IDWEEK 2020 | Poster Number: 826

Comparative Epidemiological Analysis of *Serratia marcescens* using PFGE and Whole Genome Sequence Methods

Lalitagauri M. Deshpande, Timothy D. Collingsworth, Rodrigo Mendes, Mariana Castanheira

JMI Laboratories, North Liberty, IA, USA

Introduction

- Epidemiological analyses of bacterial pathogens play an important role in the identification of potential outbreaks for infection control practices.
- Pulsed field gel electrophoresis (PFGE) has been considered the gold standard for epidemiological typing of most bacterial organisms.
  - This method is time consuming, requires considerable technical expertise, specific equipment, and software.
  - Interpretation of results can be subjective and may not be comparable across samples tested in different laboratories.
- Multilocus sequence typing (MLST) is a DNA sequencing-based typing method that uses 5 to 8 highly conserved genes and has been used for identifying clusters of various bacterial species.
  - This method has been broadly used for many bacterial species, but no scheme has been developed for *Serratia marcescens*.
- Core genome (cg)MLST was developed to use whole genome sequencing (WGS) data and analysis of hundreds of genes to determine genetic relationships with greater accuracy than MLST or PFGE.
- We evaluated the utility of WGS ad hoc schemas to predict the genetic relatedness among clinical isolates of *S. marcescens* and compared this method with PFGE results.

Materials and Methods

- A total of 19 *S. marcescens* clinical isolates from 5 countries were selected.
  - These isolates were collected worldwide as part of the SENTRY Antimicrobial Surveillance Program and other JMI collections between 2006–2011.
- Isolates were typed by PFGE and analyzed using the Bionumerics GelCompar II software (Applied Maths, Sint-Martens-Latem, Belgium).
  - Genomic DNA was prepared in agarose blocks and digested with restriction endonuclease SpeI (New England Biolabs, Ipswich, Massachusetts, USA).
  - Electrophoresis was performed on the CHEF-DR II (BioRad, Richmond, California, USA) with the following standard conditions: 0.5 x TBE, 1% agarose, 14°C and 5-60 second switch times at 200V.
  - Isolates showing 100% similarity were scored as identical (assigned with a capital letter; e.g., A), ≥85% and <100% as genetically related (e.g. differed from baseline strain A by ≤3 PFGE bands, designated as subtype A1, A2, A3, etc.) and <85% as unrelated (e.g. differed from strain A by ≥4 bands, designated as strain B, C, D, etc.).
- WGS was performed using MiSeq (Illumina, San Diego, California, USA).
  - Total genomic DNA was used as input material for library construction. DNA libraries were prepared using the Nextera XT™ library construction protocol and index kit (Illumina) and sequenced on a MiSeq Sequencer using MiSeq Reagent Kit v3 (600 cycle).
- Raw reads were used to a generate a similarity matrix and phylogenetic tree using kWIP.
  - A matrix of pairwise distances amenable to qualitative comparison of genetic distances between samples within a population were estimated by calculating the weighted inner product (WIP) based on informational entropy across each analysis set.
  - The R package 'hclust' was employed to generate a dendrogram for visual analysis.
- Assembled contigs were used to generate a similarity matrix and phylogenetic tree using ProgressiveMauve, Gubbins and RAxML (Dabos et al., 2019).
  - Error-corrected reads (BayesHammer) were used to generate contigs using the de novo assembler SPAdes version 3.11.1 for each sample set independently. K-mer values were automatically selected by read length with settings to reduce the number of mismatches, producing a FASTA format file of contiguous sequences with the best N50 value.
  - ProgressiveMauve and the Mauve-distributed script stripSubsetLCBs were used to extract all regions with a minimum length of 500 bp shared by all the contigs, generating a core genome alignment.
  - Recombinant variable positions were removed with Gubbins and a maximum likelihood phylogeny was inferred for the alignment of non-recombinant variable sites with RAxML using a general time-reversible (GTR) model incorporating a gamma distribution rate among sites.
  - The resulting dendrogram was visualized using the ETEToolkit.
- The matrices and dendrograms created by the 3 protocols were compared.

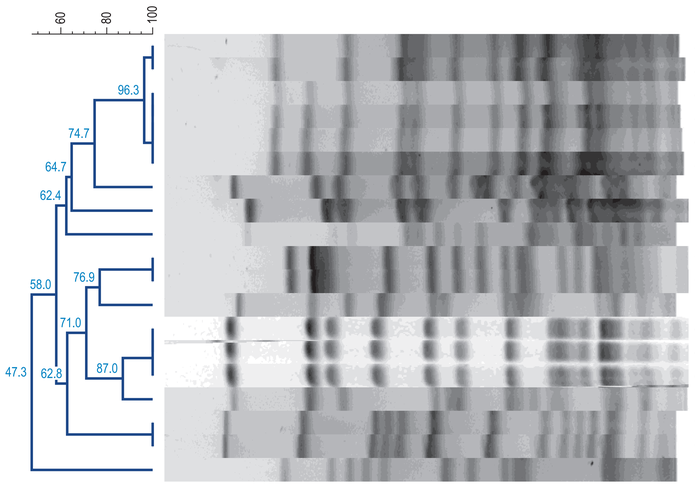
Results

- Based on PFGE analysis, 19 isolates were classified into 10 pulsotypes, A through J, and 1 subtype, A1 (Figure 1).
  - Five isolates from Japan belonged to 4 pulsotypes (C, D, H, and I).
  - Three isolates from Mexico showed identical PFGE pattern (J).
  - Four USA isolates belonged to three pulsotypes (B, E and G).
  - Six isolates from France were related (A and A1).
  - One isolate from Poland had an unrelated pattern to all other isolates (F).
- Figure 2A showed that isolates belonging to PFGE clusters A, C, G, and J had 100% similarity within the group and 96.3% similarity for the subtype A1 compared to isolates belonging to pulsotype A.
- Figure 2B, displaying the kWIP analysis that uses short k-mers for analysis, showed much lower percent homologies when compared to other methods.
  - Analysis applying this method showed that isolates from cluster A ranged from 36.5 to 84.4% similarity, isolates from cluster G had 36.3% similarity, cluster J had 83.2 to 84.0% similarity, and cluster C isolates had a similarity only of 0.3%.
- Figure 2C, displaying the matrix generated using Dabos et al., 2019 protocol that uses longer contigs for analysis, shows that isolates from clusters C and J had homology percentages ranging from 99.7% to 99.9%.
  - Isolates from pulsotype A and subtype A1 displayed homology of 99.2% to 99.9%, with the exception of 1 isolate that displayed only 40.1% homology to the other isolates from this cluster.
  - Isolates belonging to cluster G displayed only 39.1% similarity when using this method.
- The kWIP analysis (Figure 2B) did not show any isolates belonging to unique PFGE patterns, as similarity was >57% when compared to other isolates.
- The Dabos et al., 2019 protocol (Figure 2C) assigned homologies to clusters C, J, and F. Additionally, 1 isolate belonged to cluster G at >90% homology.

Conclusions

- The comparison of PFGE and 2 WGS-based phylogenetic analysis of *S. marcescens* demonstrated that the method using short k-mers, kWIP, produced discrepant results from PFGE, which is still considered the gold-standard method for epidemiologic typing.
- The method using longer contiguous sequences, the Dabos et al., 2019 protocol, displayed better correlation with PFGE, with exception of 1 cluster. Additionally, this protocol displayed relatively high cut-off values that would have to be applied due to some high similarity observed among unrelated isolates.
- WGS-based methods for epidemiologic typing can be used for the analysis of *S. marcescens*, but some optimization is still required.

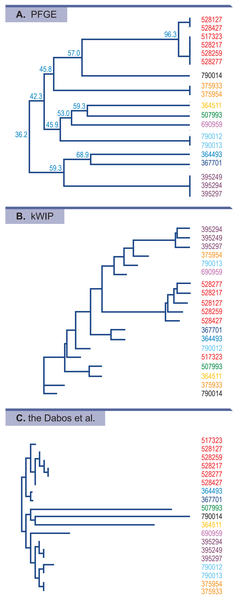
Figure 1 Gelcompar II analysis of PFGE patterns obtained from *Serratia marcescens* isolates



Pulsotype	JMI collection number	Organism
A	528127	*Serratia marcescens*
A	528427	*Serratia marcescens*
A1	517323	*Serratia marcescens*
A1	528217	*Serratia marcescens*
A1	528259	*Serratia marcescens*
A1	528277	*Serratia marcescens*
E	507993	*Serratia marcescens*
B	790014	*Serratia marcescens*
D	364511	*Serratia marcescens*
G	790012	*Serratia marcescens*
G	790013	*Serratia marcescens*
I	367701	*Serratia marcescens*
J	395249	*Serratia marcescens*
J	395294	*Serratia marcescens*
J	395297	*Serratia marcescens*
F	690959	*Serratia marcescens*
C	375933	*Serratia marcescens*
C	375954	*Serratia marcescens*
H	364493	*Serratia marcescens*

Figure 2 Comparison of PFGE and WGS- based analysis of *Serratia marcescens* isolates.

	A 528127	A 528427	A1 517323	A1 528217	A1 528259	A1 528277	B 790014	C 375933	C 375954	D 364511	E 507993	F 690959	G 790012	G 790013	H 364493	I 367701	J 395249	J 395294	J 395297
A 528127	100.0	100.0	88.9	96.3	96.3	96.3	57.2	51.9	44.5	38.5	51.9	41.7	37.0	37.0	33.3	40.0	25.0	25.0	25.0
A 528427	100.0	100.0	96.3	96.3	96.3	96.3	57.2	51.9	51.9	46.2	59.3	33.3	44.5	44.5	33.3	40.0	25.0	33.3	25.0
A1 517323	88.9	96.3	100.0	100.0	100.0	100.0	55.2	50.0	35.7	51.9	28.6	48.0	42.9	50.0	32.0	46.2	24.0	24.0	32.0
A1 528217	96.3	96.3	100.0	100.0	100.0	100.0	62.1	50.0	50.0	44.5	50.0	48.0	42.9	35.7	24.0	46.2	32.0	32.0	32.0
A1 528259	96.3	96.3	100.0	100.0	100.0	100.0	48.3	42.9	42.9	44.5	42.9	40.0	35.7	35.7	24.0	38.5	32.0	32.0	32.0
A1 528277	96.3	96.3	100.0	100.0	100.0	100.0	62.1	50.0	50.0	44.5	50.0	40.0	50.0	42.9	16.0	38.5	24.0	32.0	24.0
B 790014	57.2	57.2	55.2	62.1	48.3	62.1	100.0	34.5	34.5	50.0	48.3	38.5	34.5	27.6	30.8	37.0	46.2	46.2	46.2
C 375933	51.9	51.9	50.0	50.0	42.9	50.0	34.5	100.0	100.0	29.6	50.0	40.0	42.9	42.9	24.0	53.9	40.0	40.0	40.0
C 375954	44.5	51.9	35.7	50.0	42.9	50.0	34.5	100.0	100.0	29.6	50.0	32.0	42.9	42.9	32.0	38.5	40.0	40.0	40.0
D 364511	38.5	46.2	51.9	44.5	44.5	44.5	50.0	29.6	29.6	100.0	59.3	50.0	44.5	51.9	25.0	48.0	25.0	25.0	25.0
E 507993	51.9	59.3	28.6	50.0	42.9	50.0	48.3	50.0	50.0	59.3	100.0	56.0	42.9	50.0	40.0	46.2	48.0	48.0	48.0
F 690959	41.7	33.3	48.0	48.0	40.0	40.0	38.5	40.0	32.0	50.0	56.0	100.0	56.0	48.0	27.3	60.9	45.5	54.6	45.5
G 790012	37.0	44.5	42.9	42.9	35.7	50.0	34.5	42.9	42.9	44.5	42.9	56.0	100.0	100.0	32.0	53.9	40.0	40.0	40.0
G 790013	37.0	44.5	50.0	35.7	35.7	42.9	27.6	42.9	42.9	51.9	50.0	48.0	100.0	100.0	32.0	53.9	32.0	40.0	40.0
H 364493	33.3	33.3	32.0	24.0	24.0	16.0	30.8	24.0	32.0	25.0	40.0	27.3	32.0	32.0	100.0	60.9	45.5	45.5	45.5
I 367701	40.0	40.0	46.2	46.2	38.5	38.5	37.0	53.9	38.5	48.0	46.2	60.9	53.9	53.9	60.9	100.0	52.2	52.2	60.9
J 395249	25.0	25.0	24.0	32.0	32.0	24.0	46.2	40.0	40.0	25.0	48.0	45.5	40.0	32.0	45.5	52.2	100.0	100.0	100.0
J 395294	25.0	33.3	24.0	32.0	32.0	32.0	46.2	40.0	40.0	25.0	48.0	54.6	40.0	40.0	45.5	52.2	100.0	100.0	100.0
J 395297	25.0	25.0	32.0	32.0	32.0	24.0	46.2	40.0	40.0	25.0	48.0	45.5	40.0	40.0	45.5	60.9	100.0	100.0	100.0
A 528127	100.0	77.6	36.5	82.0	78.4	82.2	0.3	0.3	44.1	23.6	23.6	38.2	23.2	40.2	42.1	31.6	43.9	43.9	43.9
A 528427	77.6	100.0	37.0	78.1	76.7	77.4	0.2	0.4	42.3	24.4	23.4	38.1	23.1	40.1	40.6	31.6	42.5	42.4	42.3
A1 517323	36.5	37.0	100.0	37.0	37.8	36.9	0.1	0.3	23.0	15.5	13.7	21.8	14.7	23.7	22.6	19.8	23.1	23.1	23.0
A1 528217	82.0	78.1	37.0	100.0	81.2	84.4	0.2	0.3	43.9	23.8	23.6	38.4	23.2	40.2	42.0	31.7	44.1	44.0	43.9
A1 528259	78.4	76.7	37.8	81.2	100.0	81.1	0.2	0.3	42.0	23.7	23.0	37.2	22.8	39.3	40.4	31.1	42.2	42.1	42.0
A1 528277	82.2	77.4	36.9	84.4	81.1	100.0	0.2	0.3	44.0	23.6	23.6	38.2	23.1	40.1	42.0	31.6	44.0	44.0	44.0
B 790014	0.3	0.2	0.1	0.2	0.2	0.2	100.0	0.0	0.2	0.2	0.1	0.2	0.0	0.2	0.2	0.1	0.3	0.3	0.3
C 375933	0.3	0.4	0.3	0.3	0.3	0.3	0.0	100.0	0.3	0.0	0.1	0.3	0.5	0.4	0.3	0.5	0.3	0.3	0.3
C 375954	44.1	42.3	23.0	43.9	42.0	44.0	0.2	0.3	100.0	24.6	23.6	47.1	28.4	50.7	43.4	32.9	56.9	56.9	56.8
D 364511	23.6	24.4	15.5	23.8	23.7	23.6	0.2	0.0	24.6	100.0	26.6	23.3	15.5	25.9	22.9	20.1	25.2	25.1	25.2
E 507993	23.6	23.4	13.7	23.6	23.0	23.6	0.1	0.1	23.6	26.6	100.0	23.7	13.5	22.6	22.5	17.3	23.9	23.8	23.8
F 690959	38.2	38.1	21.8	38.4	37.2	38.2	0.2	0.3	47.1	23.3	23.7	100.0	25.4	44.2	38.6	29.6	47.1	47.0	46.9
G 790012	23.2	23.1	14.7	23.2	22.8	23.1	0.0	0.5	28.4	15.5	13.5	25.4	100.0	36.3	23.0	20.2	28.7	28.8	28.6
G 790013	40.2	40.1	23.7	40.2	39.3	40.1	0.2	0.4	50.7	25.9	22.6	44.2	36.3	100.0	39.9	33.9	51.8	51.8	51.6
H 364493	42.1	40.6	22.6	42.0	40.4	42.0	0.2	0.3	43.4	22.9	22.5	38.6	23.0	39.9	100.0	42.4	43.6	43.6	43.6
I 367701	31.6	31.6	19.8	31.7	31.1	31.6	0.1	0.5	32.9	20.1	17.3	29.6	20.2	33.9	42.4	100.0	33.1	33.1	33.1
J 395249	43.9	42.5	23.1	44.1	42.2	44.0	0.3	0.3	56.9	25.2	23.9	47.1	28.7	51.9	43.6	33.1	100.0	84.0	83.5
J 395294	43.9	42.4	23.1	44.0	42.1	44.0	0.3	0.3	56.9	25.1	23.8	47.0	28.8	51.8	43.6	33.1	84.0	100.0	83.2
J 395297	43.9	42.3	23.0	43.9	42.0	44.0	0.3	0.3	56.8	25.2	23.8	46.9	28.6	51.6	43.6	33.1	83.5	83.2	100.0
A 528127	100.0	99.9	40.0	99.3	99.3	99.3	79.0	84.2	84.3	78.9	76.3	84.5	33.2	85.6	83.4	57.7	86.5	86.5	86.5
A 528427	99.9	100.0	40.0	99.3	99.2	99.3	79.0	84.2	84.3	78.9	76.3	84.5	33.2	85.6	83.4	57.7	86.5	86.5	86.5
A1 517323	40.0	40.0	100.0	40.1	40.0	40.1	30.2	31.0	31.1	31.3	30.5	31.8	0.0	31.3	31.8	21.4	31.3	31.2	31.3
A1 528217	99.3	99.3	40.1	100.0	99.8	99.9	79.4	84.6	84.7	79.3	76.6	84.9	33.3	86.0	83.8	58.0	86.9	86.9	86.9
A1 528259	99.3	99.2	40.0	99.8	100.0	99.9	79.3	84.6	84.7	79.3	76.5	84.9	33.2	86.0	83.8	58.0	86.9	86.9	86.9
A1 528277	99.3	99.3	40.1	99.9	99.9	100.0	79.4	84.6	84.7	79.3	76.6	84.9	33.3	86.0	83.8	58.0	86.9	86.9	86.9
B 790014	79.0	79.0	30.2	79.4	79.3	79.4	100.0	80.5	80.6	85.4	79.3	80.8	32.2	82.2	79.0	57.8	83.2	83.2	83.2
C 375933	84.2	84.2	31.0	84.6	84.6	84.6	80.5	100.0	99.7	78.7	76.0	91.0	36.2	91.5	84.2	58.8	92.4	92.4	92.4
C 375954	84.3	84.3	31.1	84.7	84.7	84.7	80.6	99.7	100.0	78.7	76.0	91.1	36.3	91.6	84.3	58.9	92.6	92.5	92.6
D 364511	78.9	78.9	31.3	79.3	79.3	79.3	86.4	78.7	78.7	100.0	76.7	79.5	31.9	80.9	78.9	57.7	81.8	81.8	81.8
E 507993	76.3	76.3	30.5	76.6	76.5	76.6	79.3	76.0	76.0	76.7	100.0	76.7	31.6	77.2	75.0	54.8	78.0	78.0	78.1
F 690959	84.5	84.5	31.8	84.9	84.9	84.9	80.8	91.0	91.1	79.5	76.7	100.0	36.5	91.6	85.0	59.1	92.7	92.7	92.7
G 790012	33.2	33.2	0.0	33.3	33.2	33.3	32.2	36.2	36.3	31.9	31.6	36.5	100.0	39.1	33.1	22.9	37.1	37.1	37.1
G 790013	85.6	85.6	31.3	86.0	86.0	86.0	82.2	91.5	91.6	80.9	77.2	91.6	39.1	100.0	85.8	59.9	95.5	95.5	95.5
H 364493	83.4	83.4	31.8	83.8	83.8	83.8	79.0	84.2	84.3	78.9	75.0	85.0	33.1	85.8	100.0	63.0	86.9	86.9	86.9
I 367701	57.7	57.7	21.4	58.0	58.0	58.0	57.8	58.8	58.9	57.7	54.8	59.1	22.9	59.9	63.0	100.0	60.5	60.5	60.5
J 395249	86.5	86.5	31.3	86.9	86.9	86.9	83.2	92.4	92.6	81.8	78.0	92.7	37.1	95.5	86.9	60.5	100.0	99.9	99.9
J 395294	86.5	86.5	31.2	86.9	86.9	86.9	83.2	92.4	92.5	81.8	78.0	92.7	37.1	95.5	86.9	60.5	99.9	100.0	99.9
J 395297	86.5	86.5	31.3	86.9	86.9	86.9	83.2	92.4	92.6	81.8	78.1	92.7	37.1	95.5	86.9	60.5	99.9	99.9	100.0



Three methods are compared: (A) PFGE, (B) kWIP and (C) Dabos et al., 2019. For each method, a similarity matrix (left) as well as corresponding dendrogram (right) are shown. Matrix rows and columns are labeled by JMI collection number and ordered by pulsotype (A through J). Matrices are colored by similarity: from red (high) to yellow (intermediate) to green (low). Dendrogram leaves are labeled with JMI collection number and colored according to pulsotype: red (A), black (B), orange (C), yellow (D), green (E), pink (F), cyan (G), light blue (H), dark blue (I), and purple (J).

Acknowledgements

The authors appreciate the expert assistance from Sara Fort and Judy Oberholser at JMI Laboratories. This study and poster presentation were supported by JMI Laboratories.

References

Bankevich A, Nurk S, Antipov D, et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19: 455–477.

Croucher NJ, Page AJ, Connor TR, et al. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 43: e15.

Dabos L, Patino-Navarrete R, Nastro M, et al. (2019). SME-4-producing *Serratia marcescens* from Argentina belonging to clade 2 of the *S. marcescens* phylogeny. *J Antimicrob Chemother* 74: 1836–1841.

Darling AE, Mau B, Perna NT (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5: e11147.

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

Murray KD, Webers C, Ong CS, et al. (2017). kWIP: The k-mer weighted inner product, a de novo estimator of genetic similarity. *PLoS Comput Biol* 13: e1005727.

Nikolenko SI, Korobeynikov AI, Alekseyev MA (2013). BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* 14 Suppl 1: S1-S7.

Schurch AC, Arredondo-Alonso S, Willems RJL, et al. (2018). Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. *Clin Microbiol Infect* 24: 350–354.

Stamatakis A (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.

Team RC (2018). R Team (2018): A language and environment for statistical computing. R Foundation for Statistical Computing. Available at https://www.r-project.org/. Accessed December 2018.

Tenover FC, Arbeit RD, Goering RV, et al. (1995). Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol* 33: 2233–2239.

Tenover FC, Arbeit RD, Goering RV (1997). How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. Molecular Typing Working Group of the Society for Healthcare Epidemiology of America. *Infect Control Hosp Epidemiol* 18: 426–439.

Contact

Mariana Castanheira
JMI Laboratories
345 Beaver Kreek Centre, Suite A
North Liberty, Iowa 52317
Phone: (319) 665-3370
Fax: (319) 665-3371
Email: mariana-castanheira@jmilabs.com