76.92	0.149	1.28E-03	1.00E+00	7.40E-03	4	2	2	1.000	1.000
group 52		hypothetical protein	43	4	74	35	36.75	89.74	5.084	1.28E-03	1.00E+00	7.40E-03	7	5	2	0.453	0.453
group 195		hypothetical protein	11	13	106	26	9.40	66.67	0.208	1.32E-03	1.00E+00	7.57E-03	6	3	4	0.688	1.000
group 1155		hypothetical protein	42	26	75	13	35.90	33.33	0.280	1.34E-03	1.00E+00	7.69E-03	4	3	2	0.625	1.000
group 60		hypothetical protein	39	3	78	36	33.33	92.31	6.000	1.49E-03	1.00E+00	8.53E-03	3	3	3	0.250	0.250
group 610		hypothetical protein	39	3	78	36	33.33	92.31	6.000	1.49E-03	1.00E+00	8.53E-03	6	5	1	0.219	0.219
hsdR		Type-1 restriction enzyme R protein	18	16	99	23	15.38	58.97	0.261	1.50E-03	1.00E+00	8.55E-03	2	2	2	0.500	0.500
group 2945		putative type I restriction enzymeP M protein	18	16	99	23	15.38	58.97	0.261	1.50E-03	1.00E+00	8.55E-03	2	2	2	0.500	0.500
group 1004		hypothetical protein	14	14	103	25	11.97	64.10	0.243	1.51E-03	1.00E+00	8.63E-03	5	2	5	0.063	1.000
gbpA		GlcNAc-binding protein A precursor	88	38	29	1	75.21	2.56	0.080	1.70E-03	1.00E+00	9.68E-03	3	1	3	0.250	1.000
group 3562		hypothetical protein	7	10	110	29	5.98	74.36	0.185	1.72E-03	1.00E+00	9.72E-03	4	2	2	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2536		hypothetical protein	7	10	110	29	5.98	74.36	0.185	1.72E-03	1.00E+00	9.72E-03	2	2	2	0.500	0.500
group 2487		hypothetical protein	7	10	110	29	5.98	74.36	0.185	1.72E-03	1.00E+00	9.72E-03	3	2	3	0.250	1.000
group 3530		hypothetical protein	4	8	113	31	3.42	79.49	0.137	1.84E-03	1.00E+00	1.04E-02	4	1	4	0.125	0.625
group 3289		hypothetical protein	4	8	113	31	3.42	79.49	0.137	1.84E-03	1.00E+00	1.04E-02	2	2	1	0.500	1.000
xre		HTH-type transcriptional regulator Xre	51	6	66	33	43.59	84.62	4.250	1.86E-03	1.00E+00	1.05E-02	4	4	1	0.125	0.625
group 3081		hypothetical protein	51	6	66	33	43.59	84.62	4.250	1.86E-03	1.00E+00	1.05E-02	7	6	3	0.125	1.000
yjjV		putative deoxyribonuclease YjjV	35	23	82	16	29.91	41.03	0.297	1.96E-03	1.00E+00	1.08E-02	5	3	4	0.375	1.000
group 2928		hypothetical protein	16	15	101	24	13.68	61.54	0.253	1.97E-03	1.00E+00	1.08E-02	1	1	1	1.000	1.000
group 2362	inlJ_11	Internalin-J precursor	16	15	101	24	13.68	61.54	0.253	1.97E-03	1.00E+00	1.08E-02	1	1	1	1.000	1.000
group 2338		hypothetical protein	16	15	101	24	13.68	61.54	0.253	1.97E-03	1.00E+00	1.08E-02	1	1	1	1.000	1.000
kstR2		HTH-type transcriptional repressor KstR2	16	15	101	24	13.68	61.54	0.253	1.97E-03	1.00E+00	1.08E-02	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2925		hypothetical protein	16	15	101	24	13.68	61.54	0.253	1.97E-03	1.00E+00	1.08E-02	1	1	1	1.000	1.000
group 1466		hypothetical protein	16	15	101	24	13.68	61.54	0.253	1.97E-03	1.00E+00	1.08E-02	8	4	6	0.289	1.000
group 2927		DNA-binding transcriptional repressor PuuR	16	15	101	24	13.68	61.54	0.253	1.97E-03	1.00E+00	1.08E-02	1	1	1	1.000	1.000
group 376		hypothetical protein	16	15	101	24	13.68	61.54	0.253	1.97E-03	1.00E+00	1.08E-02	1	1	1	1.000	1.000
group 142		Reverse transcriptase (RNA-dependent DNA polymerase)	16	15	101	24	13.68	61.54	0.253	1.97E-03	1.00E+00	1.08E-02	1	1	1	1.000	1.000
group 1348		hypothetical protein	16	15	101	24	13.68	61.54	0.253	1.97E-03	1.00E+00	1.08E-02	1	1	1	1.000	1.000
group 580		hypothetical protein	16	15	101	24	13.68	61.54	0.253	1.97E-03	1.00E+00	1.08E-02	7	4	4	1.000	1.000
group 385		Minor capsid protein	42	4	75	35	35.90	89.74	4.900	2.12E-03	1.00E+00	1.16E-02	7	5	2	0.453	0.453
group 3451		hypothetical protein	21	0	96	39	17.95	100.00	inf	2.15E-03	1.00E+00	1.18E-02	2	2	0	0.500	0.500
group 931		hypothetical protein	21	0	96	39	17.95	100.00	inf	2.15E-03	1.00E+00	1.18E-02	2	2	0	0.500	0.500
group 901		hypothetical protein	21	0	96	39	17.95	100.00	inf	2.15E-03	1.00E+00	1.18E-02	3	3	0	0.250	0.250

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 8		Listeria-Bacteroides repeat domain (List_Bact_rpt)	21	17	96	22	17.95	56.41	0.283	2.26E-03	1.00E+00	1.23E-02	5	4	2	0.375	1.000
ssb 2		Single-stranded DNA-binding protein ssb	28	20	89	19	23.93	48.72	0.299	2.36E-03	1.00E+00	1.29E-02	8	6	3	0.289	0.727
group 1441	traG	Conjugal transfer protein TraG	13	13	104	26	11.11	66.67	0.250	2.50E-03	1.00E+00	1.36E-02	7	3	5	0.453	1.000
group 3197		hypothetical protein	13	13	104	26	11.11	66.67	0.250	2.50E-03	1.00E+00	1.36E-02	7	3	5	0.453	1.000
group 193		Mga helix-turn-helix domain protein	13	13	104	26	11.11	66.67	0.250	2.50E-03	1.00E+00	1.36E-02	6	3	4	0.688	1.000
group 1345		hypothetical protein	13	13	104	26	11.11	66.67	0.250	2.50E-03	1.00E+00	1.36E-02	6	5	3	0.219	1.000
group 1139		ABC-2 family transporter protein	57	30	60	9	48.72	23.08	0.285	2.68E-03	1.00E+00	1.45E-02	4	1	4	0.125	0.625
soj 2		Chromosome-partitioning ATPase Soj	88	19	29	20	75.21	51.28	3.194	2.85E-03	1.00E+00	1.54E-02	9	6	5	0.508	1.000
group 2811		hypothetical protein	88	19	29	20	75.21	51.28	3.194	2.85E-03	1.00E+00	1.54E-02	9	6	5	0.508	1.000
group 1475		hypothetical protein	36	3	81	36	30.77	92.31	5.333	2.87E-03	1.00E+00	1.54E-02	5	4	2	0.375	1.000
group 3515		hypothetical protein	36	3	81	36	30.77	92.31	5.333	2.87E-03	1.00E+00	1.54E-02	5	5	2	0.063	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 1474		hypothetical protein	36	3	81	36	30.77	92.31	5.333	2.87E-03	1.00E+00	1.54E-02	5	4	2	0.375	1.000
group 873		HeH/LEM domain protein	11	12	106	27	9.40	69.23	0.233	2.92E-03	1.00E+00	1.54E-02	6	3	5	0.219	1.000
group 151		Minor capsid protein	11	12	106	27	9.40	69.23	0.233	2.92E-03	1.00E+00	1.54E-02	6	3	5	0.219	1.000
group 942		hypothetical protein	11	12	106	27	9.40	69.23	0.233	2.92E-03	1.00E+00	1.54E-02	7	4	3	1.000	1.000
group 2388		Phage capsid family protein	11	12	106	27	9.40	69.23	0.233	2.92E-03	1.00E+00	1.54E-02	6	3	5	0.219	1.000
group 737		hypothetical protein	58	30	59	9	49.57	23.08	0.295	2.93E-03	1.00E+00	1.54E-02	3	2	3	0.250	1.000
mccF 2		Microcin C7 self-immunity protein MccF	58	30	59	9	49.57	23.08	0.295	2.93E-03	1.00E+00	1.54E-02	3	2	3	0.250	1.000
ecfA3		Energy-coupling factor transporter ATP-binding protein EcfA3	58	30	59	9	49.57	23.08	0.295	2.93E-03	1.00E+00	1.54E-02	3	2	3	0.250	1.000
group 1946		hypothetical protein	58	30	59	9	49.57	23.08	0.295	2.93E-03	1.00E+00	1.54E-02	3	2	3	0.250	1.000
ykfA		putative murein peptide carboxypeptidase	59	9	58	30	50.43	76.92	3.391	2.93E-03	1.00E+00	1.54E-02	3	3	2	0.250	1.000
group 3119		HTH domain protein	59	9	58	30	50.43	76.92	3.391	2.93E-03	1.00E+00	1.54E-02	4	4	2	0.125	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
dacA 2		D-alanyl-D-alanine carboxypeptidase DacA precursor	90	38	27	1	76.92	2.56	0.088	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000
group 554		HD domain protein	90	38	27	1	76.92	2.56	0.088	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000
pabB		Aminodeoxychorismate synthase component 1	90	38	27	1	76.92	2.56	0.088	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000
trpG		Anthranilate synthase component 2	90	38	27	1	76.92	2.56	0.088	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000
mtrR		HTH-type transcriptional regulator MtrR	90	38	27	1	76.92	2.56	0.088	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000
group 1919		hypothetical protein	90	38	27	1	76.92	2.56	0.088	3.02E-03	1.00E+00	1.54E-02	3	1	2	1.000	1.000
group 3448	sbcC	Nuclease SbcCD subunit C	27	1	90	38	23.08	97.44	11.400	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000
group 3458		hypothetical protein	27	1	90	38	23.08	97.44	11.400	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000
ftsH 1		ATP-dependent zinc metalloprotease FtsH	27	1	90	38	23.08	97.44	11.400	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000
group 3456		hypothetical protein	27	1	90	38	23.08	97.44	11.400	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000
group 3465		VanZ like family protein	27	1	90	38	23.08	97.44	11.400	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 117		hypothetical protein	27	1	90	38	23.08	97.44	11.400	3.02E-03	1.00E+00	1.54E-02	2	1	1	1.000	1.000
group 3461		hypothetical protein	27	1	90	38	23.08	97.44	11.400	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000
group 3464		hypothetical protein	27	1	90	38	23.08	97.44	11.400	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000
group 3459		hypothetical protein	27	1	90	38	23.08	97.44	11.400	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000
group 3455	mtrR	HTH-type transcriptional regulator MtrR	27	1	90	38	23.08	97.44	11.400	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000
group 3454	dacA_2	D-alanyl-D-alanine carboxypeptidase DacA precursor	27	1	90	38	23.08	97.44	11.400	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000
group 3453	pabB	Aminodeoxychorismate synthase component 1	27	1	90	38	23.08	97.44	11.400	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000
pabA 2		Aminodeoxychorismate synthase component 2	27	1	90	38	23.08	97.44	11.400	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000
group 2510		hypothetical protein	27	1	90	38	23.08	97.44	11.400	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000
group 3364		Helix-turn-helix domain protein	27	1	90	38	23.08	97.44	11.400	3.02E-03	1.00E+00	1.54E-02	5	4	1	0.375	0.375
group 3468		hypothetical protein	27	1	90	38	23.08	97.44	11.400	3.02E-03	1.00E+00	1.54E-02	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3466		hypothetical protein	32	2	85	37	27.35	94.87	6.965	3.05E-03	1.00E+00	1.55E-02	3	2	2	1.000	1.000
group 3467		hypothetical protein	26	1	91	38	22.22	97.44	10.857	3.08E-03	1.00E+00	1.56E-02	1	1	1	1.000	1.000
group 3460		hypothetical protein	26	1	91	38	22.22	97.44	10.857	3.08E-03	1.00E+00	1.56E-02	1	1	1	1.000	1.000
group 1339		hypothetical protein	24	18	93	21	20.51	53.85	0.301	3.14E-03	1.00E+00	1.59E-02	5	4	4	0.375	0.375
group 1371		hypothetical protein	24	18	93	21	20.51	53.85	0.301	3.14E-03	1.00E+00	1.59E-02	9	5	6	0.508	1.000
group 1463		hypothetical protein	31	2	86	37	26.50	94.87	6.669	3.21E-03	1.00E+00	1.62E-02	7	6	2	0.125	0.453
group 1343		hypothetical protein	19	16	98	23	16.24	58.97	0.279	3.25E-03	1.00E+00	1.64E-02	4	3	3	0.625	0.625
group 1341		hypothetical protein	2	6	115	33	1.71	84.62	0.096	3.31E-03	1.00E+00	1.67E-02	3	2	2	1.000	1.000
group 1681		DNA-binding transcriptional repressor PuuR	15	14	102	25	12.82	64.10	0.263	3.39E-03	1.00E+00	1.70E-02	3	3	3	0.250	0.250
inlA 14		Internalin-A precursor	15	14	102	25	12.82	64.10	0.263	3.39E-03	1.00E+00	1.70E-02	5	3	3	1.000	1.000
xre 1		HTH-type transcriptional regulator Xre	15	14	102	25	12.82	64.10	0.263	3.39E-03	1.00E+00	1.70E-02	3	3	3	0.250	0.250

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 384		Phage terminase large subunit	48	6	69	33	41.03	84.62	3.826	3.46E-03	1.00E+00	1.71E-02	9	6	4	0.508	1.000
group 3133		hypothetical protein	48	6	69	33	41.03	84.62	3.826	3.46E-03	1.00E+00	1.71E-02	4	4	1	0.125	0.625
ftsW 2		Lipid II flippase FtsW	0	4	117	35	0.00	89.74	0.000	3.46E-03	1.00E+00	1.71E-02	1	0	1	1.000	1.000
group 2549		hypothetical protein	0	4	117	35	0.00	89.74	0.000	3.46E-03	1.00E+00	1.71E-02	3	0	3	0.250	0.250
group 1369		hypothetical protein	0	4	117	35	0.00	89.74	0.000	3.46E-03	1.00E+00	1.71E-02	4	0	4	0.125	0.125
group 3022		hypothetical protein	0	4	117	35	0.00	89.74	0.000	3.46E-03	1.00E+00	1.71E-02	3	0	3	0.250	0.250
group 2386		hypothetical protein	0	4	117	35	0.00	89.74	0.000	3.46E-03	1.00E+00	1.71E-02	4	0	4	0.125	0.125
group 3712		hypothetical protein	0	4	117	35	0.00	89.74	0.000	3.46E-03	1.00E+00	1.71E-02	1	0	1	1.000	1.000
group 1174	ftsW_1	Lipid II flippase FtsW	117	35	0	4	100.00	10.26	inf	3.46E-03	1.00E+00	1.71E-02	1	1	0	1.000	1.000
group 3529	cas1	CRISPR-associated endonuclease Cas1	5	8	112	31	4.27	79.49	0.173	3.87E-03	1.00E+00	1.88E-02	3	1	3	0.250	1.000
group 4102		hypothetical protein	5	8	112	31	4.27	79.49	0.173	3.87E-03	1.00E+00	1.88E-02	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
cas9		CRISPR-associated endonuclease Cas9	5	8	112	31	4.27	79.49	0.173	3.87E-03	1.00E+00	1.88E-02	3	1	3	0.250	1.000
group 3649		hypothetical protein	5	8	112	31	4.27	79.49	0.173	3.87E-03	1.00E+00	1.88E-02	3	2	1	1.000	1.000
group 4103		hypothetical protein	5	8	112	31	4.27	79.49	0.173	3.87E-03	1.00E+00	1.88E-02	1	1	1	1.000	1.000
group 103		hypothetical protein	5	8	112	31	4.27	79.49	0.173	3.87E-03	1.00E+00	1.88E-02	1	1	1	1.000	1.000
group 4099		hypothetical protein	5	8	112	31	4.27	79.49	0.173	3.87E-03	1.00E+00	1.88E-02	1	1	1	1.000	1.000
group 3539		hypothetical protein	5	8	112	31	4.27	79.49	0.173	3.87E-03	1.00E+00	1.88E-02	3	1	3	0.250	1.000
csn2		CRISPR-associated protein Csn2	5	8	112	31	4.27	79.49	0.173	3.87E-03	1.00E+00	1.88E-02	3	1	3	0.250	1.000
group 3528	cas2	CRISPR-associated endoribonuclease Cas2	5	8	112	31	4.27	79.49	0.173	3.87E-03	1.00E+00	1.88E-02	3	1	3	0.250	1.000
group 4101		YopX protein	5	8	112	31	4.27	79.49	0.173	3.87E-03	1.00E+00	1.88E-02	1	1	1	1.000	1.000
group 4182		hypothetical protein	5	8	112	31	4.27	79.49	0.173	3.87E-03	1.00E+00	1.88E-02	1	1	1	1.000	1.000
group 4100		hypothetical protein	5	8	112	31	4.27	79.49	0.173	3.87E-03	1.00E+00	1.88E-02	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 904		hypothetical protein	19	0	98	39	16.24	100.00	inf	3.94E-03	1.00E+00	1.90E-02	4	4	0	0.125	0.125
group 253		Integrase core domain protein	19	0	98	39	16.24	100.00	inf	3.94E-03	1.00E+00	1.90E-02	1	1	0	1.000	1.000
queC		7-cyano-7-deazaguanine synthase	35	22	82	17	29.91	43.59	0.330	3.95E-03	1.00E+00	1.90E-02	4	3	2	0.625	1.000
group 85		hypothetical protein	35	22	82	17	29.91	43.59	0.330	3.95E-03	1.00E+00	1.90E-02	5	4	3	0.375	1.000
group 2962		hypothetical protein	37	23	80	16	31.62	41.03	0.322	3.99E-03	1.00E+00	1.92E-02	3	3	2	0.250	1.000
group 2125		hypothetical protein	37	23	80	16	31.62	41.03	0.322	3.99E-03	1.00E+00	1.92E-02	3	3	2	0.250	1.000
group 288		carbamoyl phosphate synthase-like protein	20	0	97	39	17.09	100.00	inf	4.04E-03	1.00E+00	1.94E-02	1	1	0	1.000	1.000
group 3198		hypothetical protein	12	12	105	27	10.26	69.23	0.257	4.10E-03	1.00E+00	1.97E-02	5	2	4	0.375	1.000
group 292		hypothetical protein	16	14	101	25	13.68	64.10	0.283	4.26E-03	1.00E+00	2.04E-02	9	5	6	0.508	1.000
group 1931		hypothetical protein	54	8	63	31	46.15	79.49	3.321	4.69E-03	1.00E+00	2.23E-02	2	2	1	0.500	1.000
group 1131		hypothetical protein	54	8	63	31	46.15	79.49	3.321	4.69E-03	1.00E+00	2.23E-02	2	2	1	0.500	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 1132		hypothetical protein	54	8	63	31	46.15	79.49	3.321	4.69E-03	1.00E+00	2.23E-02	2	2	1	0.500	1.000
group 2349		Type I restriction modification DNA specificity domain protein	7	9	110	30	5.98	76.92	0.212	4.83E-03	1.00E+00	2.29E-02	5	2	5	0.063	1.000
group 945		hypothetical protein	7	9	110	30	5.98	76.92	0.212	4.83E-03	1.00E+00	2.29E-02	4	2	2	1.000	1.000
group 266		chromosome segregation protein	7	9	110	30	5.98	76.92	0.212	4.83E-03	1.00E+00	2.29E-02	4	2	3	0.625	1.000
dinB 3		DNA polymerase IV	89	20	28	19	76.07	48.72	3.020	4.85E-03	1.00E+00	2.30E-02	9	7	4	0.180	1.000
group 279		YopX protein	10	11	107	28	8.55	71.79	0.238	4.90E-03	1.00E+00	2.32E-02	5	3	3	1.000	1.000
group 636		hypothetical protein	10	11	107	28	8.55	71.79	0.238	4.90E-03	1.00E+00	2.32E-02	3	2	2	1.000	1.000
group 3083		hypothetical protein	61	10	56	29	52.14	74.36	3.159	5.10E-03	1.00E+00	2.40E-02	3	3	2	0.250	1.000
group 398		hypothetical protein	61	10	56	29	52.14	74.36	3.159	5.10E-03	1.00E+00	2.40E-02	3	3	2	0.250	1.000
group 999		hypothetical protein	56	29	61	10	47.86	25.64	0.317	5.10E-03	1.00E+00	2.40E-02	3	2	3	0.250	1.000
group 3462		hypothetical protein	24	1	93	38	20.51	97.44	9.806	5.53E-03	1.00E+00	2.59E-02	2	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 299	pipB2	Secreted effector protein pipB2	24	1	93	38	20.51	97.44	9.806	5.53E-03	1.00E+00	2.59E-02	5	5	1	0.063	0.375
group 1140		hypothetical protein	57	29	60	10	48.72	25.64	0.328	5.56E-03	1.00E+00	2.61E-02	5	3	4	0.375	1.000
Int-Tn		Transposase from transposon Tn916	18	15	99	24	15.38	61.54	0.291	5.57E-03	1.00E+00	2.61E-02	3	3	2	0.250	1.000
group 1443		Bacterial regulatory proteins, luxR family	4	7	113	32	3.42	82.05	0.162	5.66E-03	1.00E+00	2.64E-02	2	2	2	0.500	0.500
group 3296		hypothetical protein	29	2	88	37	24.79	94.87	6.097	5.69E-03	1.00E+00	2.64E-02	2	2	1	0.500	1.000
group 3298	eccCa1	ESX-1 secretion system protein EccCa1	29	2	88	37	24.79	94.87	6.097	5.69E-03	1.00E+00	2.64E-02	2	2	1	0.500	1.000
group 3295		hypothetical protein	29	2	88	37	24.79	94.87	6.097	5.69E-03	1.00E+00	2.64E-02	2	2	1	0.500	1.000
group 1014		hypothetical protein	14	13	103	26	11.97	66.67	0.272	5.79E-03	1.00E+00	2.69E-02	4	3	3	0.625	0.625
group 2		hypothetical protein	63	11	54	28	53.85	71.79	2.970	5.82E-03	1.00E+00	2.70E-02	14	11	9	0.057	0.424
group 227		hypothetical protein	54	28	63	11	46.15	28.21	0.337	5.82E-03	1.00E+00	2.70E-02	7	2	6	0.125	0.453
group 150		Minor capsid protein	39	4	78	35	33.33	89.74	4.375	6.42E-03	1.00E+00	2.97E-02	7	5	2	0.453	0.453

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2537		hypothetical protein	39	4	78	35	33.33	89.74	4.375	6.42E-03	1.00E+00	2.97E-02	7	5	2	0.453	0.453
group 877		hypothetical protein	15	13	102	26	12.82	66.67	0.294	7.06E-03	1.00E+00	3.25E-02	8	5	4	0.727	1.000
group 2433		hypothetical protein	49	7	68	32	41.88	82.05	3.294	7.08E-03	1.00E+00	3.26E-02	4	3	3	0.625	0.625
group 111		hypothetical protein	6	8	111	31	5.13	79.49	0.209	7.33E-03	1.00E+00	3.37E-02	2	2	1	0.500	1.000
kstR2 2		HTH-type transcriptional repressor KstR2	53	8	64	31	45.30	79.49	3.209	7.58E-03	1.00E+00	3.48E-02	4	4	1	0.125	0.625
group 1473		AP2 domain protein	33	3	84	36	28.21	92.31	4.714	8.04E-03	1.00E+00	3.63E-02	4	3	3	0.625	0.625
xerC 1		Tyrosine recombinase XerC	3	6	114	33	2.56	84.62	0.145	8.06E-03	1.00E+00	3.63E-02	5	2	4	0.375	1.000
group 217		hypothetical protein	90	21	27	18	76.92	46.15	2.857	8.06E-03	1.00E+00	3.63E-02	9	3	7	0.180	0.508
group 2528	ssb_2	Single-stranded DNA-binding protein ssb	34	3	83	36	29.06	92.31	4.916	8.09E-03	1.00E+00	3.63E-02	3	2	3	0.250	1.000
group 3256		Caudovirus prohead protease	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 3244	Int-Tn	Transposase from transposon Tn916	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3252		site-specific tyrosine recombinase XerC	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 3257		Phage capsid family protein	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 3245		hypothetical protein	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 1456		hypothetical protein	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 2476		positive control sigma-like factor	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 1461	ssb_1	Single-stranded DNA-binding protein ssb	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 3254		Phage Terminase	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 3253		Phage terminase, small subunit	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 2475		hypothetical protein	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 3261		Prophage endopeptidase tail	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 3522		hypothetical protein	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	4	3	3	0.625	0.625

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2478		hypothetical protein	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 2479		hypothetical protein	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 3259		hypothetical protein	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 2477		Phage head-tail joining protein	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 2481		Phage tail protein	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 3260	smc_2	Chromosome partition protein Smc	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 1457		hypothetical protein	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 3258		Phage gp6-like head-tail connector protein	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 3255		Phage portal protein	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 1462		hypothetical protein	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 2480		hypothetical protein	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3249		hypothetical protein	9	10	108	29	7.69	74.36	0.242	8.23E-03	1.00E+00	3.63E-02	5	3	2	1.000	1.000
group 2409		hypothetical protein	41	5	76	34	35.04	87.18	3.668	8.33E-03	1.00E+00	3.67E-02	6	6	2	0.031	0.688
group 2309		hypothetical protein	88	20	29	19	75.21	48.72	2.883	8.55E-03	1.00E+00	3.77E-02	9	6	4	0.508	1.000
group 879		Phage holin	28	2	89	37	23.93	94.87	5.820	9.18E-03	1.00E+00	4.03E-02	6	4	2	0.688	0.688
group 1379		Protein gp23 (Bacteriophage A118)	28	2	89	37	23.93	94.87	5.820	9.18E-03	1.00E+00	4.03E-02	6	4	2	0.688	0.688
group 1489		Terminase-like family protein	12	11	105	28	10.26	71.79	0.291	9.53E-03	1.00E+00	4.15E-02	6	4	4	0.688	0.688
group 2678		hypothetical protein	12	11	105	28	10.26	71.79	0.291	9.53E-03	1.00E+00	4.15E-02	2	1	2	0.500	1.000
group 1029		hypothetical protein	12	11	105	28	10.26	71.79	0.291	9.53E-03	1.00E+00	4.15E-02	2	0	2	0.500	0.500
group 1715		ABC-2 type transporter	12	11	105	28	10.26	71.79	0.291	9.53E-03	1.00E+00	4.15E-02	2	1	2	0.500	1.000
group 146	znuC_3	Zinc import ATP-binding protein ZnuC	12	11	105	28	10.26	71.79	0.291	9.53E-03	1.00E+00	4.15E-02	3	2	3	0.250	1.000
group 262		Viral (Superfamily 1) RNA helicase	12	11	105	28	10.26	71.79	0.291	9.53E-03	1.00E+00	4.15E-02	4	3	2	0.625	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 1028		YcaO-like family protein	12	11	105	28	10.26	71.79	0.291	9.53E-03	1.00E+00	4.15E-02	2	1	2	0.500	1.000
group 2948		hypothetical protein	104	27	13	12	88.89	30.77	3.556	9.78E-03	1.00E+00	4.23E-02	3	3	1	0.250	1.000
group 1219		hypothetical protein	104	27	13	12	88.89	30.77	3.556	9.78E-03	1.00E+00	4.23E-02	4	4	2	0.125	1.000
group 2728		Bacillus transposase protein	13	12	104	27	11.11	69.23	0.281	9.78E-03	1.00E+00	4.23E-02	1	1	1	1.000	1.000
group 167	tipA	HTH-type transcriptional activator TipA	13	12	104	27	11.11	69.23	0.281	9.78E-03	1.00E+00	4.23E-02	2	1	2	0.500	1.000
group 2727		hypothetical protein	13	12	104	27	11.11	69.23	0.281	9.78E-03	1.00E+00	4.23E-02	1	1	1	1.000	1.000
group 2420		hypothetical protein	61	11	56	28	52.14	71.79	2.773	1.00E-02	1.00E+00	4.32E-02	2	2	2	0.500	0.500
group 665		ASCH domain protein	23	16	94	23	19.66	58.97	0.352	1.05E-02	1.00E+00	4.54E-02	4	4	2	0.125	1.000
group 3532		hypothetical protein	5	7	112	32	4.27	82.05	0.204	1.10E-02	1.00E+00	4.73E-02	2	1	2	0.500	1.000
group 2535		hypothetical protein	5	7	112	32	4.27	82.05	0.204	1.10E-02	1.00E+00	4.73E-02	2	1	2	0.500	1.000
group 2534		hypothetical protein	5	7	112	32	4.27	82.05	0.204	1.10E-02	1.00E+00	4.73E-02	2	1	2	0.500	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
ugpB		sn-glycerol-3-phosphate-binding periplasmic protein UgpB precursor	115	34	2	5	98.29	12.82	8.456	1.11E-02	1.00E+00	4.74E-02	6	4	2	0.688	0.688
group 893		hypothetical protein	16	0	101	39	13.68	100.00	inf	1.24E-02	1.00E+00	5.32E-02	1	1	0	1.000	1.000
group 2425		hypothetical protein	51	8	66	31	43.59	79.49	2.994	1.27E-02	1.00E+00	5.43E-02	3	2	3	0.250	1.000
group 3093		hypothetical protein	51	8	66	31	43.59	79.49	2.994	1.27E-02	1.00E+00	5.43E-02	3	2	3	0.250	1.000
group 3094		hypothetical protein	51	8	66	31	43.59	79.49	2.994	1.27E-02	1.00E+00	5.43E-02	3	2	3	0.250	1.000
group 2512		hypothetical protein	40	5	77	34	34.19	87.18	3.532	1.35E-02	1.00E+00	5.75E-02	3	3	3	0.250	0.250
group 3463		hypothetical protein	40	5	77	34	34.19	87.18	3.532	1.35E-02	1.00E+00	5.75E-02	3	3	3	0.250	0.250
group 2952		hypothetical protein	1	4	116	35	0.85	89.74	0.075	1.41E-02	1.00E+00	5.98E-02	3	1	3	0.250	1.000
group 1373		hypothetical protein	1	4	116	35	0.85	89.74	0.075	1.41E-02	1.00E+00	5.98E-02	5	1	4	0.375	0.375
group 593		Bacteriophage holin	1	4	116	35	0.85	89.74	0.075	1.41E-02	1.00E+00	5.98E-02	2	1	2	0.500	1.000
group 2556		Bacteriophage holin	1	4	116	35	0.85	89.74	0.075	1.41E-02	1.00E+00	5.98E-02	4	1	3	0.625	0.625

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3704		hypothetical protein	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	1	0	1	1.000	1.000
group 2598		hypothetical protein	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	2	0	2	0.500	0.500
group 2389		hypothetical protein	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	3	0	3	0.250	0.250
group 2576		hypothetical protein	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	2	0	2	0.500	0.500
group 3711	hsdR	Type-1 restriction enzyme R protein	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	1	0	1	1.000	1.000
group 3024		hypothetical protein	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	3	0	3	0.250	0.250
dgt 2		Deoxyguanosinetriphosphate triphosphohydrolase	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	2	0	2	0.500	0.500
group 1508		HEAT repeat protein	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	1	0	1	1.000	1.000
group 1506		putative type I restriction enzymeP M protein	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	1	0	1	1.000	1.000
group 2574		hypothetical protein	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	1	0	1	1.000	1.000
group 2548		Bacterial Ig-like domain (group 2)	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	2	0	2	0.500	0.500

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3023		hypothetical protein	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	3	0	3	0.250	0.250
group 2577		hypothetical protein	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	2	0	2	0.500	0.500
group 280		YopX protein	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	3	0	3	0.250	0.250
group 1492	cwlA	N-acetylmuramoyl-L-alanine amidase CwlA precursor	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	3	0	3	0.250	0.250
group 548	xerC_3	Tyrosine recombinase XerC	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	3	0	3	0.250	0.250
group 3587		hypothetical protein	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	2	0	2	0.500	0.500
group 3705		hypothetical protein	0	3	117	36	0.00	92.31	0.000	1.47E-02	1.00E+00	6.08E-02	1	0	1	1.000	1.000
group 3193		hypothetical protein	15	12	102	27	12.82	69.23	0.331	1.47E-02	1.00E+00	6.08E-02	5	4	4	0.375	0.375
group 3194		hypothetical protein	15	12	102	27	12.82	69.23	0.331	1.47E-02	1.00E+00	6.08E-02	5	4	4	0.375	0.375
group 1130		Ribonuclease	64	30	53	9	54.70	23.08	0.362	1.48E-02	1.00E+00	6.08E-02	4	3	3	0.625	0.625
group 1314		hypothetical protein	64	30	53	9	54.70	23.08	0.362	1.48E-02	1.00E+00	6.08E-02	3	2	3	0.250	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 609		hypothetical protein	11	10	106	29	9.40	74.36	0.301	1.51E-02	1.00E+00	6.10E-02	4	3	3	0.625	0.625
group 934		hypothetical protein	26	2	91	37	22.22	94.87	5.286	1.55E-02	1.00E+00	6.10E-02	2	1	2	0.500	1.000
group 3297		hypothetical protein	26	2	91	37	22.22	94.87	5.286	1.55E-02	1.00E+00	6.10E-02	3	1	2	1.000	1.000
group 2517		hypothetical protein	27	2	90	37	23.08	94.87	5.550	1.57E-02	1.00E+00	6.10E-02	2	1	2	0.500	1.000
group 3470		hypothetical protein	27	2	90	37	23.08	94.87	5.550	1.57E-02	1.00E+00	6.10E-02	1	1	1	1.000	1.000
group 3447		hypothetical protein	27	2	90	37	23.08	94.87	5.550	1.57E-02	1.00E+00	6.10E-02	2	1	2	0.500	1.000
cwlK 1		Peptidoglycan L-alanyl-D-glutamate endopeptidase CwlK precursor	27	2	90	37	23.08	94.87	5.550	1.57E-02	1.00E+00	6.10E-02	5	5	1	0.063	0.375
group 3471		hypothetical protein	27	2	90	37	23.08	94.87	5.550	1.57E-02	1.00E+00	6.10E-02	1	1	1	1.000	1.000
group 936		hypothetical protein	27	2	90	37	23.08	94.87	5.550	1.57E-02	1.00E+00	6.10E-02	2	1	2	0.500	1.000
group 3469	Int-Tn	Transposase from transposon Tn916	27	2	90	37	23.08	94.87	5.550	1.57E-02	1.00E+00	6.10E-02	1	1	1	1.000	1.000
group 3389		PD-(D/E)XK nuclease superfamily protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3366		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3386		Phage regulatory protein Rha (Phage_pRha)	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
adhR 7		HTH-type transcriptional regulator AdhR	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3380		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
polA 2		DNA polymerase I, thermostable	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
dkgA		2,5-diketo-D-gluconic acid reductase A	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3397		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3384		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3401		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3433		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3395		chromosome segregation protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3432		GTP-binding protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3392		RNA polymerase sigma factor	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3373		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3400		putative methyltransferase	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3374		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3375		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
hsdM		Type I restriction enzyme EcoKI M protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
mboIIM		Modification methylase MboII	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
clpP 3		ATP-dependent Clp protease proteolytic subunit	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3430		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3388		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3403		Helix-turn-helix	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3381		VRR-NUC domain protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3371		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3383		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3370		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3393		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3404		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3398		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3391		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3376		putative methyltransferase	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3365		Holin family protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3431		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3377		S-adenosylmethionine synthetase	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3385		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
aacA4		Aminoglycoside N(6')-acetyltransferase type 1	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3394		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3379		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3390		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 3382		hypothetical protein	21	1	96	38	17.95	97.44	8.313	1.58E-02	1.00E+00	6.10E-02	2	2	1	0.500	1.000
group 1392		hypothetical protein	56	10	61	29	47.86	74.36	2.662	1.58E-02	1.00E+00	6.10E-02	3	2	3	0.250	1.000
group 3057		hypothetical protein	56	10	61	29	47.86	74.36	2.662	1.58E-02	1.00E+00	6.10E-02	3	2	3	0.250	1.000
group 2708		hypothetical protein	61	29	56	10	52.14	25.64	0.376	1.58E-02	1.00E+00	6.10E-02	3	3	2	0.250	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3396		Helix-turn-helix	22	1	95	38	18.80	97.44	8.800	1.64E-02	1.00E+00	6.30E-02	3	3	1	0.250	1.000
group 289		carbamoyl phosphate synthase-like protein	22	1	95	38	18.80	97.44	8.800	1.64E-02	1.00E+00	6.30E-02	2	1	1	1.000	1.000
group 1236	fruA_5	PTS system fructose-specific EIIBC component	86	20	31	19	73.50	48.72	2.635	1.66E-02	1.00E+00	6.36E-02	4	1	4	0.125	0.625
group 3118		putative peptidase	59	11	58	28	50.43	71.79	2.589	1.67E-02	1.00E+00	6.39E-02	2	2	2	0.500	0.500
group 1338		hypothetical protein	62	12	55	27	52.99	69.23	2.536	1.72E-02	1.00E+00	6.58E-02	8	5	7	0.070	0.727
group 194		hypothetical protein	9	9	108	30	7.69	76.92	0.278	1.75E-02	1.00E+00	6.68E-02	5	4	4	0.375	0.375
group 823		hypothetical protein	83	35	34	4	70.94	10.26	0.279	1.81E-02	1.00E+00	6.91E-02	3	2	3	0.250	1.000
group 2489		hypothetical protein	34	4	83	35	29.06	89.74	3.584	1.81E-02	1.00E+00	6.91E-02	3	3	2	0.250	1.000
tipA		HTH-type transcriptional activator TipA	104	28	13	11	88.89	28.21	3.143	1.88E-02	1.00E+00	7.15E-02	2	2	1	0.500	1.000
tagG		Teichoic acid translocation permease protein TagG	104	28	13	11	88.89	28.21	3.143	1.88E-02	1.00E+00	7.15E-02	2	2	1	0.500	1.000
group 1112		hypothetical protein	13	11	104	28	11.11	71.79	0.318	1.88E-02	1.00E+00	7.15E-02	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
fosB		Metallothiol transferase FosB	13	11	104	28	11.11	71.79	0.318	1.88E-02	1.00E+00	7.15E-02	1	1	1	1.000	1.000
glnQ 1		Glutamine transport ATP-binding protein GlnQ	13	11	104	28	11.11	71.79	0.318	1.88E-02	1.00E+00	7.15E-02	1	1	1	1.000	1.000
group 93		Phage minor capsid protein 2	45	7	72	32	38.46	82.05	2.857	1.94E-02	1.00E+00	7.34E-02	10	6	5	0.754	1.000
group 91		Phage portal protein, SPP1 Gp6-like	45	7	72	32	38.46	82.05	2.857	1.94E-02	1.00E+00	7.34E-02	10	6	5	0.754	1.000
group 2392		Ftsk gamma domain protein	18	13	99	26	15.38	66.67	0.364	2.06E-02	1.00E+00	7.79E-02	7	4	4	1.000	1.000
group 2730		hypothetical protein	25	16	92	23	21.37	58.97	0.391	2.10E-02	1.00E+00	7.91E-02	4	4	2	0.125	1.000
group 2492		hypothetical protein	14	0	103	39	11.97	100.00	inf	2.18E-02	1.00E+00	8.21E-02	3	3	0	0.250	0.250
group 894		hypothetical protein	14	0	103	39	11.97	100.00	inf	2.18E-02	1.00E+00	8.21E-02	1	1	0	1.000	1.000
group 2403		hypothetical protein	38	5	79	34	32.48	87.18	3.271	2.19E-02	1.00E+00	8.26E-02	6	5	2	0.219	0.688
group 3183		hypothetical protein	30	3	87	36	25.64	92.31	4.138	2.21E-02	1.00E+00	8.31E-02	3	2	3	0.250	1.000
group 935		hypothetical protein	29	3	88	36	24.79	92.31	3.955	2.24E-02	1.00E+00	8.40E-02	4	2	3	0.625	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
hin 1		DNA-invertase hin	77	17	40	22	65.81	56.41	2.491	2.26E-02	1.00E+00	8.46E-02	7	5	4	0.453	1.000
comK 1		Competence transcription factor	77	17	40	22	65.81	56.41	2.491	2.26E-02	1.00E+00	8.46E-02	7	5	4	0.453	1.000
group 461	inlJ_13	Internalin-J precursor	15	0	102	39	12.82	100.00	inf	2.32E-02	1.00E+00	8.69E-02	1	1	0	1.000	1.000
group 1359		hypothetical protein	103	28	14	11	88.03	28.21	2.890	2.33E-02	1.00E+00	8.71E-02	3	2	2	1.000	1.000
group 899		hypothetical protein	14	11	103	28	11.97	71.79	0.346	2.33E-02	1.00E+00	8.71E-02	7	5	4	0.453	1.000
group 2470		hypothetical protein	20	14	97	25	17.09	64.10	0.368	2.35E-02	1.00E+00	8.78E-02	4	3	4	0.125	0.625
group 3076		hypothetical protein	10	9	107	30	8.55	76.92	0.312	2.37E-02	1.00E+00	8.80E-02	4	3	2	0.625	1.000
group 575		hypothetical protein	10	9	107	30	8.55	76.92	0.312	2.37E-02	1.00E+00	8.80E-02	4	3	2	0.625	1.000
group 2464		hypothetical protein	3	5	114	34	2.56	87.18	0.179	2.40E-02	1.00E+00	8.89E-02	5	1	4	0.375	0.375
group 1376		Phage minor structural protein GP20	3	5	114	34	2.56	87.18	0.179	2.40E-02	1.00E+00	8.89E-02	5	1	5	0.063	0.375
group 874	cgkA	Bacterial Ig-like domain (group 2)	3	5	114	34	2.56	87.18	0.179	2.40E-02	1.00E+00	8.89E-02	5	1	5	0.063	0.375

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2504		hypothetical protein	25	2	92	37	21.37	94.87	5.027	2.58E-02	1.00E+00	9.55E-02	2	1	2	0.500	1.000
cwlK		Peptidoglycan L-alanyl-D-glutamate endopeptidase CwlK precursor	41	6	76	33	35.04	84.62	2.967	2.59E-02	1.00E+00	9.59E-02	7	5	3	0.453	1.000
group 1929		hypothetical protein	56	27	61	12	47.86	30.77	0.408	2.60E-02	1.00E+00	9.62E-02	4	3	3	0.625	0.625
group 3071		hypothetical protein	19	1	98	38	16.24	97.44	7.367	2.69E-02	1.00E+00	9.92E-02	3	3	1	0.250	1.000
group 3486		hypothetical protein	20	1	97	38	17.09	97.44	7.835	2.72E-02	1.00E+00	1.00E-01	2	2	1	0.500	1.000
group 3354		hypothetical protein	20	1	97	38	17.09	97.44	7.835	2.72E-02	1.00E+00	1.00E-01	3	3	1	0.250	1.000
group 3549		Acetyltransferase (GNAT) family protein	5	6	112	33	4.27	84.62	0.246	2.92E-02	1.00E+00	1.06E-01	1	1	1	1.000	1.000
group 3543		hypothetical protein	5	6	112	33	4.27	84.62	0.246	2.92E-02	1.00E+00	1.06E-01	1	1	1	1.000	1.000
yofA		HTH-type transcriptional regulator YofA	5	6	112	33	4.27	84.62	0.246	2.92E-02	1.00E+00	1.06E-01	1	1	1	1.000	1.000
group 2541	hsdM	Type I restriction enzyme EcoKI M protein	5	6	112	33	4.27	84.62	0.246	2.92E-02	1.00E+00	1.06E-01	1	1	1	1.000	1.000
group 3531		hypothetical protein	5	6	112	33	4.27	84.62	0.246	2.92E-02	1.00E+00	1.06E-01	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 3548	hsdR	Type-1 restriction enzyme R protein	5	6	112	33	4.27	84.62	0.246	2.92E-02	1.00E+00	1.06E-01	1	1	1	1.000	1.000
znuC 4		Zinc import ATP-binding protein ZnuC	5	6	112	33	4.27	84.62	0.246	2.92E-02	1.00E+00	1.06E-01	1	1	1	1.000	1.000
group 3544		hypothetical protein	5	6	112	33	4.27	84.62	0.246	2.92E-02	1.00E+00	1.06E-01	1	1	1	1.000	1.000
group 3547		EcoKI restriction-modification system protein HsdS	5	6	112	33	4.27	84.62	0.246	2.92E-02	1.00E+00	1.06E-01	1	1	1	1.000	1.000
group 3542		hypothetical protein	5	6	112	33	4.27	84.62	0.246	2.92E-02	1.00E+00	1.06E-01	1	1	1	1.000	1.000
group 3541		hypothetical protein	5	6	112	33	4.27	84.62	0.246	2.92E-02	1.00E+00	1.06E-01	1	1	1	1.000	1.000
group 3545		Recombinase	5	6	112	33	4.27	84.62	0.246	2.92E-02	1.00E+00	1.06E-01	1	1	1	1.000	1.000
group 3546		hypothetical protein	5	6	112	33	4.27	84.62	0.246	2.92E-02	1.00E+00	1.06E-01	1	1	1	1.000	1.000
cysL		HTH-type transcriptional regulator CysL	112	33	5	6	95.73	15.38	4.073	2.92E-02	1.00E+00	1.06E-01	1	1	1	1.000	1.000
gltR 2		HTH-type transcriptional regulator GltR	112	33	5	6	95.73	15.38	4.073	2.92E-02	1.00E+00	1.06E-01	1	1	1	1.000	1.000
nfrA1 2		FMN reductase (NADPH)	112	33	5	6	95.73	15.38	4.073	2.92E-02	1.00E+00	1.06E-01	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 634		CRISPR-associated protein (Cas_CXXC_CXXC)	32	4	85	35	27.35	89.74	3.294	2.93E-02	1.00E+00	1.06E-01	3	2	3	0.250	1.000
group 82		hypothetical protein	84	20	33	19	71.79	48.72	2.418	3.00E-02	1.00E+00	1.08E-01	12	6	12	0.000	1.000
ykoT		putative glycosyltransferase YkoT	12	10	105	29	10.26	74.36	0.331	3.04E-02	1.00E+00	1.10E-01	2	1	2	0.500	1.000
group 676		Nitroreductase family protein	12	10	105	29	10.26	74.36	0.331	3.04E-02	1.00E+00	1.10E-01	3	2	2	1.000	1.000
group 251		Integrase core domain protein	12	10	105	29	10.26	74.36	0.331	3.04E-02	1.00E+00	1.10E-01	7	6	3	0.125	1.000
group 1673		hypothetical protein	17	12	100	27	14.53	69.23	0.383	3.24E-02	1.00E+00	1.16E-01	9	6	6	0.508	0.508
group 307		hypothetical protein	17	12	100	27	14.53	69.23	0.383	3.24E-02	1.00E+00	1.16E-01	9	6	6	0.508	0.508
group 550	inlB_5	Muramidase-2 precursor	93	24	24	15	79.49	38.46	2.422	3.28E-02	1.00E+00	1.18E-01	6	5	4	0.219	0.688
group 31	cwlK	Peptidoglycan L-alanyl-D-glutamate endopeptidase CwlK precursor	2	4	115	35	1.71	89.74	0.152	3.46E-02	1.00E+00	1.20E-01	5	1	4	0.375	0.375
group 2951		hypothetical protein	2	4	115	35	1.71	89.74	0.152	3.46E-02	1.00E+00	1.20E-01	3	1	2	1.000	1.000
group 2953		hypothetical protein	2	4	115	35	1.71	89.74	0.152	3.46E-02	1.00E+00	1.20E-01	3	1	2	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 328		hypothetical protein	2	4	115	35	1.71	89.74	0.152	3.46E-02	1.00E+00	1.20E-01	3	1	2	1.000	1.000
group 2950		hypothetical protein	2	4	115	35	1.71	89.74	0.152	3.46E-02	1.00E+00	1.20E-01	3	1	2	1.000	1.000
group 2311		hypothetical protein	37	20	80	19	31.62	48.72	0.439	3.48E-02	1.00E+00	1.20E-01	9	4	6	0.508	1.000
lytG 3		Exo-glucosaminidase LytG precursor	36	5	81	34	30.77	87.18	3.022	3.48E-02	1.00E+00	1.20E-01	7	5	4	0.453	1.000
group 169		hypothetical protein	82	34	35	5	70.09	12.82	0.345	3.59E-02	1.00E+00	1.20E-01	4	4	2	0.125	1.000
group 741		hypothetical protein	90	36	27	3	76.92	7.69	0.278	3.63E-02	1.00E+00	1.20E-01	2	2	1	0.500	1.000
group 3457		hypothetical protein	27	3	90	36	23.08	92.31	3.600	3.63E-02	1.00E+00	1.20E-01	2	1	2	0.500	1.000
group 2546		hypothetical protein	27	3	90	36	23.08	92.31	3.600	3.63E-02	1.00E+00	1.20E-01	3	1	3	0.250	1.000
rmlC		dTDP-4-dehydrorhamnose 3,5-epimerase	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
epsJ 3		putative glycosyltransferase EpsJ	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
rlmA		23S rRNA (guanine(745)-N(1))-methyltransferase	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
malL 2		Oligo-1,6-glucosidase	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	3	3	2	0.250	1.000
efeM		putative iron uptake system component EfeM precursor	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
tatAy		Sec-independent protein translocase protein TatAy	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 858	inlB_5	Internalin B precursor	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
epsJ 4		putative glycosyltransferase EpsJ	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 1129		hypothetical protein	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	4	4	2	0.125	1.000
malG 2		Maltose transport system permease protein MalG	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	3	3	2	0.250	1.000
tagH		Teichoic acids export ATP-binding protein TagH	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
efeN		putative deferrochelatase/peroxidase EfeN precursor	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2344		hypothetical protein	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
tagB 2		Putative CDP-glycerol:glycerophosphate glycerophosphotransferase	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
degA		HTH-type transcriptional regulator DegA	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	3	3	2	0.250	1.000
group 2341		Bacterial membrane protein YfhO	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
rmlA1		Glucose-1-phosphate thymidylyltransferase 1	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
gtaB		UTP--glucose-1-phosphate uridylyltransferase	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 857		Endonuclease/Exonuclease/phosphatase family protein	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
rmlD		dTDP-4-dehydrorhamnose reductase	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2346		hypothetical protein	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
efeU		Ferrous iron permease EfeU precursor	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2972		hypothetical protein	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
ppdK		Pyruvate, phosphate dikinase	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
malF 2		Maltose transport system permease protein MalF	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	3	3	2	0.250	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
epsJ 1		putative glycosyltransferase EpsJ	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
tatC2		Sec-independent protein translocase protein TatCy	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
ispD2 1		Putative 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase 2	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2924		hypothetical protein	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2930	gutB	Sorbitol dehydrogenase	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
tagB 1		Putative CDP-glycerol:glycerophosphate glycerophosphotransferase	104	29	13	10	88.89	25.64	2.759	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
inlB 6		Internalin B precursor	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 472	inlB_5	hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2704		hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2697	gutB	Sorbitol dehydrogenase	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 1933		hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	2	1	2	0.500	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2755		hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2153		hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2680		hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
tagF		CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2703	tagG	Teichoic acid translocation permease protein TagG	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2791		putative peptidase	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2699		hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 415	inlA_4	Internalin-A precursor	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	3	2	3	0.250	1.000
epsJ		putative glycosyltransferase EpsJ	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2701	lytG_1	Exo-glucosaminidase LytG precursor	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2747		hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 1952	inlB_2	Internalin B precursor	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2754		Undecaprenyl-phosphate mannosyltransferase	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 1719	ispD2_1	Putative 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase 2	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
arnT		Undecaprenyl phosphate-alpha-4-amino-4-deoxy-L-arabinose arabinosyl transferase	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 833	murAB	UDP-N-acetylglucosamine 1-carboxyvinyl transferase 2	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 1005		hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2808		hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 664		hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 1720		hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
wapA 1		tRNA3(Ser)-specific nuclease WapA precursor	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2698	gtaB	UTP--glucose-1-phosphate uridylyltransferase	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2788		Endonuclease/Exonuclease/phosphatase family protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2522		hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	4	2	4	0.125	1.000
group 1935		hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	2	1	2	0.500	1.000
group 663		hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2776		hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2702	tagH	Teichoic acids export ATP-binding protein TagH	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 1993		dihydroneopterin triphosphate pyrophosphatase	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2732		hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2738		Polysaccharide pyruvyl transferase	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2805	blaSE	Glutamyl endopeptidase precursor	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 2134		HTH domain protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2693	tagD	Glycerol-3-phosphate cytidylyltransferase	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 140		Type I restriction modification DNA specificity domain protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	4	3	2	0.625	1.000
group 2694	tagB_1	Putative CDP-glycerol: glycerophosphate glycerophospho-transferase	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 1932		hypothetical protein	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	2	1	2	0.500	1.000
group 2700	tagB_2	Putative CDP-glycerol: glycerophosphate glycerophospho-transferase	13	10	104	29	11.11	74.36	0.363	3.66E-02	1.00E+00	1.20E-01	1	1	1	1.000	1.000
group 3116	sbcC	Nuclease SbcCD subunit C	53	10	64	29	45.30	74.36	2.402	3.80E-02	1.00E+00	1.25E-01	3	3	3	0.250	0.250
ytpP		Thioredoxin-like protein YtpP	105	39	12	0	89.74	0.00	0.000	3.83E-02	1.00E+00	1.25E-01	1	0	1	1.000	1.000
group 4249	ytpP	Thioredoxin-like protein YtpP	12	0	105	39	10.26	100.00	inf	3.83E-02	1.00E+00	1.25E-01	1	1	0	1.000	1.000
group 606		Integrase core domain protein	12	0	105	39	10.26	100.00	inf	3.83E-02	1.00E+00	1.25E-01	2	2	0	0.500	0.500
group 2405		hypothetical protein	27	16	90	23	23.08	58.97	0.431	3.86E-02	1.00E+00	1.26E-01	8	5	5	0.727	0.727
npr		NADH peroxidase	56	11	61	28	47.86	71.79	2.337	3.99E-02	1.00E+00	1.30E-01	7	6	4	0.125	1.000

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
zosA 2		Zinc-transporting ATPase	56	11	61	28	47.86	71.79	2.337	3.99E-02	1.00E+00	1.30E-01	7	6	4	0.125	1.000
group 1178		hypothetical protein	55	11	62	28	47.01	71.79	2.258	4.20E-02	1.00E+00	1.37E-01	5	5	3	0.063	1.000
group 585		hypothetical protein	38	6	79	33	32.48	84.62	2.646	4.21E-02	1.00E+00	1.37E-01	6	4	3	0.688	1.000
group 2432		hypothetical protein	38	6	79	33	32.48	84.62	2.646	4.21E-02	1.00E+00	1.37E-01	5	3	3	1.000	1.000
group 1661		hypothetical protein	4	5	113	34	3.42	87.18	0.241	4.39E-02	1.00E+00	1.42E-01	5	2	3	1.000	1.000
group 409		hypothetical protein	4	5	113	34	3.42	87.18	0.241	4.39E-02	1.00E+00	1.42E-01	4	2	3	0.625	1.000
group 125	inlA_2	Internalin-A precursor	4	5	113	34	3.42	87.18	0.241	4.39E-02	1.00E+00	1.42E-01	5	2	3	1.000	1.000
group 86	inlA_3	Internalin-A precursor	102	28	15	11	87.18	28.21	2.671	4.44E-02	1.00E+00	1.43E-01	3	1	3	0.250	1.000
group 3237		hypothetical protein	15	11	102	28	12.82	71.79	0.374	4.44E-02	1.00E+00	1.43E-01	4	3	4	0.125	0.625
group 515	rlmA	23S rRNA (guanine(745)-N(1))-methyltransferase	15	11	102	28	12.82	71.79	0.374	4.44E-02	1.00E+00	1.43E-01	3	2	2	1.000	1.000
group 927		hypothetical protein	15	11	102	28	12.82	71.79	0.374	4.44E-02	1.00E+00	1.43E-01	4	3	4	0.125	0.625

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 2579		hypothetical protein	17	1	100	38	14.53	97.44	6.460	4.51E-02	1.00E+00	1.45E-01	1	1	1	1.000	1.000
group 3785		hypothetical protein	17	1	100	38	14.53	97.44	6.460	4.51E-02	1.00E+00	1.45E-01	1	1	1	1.000	1.000
group 3783		hypothetical protein	17	1	100	38	14.53	97.44	6.460	4.51E-02	1.00E+00	1.45E-01	1	1	1	1.000	1.000
group 3784		hypothetical protein	17	1	100	38	14.53	97.44	6.460	4.51E-02	1.00E+00	1.45E-01	1	1	1	1.000	1.000
group 3070		hypothetical protein	30	4	87	35	25.64	89.74	3.017	4.63E-02	1.00E+00	1.49E-01	6	5	1	0.219	0.219
group 2393		hypothetical protein	1	3	116	36	0.85	92.31	0.103	4.85E-02	1.00E+00	1.55E-01	4	1	3	0.625	0.625
group 2531		hypothetical protein	1	3	116	36	0.85	92.31	0.103	4.85E-02	1.00E+00	1.55E-01	3	1	3	0.250	1.000
group 3520		hypothetical protein	1	3	116	36	0.85	92.31	0.103	4.85E-02	1.00E+00	1.55E-01	2	1	2	0.500	1.000
group 2384		hypothetical protein	1	3	116	36	0.85	92.31	0.103	4.85E-02	1.00E+00	1.55E-01	3	1	3	0.250	1.000
group 1721		hypothetical protein	11	9	106	30	9.40	76.92	0.346	4.88E-02	1.00E+00	1.55E-01	4	2	3	0.625	1.000
group 1496		hypothetical protein	11	9	106	30	9.40	76.92	0.346	4.88E-02	1.00E+00	1.55E-01	6	4	2	0.688	0.688

Gene name (or Prokka cluster ID)	Non-unique gene name	Annotation	No. repeatedly isolated subtype isolates present in	No. non-repeatedly isolated subtype isolates present in	No. of repeatedly isolated subtype isolates not present in	No. non-repeatedly isolated subtype isolates not present in	Sensitivity	Specificity	Odds ratio	Naive p-value	Bonferroni-corrected p-value	Benjamini-Hochberg-corrected p-value	Max pairwise comparisons	Max supporting pairs	Max opposing pairs	Best pairwise comparison p-value	Worst pairwise comparison p-value
group 211		hypothetical protein	11	9	106	30	9.40	76.92	0.346	4.88E-02	1.00E+00	1.55E-01	7	3	5	0.453	1.000
group 1677		hypothetical protein	11	9	106	30	9.40	76.92	0.346	4.88E-02	1.00E+00	1.55E-01	5	2	4	0.375	1.000
group 2513		hypothetical protein	11	9	106	30	9.40	76.92	0.346	4.88E-02	1.00E+00	1.55E-01	4	2	2	1.000	1.000

[a] Infinity

CHAPTER 5

CONCLUSIONS

The results of the study described in Chapter 3 paint a picture of the ecology and types of strains of *L. monocytogenes* that are present in poultry further processing plants, including genetic diversity, lineages, MLST types, It also gives insight into potential resistance profiles to be concerned about, including the high prevalence of the benzalkonium chloride resistance gene cassette in these isolates, which could indicate that this sanitizer may not be an effective agent to use in food processing plants to eradicate *L. monocytogenes*. The resistance profiles could also be used to evaluate current *Listeria* environmental control strategies in food processing plants and/or create more effective strategies and aid in making risk-based food safety decisions.

The results of the study described in Chapter 4 provides a list of genes that were associated with a "repeatedly isolated" phenotype of the *L. monocytogenes* isolates analyzed in this research. It sets the stage for further research into this area, including delving further into this list of genes to determine which genes may play an integral role in *L. monocytogenes* persistence in food processing environments. This could be accomplished by creating gene knockout mutants for these genes to see if they have a biologically significant role that may play a part in increased survival in food processing environments, including possible increased biofilm formation or attachment, increased resistance to and cleaning and sanitizing agents and environmental stresses (i.e., heat, desiccation, acid, etc.). In addition, these genes could also be evaluated as potential genetic markers for persistence, which could be helpful to distinguish

between potentially-persistent strains isolated from food processing facilities during environmental sampling.

APPENDIX A

NCBI GENBANK ACCESSION NUMBERS

The following table (Table S1) contains the NCBI Genbank accession numbers to access and download the raw read files for each of the *Listeria monocytogenes* isolates used in this study. They are all listed under NCBI BioProject ID PRJNA450812 as well.

Table S1. NCBI Genbank accession numbers for *L. monocytogenes* isolates used in this study.

Isolate	Genbank Accession Number	Isolate	Genbank Accession Number
232	SAMN08947975	401	SAMN08948053
233	SAMN08947976	403	SAMN08948054
235	SAMN08947977	405	SAMN08948055
237	SAMN08947978	406	SAMN08948056
239	SAMN08947979	411	SAMN08948057
241	SAMN08947980	412	SAMN08948058
243	SAMN08947981	431	SAMN08948059
246	SAMN08947982	432	SAMN08948060
247	SAMN08947983	433	SAMN08948061
249	SAMN08947984	437	SAMN08948062
251	SAMN08947985	439	SAMN08948063
253	SAMN08947986	441	SAMN08948064
255	SAMN08947987	442	SAMN08948065
257	SAMN08947988	443	SAMN08948066
259	SAMN08947989	447	SAMN08948067
261	SAMN08947990	456	SAMN08948068
262	SAMN08947991	458	SAMN08948069
263	SAMN08947992	459	SAMN08948070
265	SAMN08947993	460	SAMN08948071
267	SAMN08947994	461	SAMN08948072
277	SAMN08947995	462	SAMN08948073
279	SAMN08947996	464	SAMN08948074
281	SAMN08947997	469	SAMN08948075
282	SAMN08947998	471	SAMN08948076
283	SAMN08947999	472	SAMN08948077

Isolate	Genbank Accession Number	Isolate	Genbank Accession Number
285	SAMN08948000	474	SAMN08948078
287	SAMN08948001	475	SAMN08948079
288	SAMN08948002	480	SAMN08948080
293	SAMN08948003	481	SAMN08948081
295	SAMN08948004	494	SAMN08948082
296	SAMN08948005	497	SAMN08948083
298	SAMN08948006	498	SAMN08948084
299	SAMN08948007	502	SAMN08948085
301	SAMN08948008	508	SAMN08948086
303	SAMN08948009	512	SAMN08948087
305	SAMN08948010	517	SAMN08948088
307	SAMN08948011	522	SAMN08948089
308	SAMN08948012	529	SAMN08948090
311	SAMN08948013	532	SAMN08948091
325	SAMN08948014	533	SAMN08948092
327	SAMN08948015	535	SAMN08948093
329	SAMN08948016	537	SAMN08948094
330	SAMN08948017	538	SAMN08948095
331	SAMN08948018	541	SAMN08948096
333	SAMN08948019	544	SAMN08948097
337	SAMN08948020	546	SAMN08948098
339	SAMN08948021	547	SAMN08948099
340	SAMN08948022	549	SAMN08948100
341	SAMN08948023	550	SAMN08948101
343	SAMN08948024	552	SAMN08948102
345	SAMN08948025	555	SAMN08948103
346	SAMN08948026	557	SAMN08948104
347	SAMN08948027	559	SAMN08948105
348	SAMN08948028	561	SAMN08948106
350	SAMN08948029	565	SAMN08948107
351	SAMN08948030	566	SAMN08948108
357	SAMN08948031	567	SAMN08948109
363	SAMN08948032	569	SAMN08948110
370	SAMN08948033	570	SAMN08948111
371	SAMN08948034	572	SAMN08948112
373	SAMN08948035	574	SAMN08948113
375	SAMN08948036	575	SAMN08948114
377	SAMN08948037	576	SAMN08948115
378	SAMN08948038	578	SAMN08948116
379	SAMN08948039	580	SAMN08948117
380	SAMN08948040	581	SAMN08948118
381	SAMN08948041	582	SAMN08948119
382	SAMN08948042	583	SAMN08948120

Isolate	Genbank Accession Number	Isolate	Genbank Accession Number
383	SAMN08948043	584	SAMN08948121
386	SAMN08948044	585	SAMN08948122
388	SAMN08948045	586	SAMN08948123
390	SAMN08948046	587	SAMN08948124
391	SAMN08948047	588	SAMN08948125
392	SAMN08948048	589	SAMN08948126
393	SAMN08948049	590	SAMN08948127
395	SAMN08948050	591	SAMN08948128
396	SAMN08948051	592	SAMN08948129
398	SAMN08948052	593	SAMN08948130