

Complete Genome Sequence of *emm28* Type *Streptococcus pyogenes* MEW123, a Streptomycin-Resistant Derivative of a Clinical Throat Isolate Suitable for Investigation of Pathogenesis

Kristin M. Jacob,^a Theodore Spilker,^a John J. LiPuma,^a Suzanne R. Dawid,^{a,b}  Michael E. Watson, Jr.^a

Division of Pediatric Infectious Diseases, Department of Pediatrics and Communicable Diseases,^a and Department of Microbiology and Immunology,^b University of Michigan Medical School, Ann Arbor, Michigan, USA

We present here the complete genome sequence of *Streptococcus pyogenes* type *emm28* strain MEW123, a streptomycin-resistant derivative of a pediatric throat isolate. The genome length is 1,878,699 bp, with 38.29% G+C% content. The genome sequence adds value to this virulent *emm28* representative strain and will aid in the investigation of streptococcal pathogenesis.

Received 29 January 2016 Accepted 5 February 2016 Published 17 March 2016

Citation Jacob KM, Spilker T, LiPuma JJ, Dawid SR, Watson ME, Jr. 2016. Complete genome sequence of *emm28* type *Streptococcus pyogenes* MEW123, a streptomycin-resistant derivative of a clinical throat isolate suitable for investigation of pathogenesis. *Genome Announc* 4(2):e00136-16. doi:10.1128/genomeA.00136-16.

Copyright © 2016 Jacob et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Michael E. Watson, Jr., mewats@med.umich.edu.

Streptococcus pyogenes is a Gram-positive bacterial pathogen responsible for a great diversity of human disease manifestations (1). Genotyping by DNA sequencing of the variable region of the *emm* gene distinguishes *S. pyogenes* isolates; the four most prevalent *emm* types causing pharyngitis and invasive disease in North America are types 1, 3, 12, and 28 (2-4). *S. pyogenes* type *emm28* has a particular association with female urogenital infections, including vulvovaginitis, endometritis, and puerperal sepsis (5-8). We previously described the *emm28* strain MEW123 as a streptomycin-resistant derivative of a pediatric throat isolate, which is amenable to genetic manipulation and establishes prolonged carriage in a murine vaginal colonization model (9). We report here the complete MEW123 genome sequence to provide a reference for future studies.

Chromosomal DNA was isolated using the Wizard genomic DNA purification kit (Promega, Madison, WI), and sequenced using the PacBio RS II sequencer (Pacific Biosciences, Menlo Park, CA). Samples were prepared according to the manufacturer's protocols with the P6 polymerase kit and C4 sequencing reagents, with the exception of an increase to 1 h polymerase binding and 1 h binding to magnetic beads. For library construction, DNA was sheared to fragments of ~23,700 bp and isolated using the BluePippin electrophoresis system (Sage Science, Inc., Beverly, MA). Sequences were collected using one single-molecule real-time (SMRT) cell. This generated 70,159 reads, each with a length of ~16,100 bp, for a total of 1,131.6 Mb of sequence data (~300- to 500-fold coverage). The sequence was assembled using Celera version 8.3rc2 (10, 11). The resulting single scaffold was indexed and aligned against the fastq reads with BWA version 0.7.12, using BWA-MEM. The resulting .SAM file was sorted, indexed, and converted to .bam and .bai files using SAMtools version 1.2 (12, 13). Error correction was performed with Pilon version 1.12 (14) and Harvest tools version 1.2, employing parsnp and gngir (15). Genome overlap at the ends was identified with SeqEdit (DNASar, Madison, WI), and trimmed manually to have position *dnaA* as the starting point.

Preliminary annotation was performed using Prokka version 1.11, with a reference library generated from *emm28* strain MGAS6180 (accession no. NC_007296.1) (16). Upon submission to GenBank, the annotations were repeated using the NCBI Prokaryotic Genome Annotation Pipeline for database consistency.

The genome sequence contains 1,878,699 bp, with 38.29% G+C% content. The number of predicted coding regions is 1,827, with 18 rRNA and 67 tRNA genes. The PHAge Search Tool (PHAST) identified 1 intact and 2 incomplete prophage regions (17). The Web-based tool CRISPRFinder identified 3 candidate clustered regularly interspaced short palindromic repeat (CRISPR) regions (18). By multilocus sequence typing (MLST), MEW123 was recognized as sequence type 52 (19). MEW123 *emm28* (cluster E4) has 100% nucleotide identity to *emm28* of MGAS6180 (8). The MEW123 R28 adhesin gene contained 7 tandem nucleotide repeats of 237 bp each, which is shorter than the MGAS6180 R28 sequence containing 13 repeats (8, 20). Sequence examination identified the pyrogenic exotoxins and superantigens SpeB, SpeC, SpeF, SpeG, SpeJ, and SmeZ. Altogether, these sequence data will benefit future investigations of *S. pyogenes* molecular biology and pathogenesis.

Nucleotide sequence accession number. This genome sequence has been deposited in GenBank under the accession no. CP014139. The version described in this paper is the first version.

ACKNOWLEDGMENTS

We thank Robert Lyon, Christina McHenry, Katherine Borysko, and the University of Michigan DNA Sequencing Core Facility for their technical expertise.

This work was supported by the University of Michigan Department of Pediatrics and Communicable Diseases and grant NIH K12 HD028820. We acknowledge the use of the *S. pyogenes* MLST database, which is located at Imperial College London and is funded by the Wellcome Trust.

FUNDING INFORMATION

This work, including the efforts of Michael Edmund Watson, was funded by HHS | NIH | National Institute of Child Health and Human Development (NICHD) (5 K12 HD 028820).

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

- Walker MJ, Barnett TC, McArthur JD, Cole JN, Gillen CM, Henningham A, Sriprakash KS, Sanderson-Smith ML, Nizet V. 2014. Disease manifestations and pathogenic mechanisms of group A streptococcus. *Clin Microbiol Rev* 27:264–301. <http://dx.doi.org/10.1128/CMR.00101-13>.
- Shulman ST, Tanz RR, Dale JB, Beall B, Kabat W, Kabat K, Cederlund E, Patel D, Rippe J, Li Z, Sakota V, North American Streptococcal Pharyngitis Surveillance Group. 2009. Seven-year surveillance of North American pediatric group A streptococcal pharyngitis isolates. *Clin Infect Dis* 49:78–84.
- O'Brien KL, Beall B, Barrett NL, Cieslak PR, Reingold A, Farley MM, Danila R, Zell ER, Facklam R, Schwartz B, Schuchat A. 2002. Epidemiology of invasive group A streptococcus disease in the United States, 1995–1999. *Clin Infect Dis* 35:268–276. <http://dx.doi.org/10.1086/341409>.
- O'Loughlin RE, Roberson A, Cieslak PR, Lynfield R, Gershman K, Craig A, Albanese BA, Farley MM, Barrett NL, Spina NL, Beall B, Harrison LH, Reingold A, Van Beneden C, Active Bacterial Core Surveillance Team. 2007. The epidemiology of invasive group A streptococcal infection and potential vaccine implications: United States, 2000–2004. *Clin Infect Dis* 45:853–862. <http://dx.doi.org/10.1086/521264>.
- Eriksson BK, Norgren M, McGregor K, Spratt BG, Normark BH. 2003. Group A streptococcal infections in Sweden: a comparative study of invasive and noninvasive infections and analysis of dominant T28 *emm*28 isolates. *Clin Infect Dis* 37:1189–1193. <http://dx.doi.org/10.1086/379013>.
- Chuang I, Van Beneden C, Beall B, Schuchat A. 2002. Population-based surveillance for postpartum invasive group A streptococcus infections, 1995–2000. *Clin Infect Dis* 35:665–670. <http://dx.doi.org/10.1086/342062>.
- Green NM, Beres SB, Graviss EA, Allison JE, McGeer AJ, Vuopio-Varkila J, LeFebvre RB, Musser JM. 2005. Genetic diversity among type *emm*28 group A streptococcus strains causing invasive infections and pharyngitis. *J Clin Microbiol* 43:4083–4091. <http://dx.doi.org/10.1128/JCM.43.8.4083-4091.2005>.
- Green NM, Zhang S, Porcella SF, Nagiec MJ, Barbian KD, Beres SB, LeFebvre RB, Musser JM. 2005. Genome sequence of a serotype M28 strain of group A streptococcus: potential new insights into puerperal sepsis and bacterial disease specificity. *J Infect Dis* 192:760–770. <http://dx.doi.org/10.1086/430618>.
- Watson ME, Jr, Nielsen HV, Hultgren SJ, Caparon MG. 2013. Murine vaginal colonization model for investigating asymptomatic mucosal carriage of *Streptococcus pyogenes*. *Infect Immun* 81:1606–1617. <http://dx.doi.org/10.1128/IAI.00021-13>.
- Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KH, Remington KA, Anson EL, Bolanos RA, Chou HH, Jordan CM, Halpern AL, Lonardi S, Beasley EM, Brandon RC, Chen L, Dunn PJ, Lai Z, Liang Y, Nusskern DR, Zhan M, Zhang Q, Zheng X, Rubin GM, Adams MD, Venter JC. 2000. A whole-genome assembly of *Drosophila*. *Science* 287:2196–2204. <http://dx.doi.org/10.1126/science.287.5461.2196>.
- Berlin K, Koren S, Chin CS, Drake JP, Landolin JM, Phillippy AM. 2015. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nat Biotechnol* 33:623–630. <http://dx.doi.org/10.1038/nbt.3238>.
- Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993. <http://dx.doi.org/10.1093/bioinformatics/btr509>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing Subgroup. 2009. The Sequence Alignment of Map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <http://dx.doi.org/10.1371/journal.pone.0112963>.
- Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol* 15:524. <http://dx.doi.org/10.1186/PREACCEPT-2573980311437212>.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <http://dx.doi.org/10.1093/bioinformatics/btu153>.
- Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. *Nucleic Acids Res* 39:W347–W352. <http://dx.doi.org/10.1093/nar/gkr485>.
- Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a Web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35:W52–W57. <http://dx.doi.org/10.1093/nar/gkm360>.
- McGregor KF, Spratt BG, Kalia A, Bennett A, Bilek N, Beall B, Bessen DE. 2004. Multilocus sequence typing of *Streptococcus pyogenes* representing most known *emm* types and distinctions among subpopulation genetic structures. *J Bacteriol* 186:4285–4294. <http://dx.doi.org/10.1128/JB.186.13.4285-4294.2004>.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27:573–580. <http://dx.doi.org/10.1093/nar/27.2.573>.