



# Construction of a Pan-Genome Allele Database of *Salmonella enterica* Serovar Enteritidis for Molecular Subtyping and Disease Cluster Identification

Yen-Yi Liu<sup>†</sup>, Chih-Chieh Chen<sup>2,3†</sup> and Chien-Shun Chiou<sup>1\*</sup>

<sup>1</sup> Central Regional Laboratory, Center for Diagnostics and Vaccine Development, Centers for Disease Control, Taichung, Taiwan, <sup>2</sup> Institute of Medical Science and Technology, National Sun Yat-sen University, Kaohsiung, Taiwan, <sup>3</sup> Medical Science and Technology Center, National Sun Yat-sen University, Kaohsiung, Taiwan

## OPEN ACCESS

### Edited by:

John W. A. Rossen,  
University Medical Center Groningen,  
Netherlands

### Reviewed by:

Jozsef Soki,  
University of Szeged, Hungary  
Ivo Van Walle,  
European Centre for Disease  
Prevention and Control, Sweden

### \*Correspondence:

Chien-Shun Chiou  
nipmcs@cdc.gov.tw

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
Frontiers in Microbiology

**Received:** 23 September 2016

**Accepted:** 30 November 2016

**Published:** 15 December 2016

### Citation:

Liu Y-Y, Chen C-C and Chiou C-S  
(2016) Construction of a  
Pan-Genome Allele Database  
of *Salmonella enterica* Serovar  
Enteritidis for Molecular Subtyping  
and Disease Cluster Identification.  
*Front. Microbiol.* 7:2010.  
doi: 10.3389/fmicb.2016.02010

We built a pan-genome allele database with 395 genomes of *Salmonella enterica* serovar Enteritidis and developed computer tools for analysis of whole genome sequencing (WGS) data of bacterial isolates for disease cluster identification. A web server (<http://wgmlst.imst.nsysu.edu.tw>) was set up with the database and the tools, allowing users to upload WGS data to generate whole genome multilocus sequence typing (wgMLST) profiles and to perform cluster analysis of wgMLST profiles. The usefulness of the database in disease cluster identification was demonstrated by analyzing a panel of genomes from 55 epidemiologically well-defined *S. Enteritidis* isolates provided by the Minnesota Department of Health. The wgMLST-based cluster analysis revealed distinct clades that were concordant with the epidemiologically defined outbreaks. Thus, using a common pan-genome allele database, wgMLST can be a promising WGS-based subtyping approach for disease surveillance and outbreak investigation across laboratories.

**Keywords:** next generation sequencing (NGS), pan-genome allele database, whole genome multilocus sequence typing (wgMLST), typing, molecular epidemiology

## INTRODUCTION

Characterization of bacterial isolates using various subtyping methods has been a fundamental work for epidemiologic study of infectious diseases. Pulsed-field gel electrophoresis (PFGE) is currently the gold standard of molecular subtyping tool in discriminating between bacterial strains for disease surveillance and outbreak investigation. PFGE has been the common subtyping tool employed in the laboratories of PulseNet, a molecular subtyping network for foodborne bacterial disease surveillance (Swaminathan et al., 2001). However, PFGE is labor-intensive and time-consuming and lacks sufficient resolution for highly clonal bacterial strains (Boxrud et al., 2007; Liang et al., 2007). Over the past two decades, public health laboratories have become eager for a new subtyping tool to provide better resolution than PFGE for examining highly clonal bacterial strains. With the advance of next generation sequencing (NGS) techniques, whole genome sequencing (WGS) has become a rapid and inexpensive method for the characterization of bacteria. WGS-based analysis has been increasingly used in public health laboratories to characterize bacterial pathogens and perform subtyping for epidemiologic study (Deng et al., 2016; Fratamico et al., 2016; Lindsey et al., 2016).

The current second generation sequencing platforms usually generate millions of short sequences (reads) from a bacterial genome. It has been a great challenge to analyze the vast number of short sequences to obtain specific information. Several sequence analysis tools and bioinformatic pipelines have been developed, and some have been installed on the website of the Center for Genome Epidemiology<sup>1</sup> so WGS data may be used for identification of bacterial species, resistance genes, virulence factors, serotypes, and analysis of phylogeny. However, public health laboratories are in need of tools that can analyze WGS data to generate a type of portable genetic fingerprint (genotype) for use in discriminating among closely related strains for inter-laboratory perspective disease surveillance and outbreak investigation.

The single-nucleotide polymorphism (SNP)-based method is a widely used approach in which WGS data are employed for high-resolution subtyping of bacterial strains (Taylor et al., 2015; Bekal et al., 2016). To apply the SNP-based approach, a reference genome sequence is required for calling SNPs from WGS data of strains. A potential drawback is that the use of different reference genomes between studies would result in different SNP profiles, making it difficult to compare results between studies. Whole genome multilocus sequence typing (wgMLST), an extension of traditional MLST (Maiden et al., 1998), is a genome wide gene-by-gene comparison approach (Maiden et al., 2013). This wgMLST-based approach has been applied to analyze WGS data for detection of disease clusters and outbreak investigation (de Been et al., 2015; Jackson et al., 2016; Jonathan et al., 2016). To make the wgMLST scheme a standard subtyping tool, a pan-genome allele database that contains genes present in the population of a bacterial organism has to be established first. wgMLST profiles generated from a common pan-genome allele database can be portable and comparable across laboratories.

In this study, we constructed an *S. Enteritidis* pan-genome allele database for analysis of WGS data of bacterial isolates. A web server with the database was built, and computer tools were installed in the website to allow users to generate wgMLST profiles and compare genetic relatedness among bacterial isolates. The usefulness of the database in identification of disease clusters was assessed using genomes from a panel of epidemiologically well-characterized *S. Enteritidis* isolates provided by the Minnesota Department of Health.

## MATERIALS AND METHODS

### Building an *S. Enteritidis* Pan-Genome Allele Database

*Salmonella enterica* serovar Enteritidis pan-genome allele database was constructed with 340 *S. Enteritidis* genomic sequences retrieved from the NCBI Assembly database<sup>2</sup> and 55 *S. Enteritidis* genomic sequences provided by Minnesota Department of Health<sup>3</sup> (Supplementary Table S1). The 395

complete genomic sequences or genomic contigs were first annotated using Prokka (Seemann, 2014), a rapid bacterial genome annotation pipeline, to generate output gff files. The gff files were processed using the PGADB\_builder, a pipeline for construction of a bacterial pan-genome allele database (Liu et al., 2016). In this study, paralogous genes were excluded from the dataset, and proteins sharing  $\geq 95\%$  amino acid sequence identity were grouped in an orthologous cluster (a protein family). A protein family was assigned to be a locus, and each protein in a locus was transferred back to its nucleotide sequence by referring to the ffn file that was created in the annotation step. Nucleotide sequences in a locus differing by a nucleotide or more from each other are defined as different alleles. Loci and alleles of a pan-genome allele dataset are encoded with a standardized numbering system.

### wgMLST Profiling

The wgMLST\_profiling tool was developed to generate wgMLST profiles from bacterial genomes in use of the *S. Enteritidis* pan-genome allele database. A wgMLST\_profiling was installed on the website, <http://wgmlst.imst.nsysu.edu.tw>, to allow users to upload assembled whole genome contigs to generate wgMLST profiles. For profiling, the longest allele (or the first one as two or more allelic sequences having the same length) for each locus was selected to make up a reference sequence set (RSS). Query genomic contigs were compared with the RSS using BLASTN algorithm (Altschul et al., 1990). In the present study, the presence of a locus in a query genome was defined as existence of a sequence that shared  $\geq 90\%$  length coverage and  $\geq 90\%$  sequence identity with one reference sequence of the RSS. Sequences from the query genome were subsequently compared with all allelic sequences in the locus and given a digital number for the locus by a designated numbering system.

### Analysis of Genetic Relatedness

A DendroMaker tool was developed and installed in the web service site for genetic relatedness analysis of bacterial strains. The program used Manhattan distance coefficient and unweighted pair group method with arithmetic mean (UPGMA) algorithm for cluster analysis of wgMLST profiles.

### Web Service

A web server (SE-PGADB) was built with *S. Enteritidis* pan-genome allele database and wgMLST\_Profiling and DendroMaker tools written in PHP scripts. The web page<sup>4</sup> was constructed in HTML, JavaScript, and PHP formats. The server runs on a Linux cluster with 2.40 GHz Intel Xeon processors comprising 24 cores.

### Input Format

The wgMLST\_Profiling module accepts genome contigs in FASTA format. Sequence comparison is performed using the BLASTN algorithm. The maximal number of genomes uploaded for profiling is set at 99. The DendroMaker module accepts wgMLST profiles generated from the wgMLST\_Profiling module.

<sup>1</sup><http://www.genomicepidemiology.org/>

<sup>2</sup><http://www.ncbi.nlm.nih.gov/assembly/>

<sup>3</sup><http://www.ncbi.nlm.nih.gov/bioproject/PRJNA237212/>

<sup>4</sup><http://wgmlst.imst.nsysu.edu.tw>

The maximal number for constructing a dendrogram is set at 999 wgMLST profiles.

## RESULTS

### *S. Enteritidis* Pan-Genome Allele Database

The database constructed with 395 genomes contained 10,704 loci (genes) of which 2,149 loci were shared by  $\geq 95\%$  of the genomes, 3,377 loci by  $\geq 90\%$  of the genomes, and 4,820 loci by  $\geq 5\%$  of the genomes (Figure 1). No genes were shared by all of the genomes, but as many as 2,830 loci (26.4%) existed in only 1 genome. Two loci were most common and shared by 391 (99.0%) genomes. The distribution of number of loci over number of genomes increased and reached a peak at which 225 loci were present in 379 (96.0%) genomes. The number of loci accumulated slowly from 90 to 5% of the genomes, but a large number of loci were present in very few genomes. There were 5,884 (55.0%) loci present in only 1 to 19 ( $\leq 5\%$ ) genomes.

In the present study, core genes were designated to those present in  $\geq 90\%$  of the 395 genomes. Dispensable genes were those present in two or more genomes, but less than 90% of the genomes. Unique genes were those present in one genome. In this database, 31.5% (3,377 loci) belonged to core genes, 42.0% (4,497 loci) dispensable genes, and 26.4% (2,830 loci) unique genes.

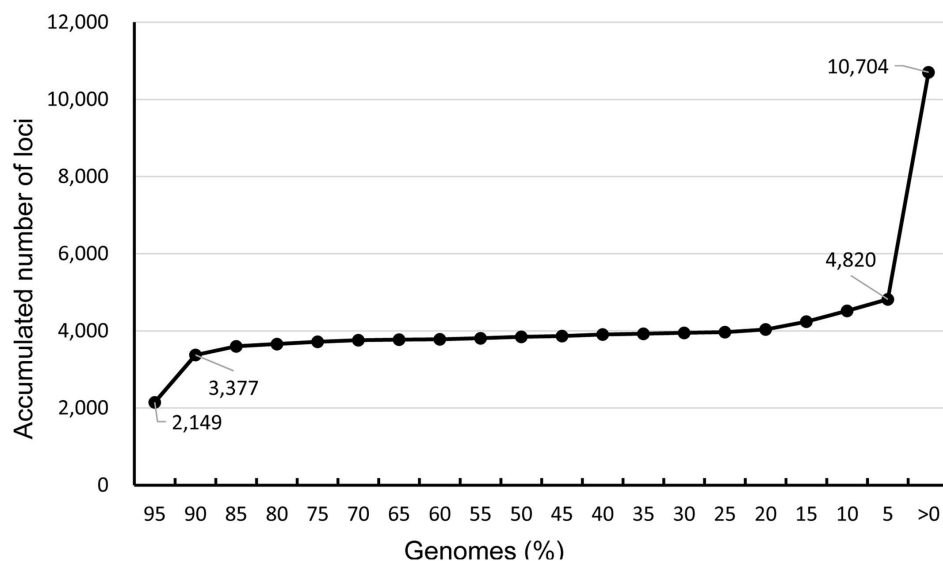
### Performance of wgMLST on Identifying Disease Clusters

The *S. Enteritidis* pan-genome allele database was applied to generate wgMLST profiles for 55 epidemiologically well-defined *S. Enteritidis* isolates from the Minnesota Department of Health. The panel included isolates from seven outbreaks and

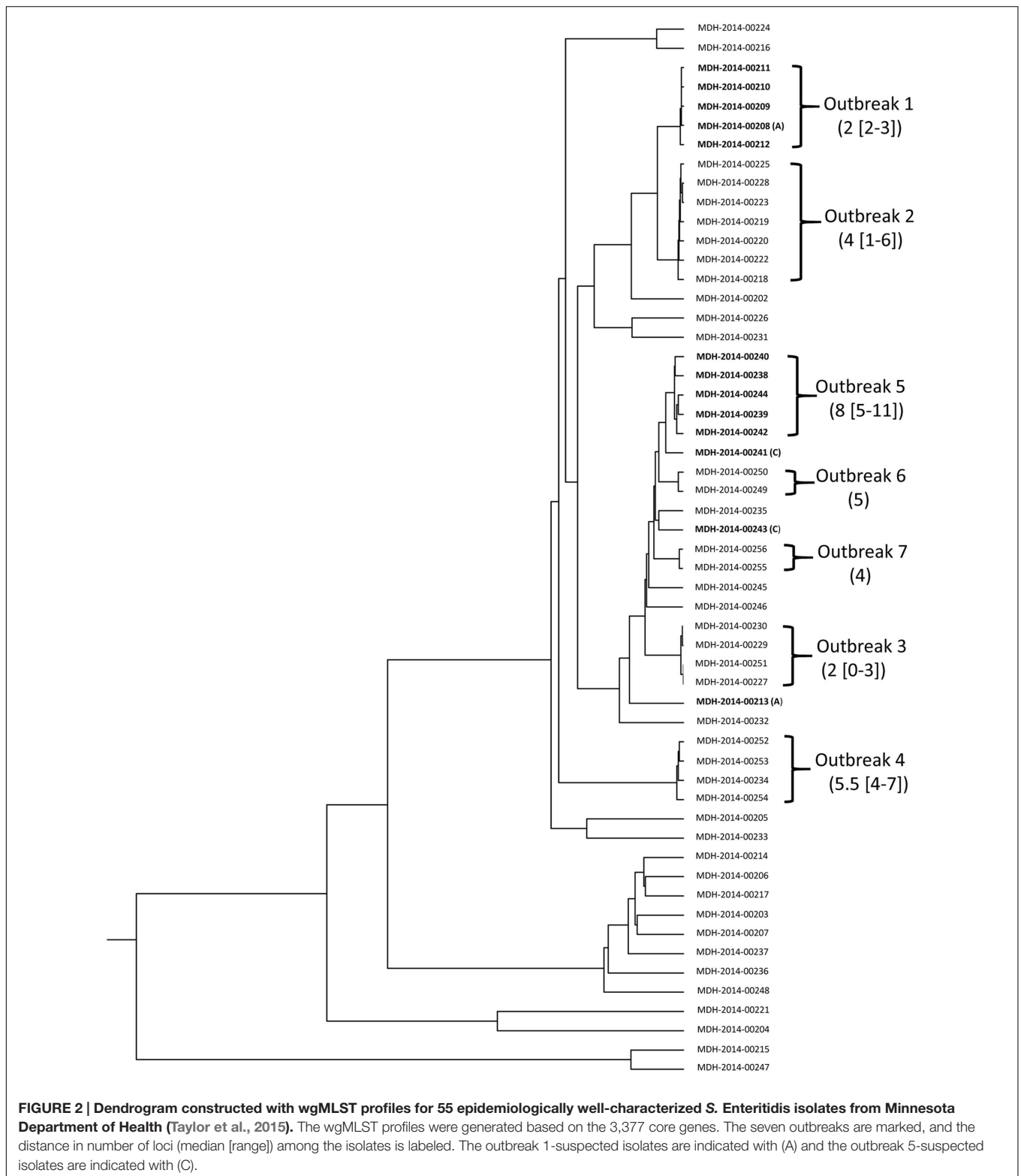
epidemiologically unrelated isolates, and the genetic relatedness among the isolates had been analyzed using the SNP-based approach (Taylor et al., 2015). Cluster analysis of wgMLST profiles generated a genetic relatedness tree which had a dendrogram topology highly similar to that constructed with the SNP profiles (Taylor et al., 2015). The dendrogram revealed distinct clades that were concordant with the seven disease outbreaks (Figure 2). Isolates for each outbreak varied by 0 to 11 loci. Consistent with the SNP-based analysis, wgMLST-based analysis excluded one of the 2 outbreak 1-suspected isolates (MDH-2014-00213) and 2 outbreak 5-suspected isolates (MDH-2014-00241 and MDH-2014-00243) from the related outbreaks (Figure 2). MDH-2014-00213 had a distance of 92–95 loci to the outbreak 1 isolates. The outbreak 5-suspected isolates were genetically close to outbreak 5 isolates; MDH-2014-00241 and MDH-2014-00243 had a distance of 17–19 loci and 23–30 loci to the outbreak 5 isolates, respectively.

## DISCUSSION

We constructed an *S. Enteritidis* pan-genome allele database and developed a wgMLST\_profiling tool to analyze WGS data for disease cluster identification. Using a pan-genome allele database, the wgMLST\_profiling program generates wgMLST profiles that can be used to assess the genetic relatedness between bacterial strains. The evaluation with 55 epidemiologically well-defined isolates from the Minnesota Department of Health indicates that wgMLST has a high resolution in discriminating between isolates for identification of disease clusters. Another advantage is that the wgMLST file is relatively small in size and portable; it can be a promising WGS-based standard subtyping tool for disease surveillance and outbreak investigation across laboratories.



**FIGURE 1 | Distribution of accumulated number of loci over genomes in an *S. Enteritidis* pan-genome database.** The database is constructed with 395 *S. Enteritidis* genomes and contains 10,704 loci (genes) of which 3,377 loci are shared by  $\geq 90\%$  genomes used to make wgMLST profiles.



The wgMLST approach has been used as a subtyping tool in the epidemiological study of various bacterial pathogens (Cody et al., 2013; Bratcher et al., 2014; Kohl et al., 2014; Leopold et al., 2014; Antwerpen et al., 2015; de Been et al.,

2015; Jackson et al., 2016; Lindsey et al., 2016; Raphael et al., 2016). This gene-by-gene comparison method is able to provide high-resolution typing results to allow accurate discrimination between epidemiologically related isolates and unrelated isolates.

wgMLST uses the same principle as the SNP-based approach; it converts SNPs in target genes into a standardized and portable allele numbering system. Our study and others indicate that the performance of wgMLST is equivalent to that of the SNP-based approach (de Been et al., 2015; Raphael et al., 2016). One great advantage is that the wgMLST profile consists of serial digital numbers in numeric order that represent alleles of the target genes. Thus, wgMLST is far less computationally intensive than an SNP-based approach for using WGS data to investigate genetic relatedness among bacterial strains.

The *S. Enteritidis* pan-genome allele database constructed with 395 genomes comprises 10,704 loci of which 31.5% (3,377 loci) exist in  $\geq 90\%$  of the genomes, only 13.5% (1,443 loci) are distributed in 5 to 90% of the genomes, and 55% (5,884 loci) are present in 1 to 19 (5%) of the genomes. Theoretically, the more target genes used for comparison, the higher the resolution in discriminating between bacterial strains. In the present study, we used the 3,377 loci (core genes) to be the target genes for wgMLST profiling. Our evaluation with the panel of epidemiologically well-defined *S. Enteritidis* isolates indicates that wgMLST profiling based on the 3,377 loci is able to provide sufficient resolution in discerning the outbreak isolates from non-outbreak isolates. We also compared the performance of wgMLST by profiling from 2,149 loci (distributed in  $\geq 95\%$  of the genomes), 3,602 loci ( $\geq 85\%$ ), 3,968 loci ( $\geq 25\%$ ), 4,241 loci ( $\geq 15\%$ ), 4,820 loci ( $\geq 5\%$ ), and 10,704 loci on cluster identification. wgMLST profiling based on various numbers of loci performed equivalently in discerning the outbreak isolates from the non-outbreak isolates for the panel of isolates from the Minnesota Department of Health. However, as the number of loci used in profiling increased, we observed a higher number of loci variations among isolates within an outbreak. Whether wgMLST profiling based on more target genes can provide better resolution for identifying outbreak isolates requires further assessment.

In the present study, the pan-genome allele database was constructed with *S. Enteritidis* genomes, and only the core genes were applied for wgMLST profiling. However, it is important to consider whether this *S. Enteritidis* pan-genome allele database can be applied for wgMLST profiling of other *Salmonella* serovars. One would expect that the core genes of *S. Enteritidis* should also exist in other *Salmonella* serovars; therefore, a pan-genome allele database constructed from genomes of a certain serovar such as *S. Enteritidis*, should also be applicable for wgMLST profiling of other serovars. Our preliminary test found that wgMLST profiling using the *S. Enteritidis* pan-genome allele database could discriminate between three *S. Typhimurium* infection outbreaks. Because the *Salmonella* genus consists of a wide variety of serovars ( $> 2,600$  serovars), constructing a universal *Salmonella* pan-genome allele database applicable to all serovars is necessary for public health laboratories.

The *S. Enteritidis* pan-genome allele database was constructed using the PGAdb-builder (Liu et al., 2016). The program grouped proteins that shared  $\geq 95\%$  amino acid sequence identity as a gene cluster (a locus). PGAdb-builder doesn't provide options for users to adjust the sequence coverage, thus alleles of a

locus might have great length differences (i.e., the shorter allele could be only a small portion of the longest one). In this study, the longest allele "nucleotide" sequence of each locus in the database is chosen as reference for wgMLST profiling. In the profiling process, putative allele has to have  $\geq 90\%$  length coverage and  $\geq 90\%$  sequence identity with the reference allele that allows to exclude alleles with variations results from a large deletion or insertion. However, the use of the longest allele of each locus as reference may not be the best choice. Assume the longest allele sequence of a particular locus is 1,000 bp, putative allele sequences less than 900 bp will be excluded. In comparison, when a reference of 800 bp is used, only putative allele sequences less than 720 bp will be excluded. Since *S. Enteritidis* is a very clonal organism, alleles of a gene should have only little variations in length. To choose the longest allele sequence as reference should not be a problem for the clonal *S. Enteritidis* but could exclude most alleles of a gene for a more diversified organism. To choose a sequence with a length that occurs most frequently in the alleles of a gene (i.e., mode) as reference may be more appropriate for a more diversified organism.

## CONCLUSION

An *S. Enteritidis* pan-genome allele database has been constructed, and the core genes can be used as target genes for wgMLST profiling of isolates. wgMLST profiles are small in size and portable and may be readily standardized and compared across laboratories; therefore, wgMLST may be a superior WGS-based subtyping tool for disease surveillance and outbreak investigation. The database and tools used for wgMLST profiling and cluster analysis of wgMLST profiles have been installed on a website<sup>4</sup> for public use.

## AUTHOR CONTRIBUTIONS

Y-YL and C-CC contributed equally to construction of the wgMLST database and the web server and development of tools. C-SC designed the study, analyzed and interpreted the data and drafted the manuscript. All authors approved the final version.

## FUNDING

This study was supported by a grant (MOHW105-CDC-C-315-123301) from Centers for Disease Control, the Ministry of Health and Welfare, Taiwan.

## ACKNOWLEDGMENTS

We are grateful to the Institute of Medical Science and Technology, National Sun Yat-sen University for providing the



hardware and software resources. We also thank the researchers in the Minnesota Department of Health for granting us to use the 55 *S. Enteritidis* genomes for assessing the usefulness of the database in fine typing of bacterial strains for disease outbreak investigation.

## REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Antwerpen, M. H., Prior, K., Mellmann, A., Hoppner, S., Spletstoesser, W. D., and Harmsen, D. (2015). Rapid high resolution genotyping of *Francisella tularensis* by whole genome sequence comparison of annotated genes (“MLST+”). *PLoS ONE* 10:e0123298. doi: 10.1371/journal.pone.0123298
- Bekal, S., Berry, C., Reimer, A. R., Van Domselaar, G., Beaudry, G., Fournier, E., et al. (2016). Usefulness of high-quality core genome single-nucleotide variant analysis for subtyping the highly clonal and the most prevalent *Salmonella enterica* serovar Heidelberg clone in the context of outbreak investigations. *J. Clin. Microbiol.* 54, 289–295. doi: 10.1128/JCM.02200-15
- Boxrud, D., Pederson-Gulrud, K., Wotton, J., Medus, C., Lyszkowicz, E., Besser, J., et al. (2007). Comparison of multiple-locus variable-number tandem repeat analysis, pulsed-field gel electrophoresis, and phage typing for subtype analysis of *Salmonella enterica* serotype Enteritidis. *J. Clin. Microbiol.* 45, 536–543. doi: 10.1128/JCM.01595-06
- Bratcher, H. B., Corton, C., Jolley, K. A., Parkhill, J., and Maiden, M. C. (2014). A gene-by-gene population genomics platform: de novo assembly, annotation and genealogical analysis of 108 representative *Neisseria meningitidis* genomes. *BMC Genomics* 15:1138. doi: 10.1186/1471-2164-15-1138
- Cody, A. J., McCarthy, N. D., Jansen van Rensburg, M., Isinkaye, T., Bentley, S. D., Parkhill, J., et al. (2013). Real-time genomic epidemiological evaluation of human *Campylobacter* isolates by use of whole-genome multilocus sequence typing. *J. Clin. Microbiol.* 51, 2526–2534. doi: 10.1128/JCM.00066-13
- de Been, M., Pinholt, M., Top, J., Bletz, S., Mellmann, A., van Schaik, W., et al. (2015). Core genome multilocus sequence typing scheme for high-resolution typing of *Enterococcus faecium*. *J. Clin. Microbiol.* 53, 3788–3797. doi: 10.1128/JCM.01946-15
- Deng, X., den Bakker, H. C., and Hendriksen, R. S. (2016). Genomic epidemiology: whole-genome-sequencing-powered surveillance and outbreak investigation of foodborne bacterial pathogens. *Annu. Rev. Food Sci. Technol.* 7, 353–374. doi: 10.1146/annurev-food-041715-033259
- Fratamico, P. M., DeRoy, C., Liu, Y., Needleman, D. S., Baranzoni, G. M., and Feng, P. (2016). Advances in molecular serotyping and subtyping of *Escherichia coli*. *Front. Microbiol.* 7:644. doi: 10.3389/fmicb.2016.00644
- Jackson, B. R., Tarr, C., Strain, E., Jackson, K. A., Conrad, A., Carleton, H., et al. (2016). Implementation of nationwide real-time whole-genome sequencing to enhance *Listeria* outbreak detection and investigation. *Clin. Infect. Dis.* 63, 380–386. doi: 10.1093/cid/ciw242
- Jonathan, S. B., Michael, G., and Adam, M. (2016). Whole-genome sequencing detection of ongoing contamination at a restaurant, Rhode Island, USA, 2014. *Emerg. Infect. Dis.* 22, 1474–1476. doi: 10.3201/eid2208.151917
- Kohl, T. A., Diel, R., Harmsen, D., Rothganger, J., Walter, K. M., Merker, M., et al. (2014). Whole-genome-based *Mycobacterium tuberculosis* surveillance: a standardized, portable, and expandable approach. *J. Clin. Microbiol.* 52, 2479–2486. doi: 10.1128/JCM.00567-14
- Leopold, S. R., Goering, R. V., Witten, A., Harmsen, D., and Mellmann, A. (2014). Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. *J. Clin. Microbiol.* 52, 2365–2370. doi: 10.1128/JCM.00262-14
- Liang, S. Y., Watanabe, H., Terajima, J., Li, C. C., Liao, J. C., Tung, S. K., et al. (2007). Multilocus variable-number tandem-repeat analysis for molecular typing of *Shigella sonnei*. *J. Clin. Microbiol.* 45, 3574–3580. doi: 10.1128/JCM.00675-07
- Lindsey, R. L., Pouseele, H., Chen, J. C., Strockbine, N. A., and Carleton, H. A. (2016). Implementation of whole genome sequencing (WGS) for identification and characterization of shiga toxin-producing *Escherichia coli* (STEC) in the United States. *Front. Microbiol.* 7:766. doi: 10.3389/fmicb.2016.00766
- Liu, Y. Y., Chiou, C. S., and Chen, C. C. (2016). PGAdb-builder: a web service tool for creating pan-genome allele database for molecular fine typing. *Sci. Rep.* 6:36213. doi: 10.1038/srep36213
- Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., et al. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* 95, 3140–3145. doi: 10.1073/pnas.95.6.3140
- Maiden, M. C., Jansen van Rensburg, M. J., Bray, J. E., Earle, S. G., Ford, S. A., Jolley, K. A., et al. (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat. Rev. Microbiol.* 11, 728–736. doi: 10.1038/nrmicro3093
- Raphael, B. H., Baker, D. J., Nazarian, E., Lapiere, P., Bopp, D., Kozak-Muiznieks, N. A., et al. (2016). Genomic resolution of outbreak-associated *Legionella pneumophila* serogroup 1 isolates from New York State. *Appl. Environ. Microbiol.* 82, 3582–3590. doi: 10.1128/AEM.00362-16
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Swaminathan, B., Barrett, T. J., Hunter, S. B., and Tauxe, R. V. (2001). PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. *Emerg. Infect. Dis.* 7, 382–389. doi: 10.3201/eid0703.017303
- Taylor, A. J., Lappi, V., Wolfgang, W. J., Lapiere, P., Palumbo, M. J., Medus, C., et al. (2015). Characterization of foodborne outbreaks of *Salmonella enterica* serovar Enteritidis with whole-genome sequencing single nucleotide polymorphism-based analysis for surveillance and outbreak detection. *J. Clin. Microbiol.* 53, 3334–3340. doi: 10.1128/JCM.01280-15

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.02010/full#supplementary-material>

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Liu, Chen and Chiou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.