2015

Complete genome sequences of emm6 Streptococcus pyogenes JRS4 and parental strain D471

Gary C. Port
Washington University School of Medicine in St. Louis

Elyse Paluscio
Washington University School of Medicine in St. Louis

Michael G. Caparon
Washington University School of Medicine in St. Louis

genomeA
Journals.ASM.org

Complete Genome Sequences of *emm6 Streptococcus pyogenes* JRS4 and Parental Strain D471

Gary C. Port, Elyse Paluscio, Michael G. Caparon

Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, Missouri, USA

We report the complete genome assemblies of the group A *Streptococcus pyogenes* serotype *emm*6 strain D471 and its streptomycin-resistant derivative JRS4. Both of these well-studied laboratory strains have been extensively characterized over the past three decades and have been instrumental in the discovery of multiple aspects of streptococcal pathogenesis.

Received 26 May 2015 **Accepted** 1 June 2015 **Published** 2 July 2015

Citation Port GC, Paluscio E, Caparon MG. 2015. Complete genome sequences of *emm*6 *Streptococcus pyogenes* JRS4 and parental strain D471. Genome Announc 3(4):e00725-15. doi:10.1128/genomeA.00725-15.

Address correspondence to Michael G. Caparon, caparon@wusm.wustl.edu.

Streptococcus pyogenes is a Gram-positive bacterial pathogen responsible for a broad range of human diseases (1). The genome of *S. pyogenes* encodes an arsenal of adhesins and toxins that enable this strict human pathogen to infect a wide range of human tissues. Immunity toward *S. pyogenes* is strain specific, as each strain encodes a unique set of surface antigens known as M-protein (2) and T-antigen (3). Advances in streptococcal genetics over the past several decades have facilitated the detailed characterization of numerous virulence factors. Much of the pioneering work in this field has utilized a strain from the Rockefeller University Lancefield collection known as D471, a rheumatic fever-associated M6 isolate, as well as its streptomycin-resistant derivative, JRS4 (4). These studies include the first targeted gene deletion (5), chromosomal complementation (6), and isogenic replacement of different M-protein encoding genes (7). Additionally, the M-protein regulator Mga was first identified (8) and episomally complemented (9) in these strains. Furthermore, the alternative sortases that covalently link T-antigen (pilus) to the cell wall were discovered in these strains (10, 11). Finally, these strains were used to first describe cytolysin-mediated translocation (CMT), whereby the secreted effector SPN is translocated into host cells by the pore-forming cytolysin SLO (12). As these strains were and continue to be heavily investigated, we sought to determine the complete genome sequences of JRS4 and D471 in order to provide a framework for future genetic studies on these classic strains.

Genomic DNA (gDNA) from JRS4 was purified by phenol chloroform extraction (13), and sequenced using a 454-GS FLX sequencer (MOgene LC, St. Louis, MO) by collecting shotgun reads and 8-kb paired-end reads as previously described (14). A total of 211,893 reads (67,091,661 nucleotides) were generated, reaching 37-fold genome coverage depth. Sequences were assembled into 26 contigs using Newbler v2.5.3, and were aligned to the SF370 *S. pyogenes* genome (15), generating a single scaffold that was 97% complete. The remaining gaps (ranging from 0.3 kb to 15 kb, total of 58 kb) were filled in by primer walking (IDT, Coralville, IA) and Sanger sequencing (GENEWIZ, South Plainfield, NJ). To correct sequencing errors, gDNA was resequenced by Illumina HiSeq 2000 (GTAC, Washington University, St. Louis, MO) by collecting 50-bp single-end reads generating a total of 7,763,695 reads (301,814,052 nucleotides) reaching 167-fold genome coverage depth. Illumina data were aligned to the reference JRS4 scaffold sequence using DNASTAR SeqMan NGen 4.0.0 (DNASTAR) to generate a final consensus sequence. gDNA from D471 was purified and sequenced by Illumina HiSeq2000 generating a total of 4,359,256 reads (214,875,135 nucleotides) reaching 119-fold genome coverage depth, and aligned to the reference JRS4 scaffold sequence as described above. The JRS4 and D471 genomes are composed of 1,811,968 bp, with an average G+C content of 38.6%. JRS4 contains 6 single-nucleotide polymorphisms (SNPs) compared to its parent D471 including a nonsynonymous substitution in *rpsL* (N56K), which confers streptomycin resistance. The remaining SNPs are in *cypB* (S233T), *rplS* (S40I), *fabT* (F35L, T51I) (16, 17), and an SNP in a noncoding intergenic region 175 bp upstream of *prfC*.

Nucleotide sequence accession numbers. The complete whole-genome sequences of *S. pyogenes* strains JRS4 and D471 have been deposited at NCBI GenBank under the accession numbers CP011414 and CP011415 with locus tags SpyM6JRS4 and SpyM6D471, respectively.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grants AI046433 and AI064721 from the NIH.

We thank Shaukat Rangwala and William Curtiss at MOgene LC for providing their technical expertise. We also thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for help with genomic analysis. The center is partially supported by NCI Cancer Center Support grant P30 CA91842 to the Siteman Cancer Center and by ICTS/CTSA grant UL1RR024992 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research.

This publication is solely the responsibility of the authors and does not necessarily represent the official view of the NCRR or NIH.

ADDENDUM IN PROOF

During the preparation of this manuscript, a JRS4 genome sequence was deposited in GenBank by the Tokyo Medical and Dental University with the locus tag SPYJRS4 under the accession no. AP012335.1. Remarkably, comparison between the two JRS4 genomes revealed only 44 regions of difference, including 39 indels and 5 SNPs throughout the > 1.8 million nucleotides. The majority of differences (37 indels) reside in homopolymeric stretches of adenines/thymidines (between 4- and 9-nucleotide-long stretches), 23 of which occur in intergenic regions, while the remaining reside in open reading frames (ORFs) and would primarily result in frameshift mutations and early stop codons. Comparison of these ORFs with the 20 *S. pyogenes* genomes available in the KEGG database reveals these mutations to be unique to the SPYJRS4 genome sequence. Although homopolymeric nucleotide stretches are subject to slip-strand mutation (16), these indel differences more likely reflect the use of multiple sequencing platforms utilized in this current study (454 and Illumina data), thereby resulting in the slightly more accurate genome sequence described here. Five SNPs were also identified between genomes, which may be the result of extensive genetic drift between the strains.

REFERENCES

1. **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.
2. **Lancefield RC.** 1962. Current knowledge of type-specific M antigens of group A streptococci. J Immunol **89:**307–313.
3. **Mora M, Bensi G, Capo S, Falugi F, Zingaretti C, Manetti AG, Maggi T, Taddei AR, Grandi G, Telford JL.** 2005. Group A *Streptococcus* produce pilus-like structures containing protective antigens and Lancefield T antigens. Proc Natl Acad Sci USA **102:**15641–15646. http://dx.doi.org/10.1073/pnas.0507808102.
4. **Scott JR, Guenthner PC, Malone LM, Fischetti VA.** 1986. Conversion of an M- group A streptococcus to M+ by transfer of a plasmid containing an M6 gene. J Exp Med **164:**1641–1651. http://dx.doi.org/10.1084/jem.164.5.1641.
5. **Norgren M, Caparon MG, Scott JR.** 1989. A method for allelic replacement that uses the conjugative transposon Tn*916*: deletion of the emm6.1 allele in *Streptococcus pyogenes* JRS4. Infect Immun **57:**3846–3850.
6. **Perez-Casal J, Caparon MG, Scott JR.** 1992. Introduction of the *emm*6 gene into an emm-deleted strain of *Streptococcus pyogenes* restores its ability to resist phagocytosis. Res Microbiol **143:**549–558. http://dx.doi.org/10.1016/0923-2508(92)90112-2.
7. **Berkower C, Ravins M, Moses AE, Hanski E.** 1999. Expression of different group A streptococcal M proteins in an isogenic background demonstrates diversity in adherence to and invasion of eukaryotic cells. Mol Microbiol **31:**1463–1475. http://dx.doi.org/10.1046/j.1365-2958.1999.01289.x.
8. **Caparon MG, Scott JR.** 1987. Identification of a gene that regulates expression of M protein, the major virulence determinant of group A streptococci. Proc Natl Acad Sci USA **84:**8677–8681. http://dx.doi.org/10.1073/pnas.84.23.8677.
9. **Perez-Casal J, Caparon MG, Scott JR.** 1991. Mry, a *trans*-acting positive regulator of the M protein gene of *Streptococcus pyogenes* with similarity to the receptor proteins of two-component regulatory systems. J Bacteriol **173:**2617–2624.
10. **Barnett TC, Patel AR, Scott JR.** 2004. A novel sortase, SrtC2, from *Streptococcus pyogenes* anchors a surface protein containing a QVPTGV motif to the cell wall. J Bacteriol **186:**5865–5875. http://dx.doi.org/10.1128/JB.186.17.5865-5875.2004.
11. **Barnett TC, Scott JR.** 2002. Differential recognition of surface proteins in *Streptococcus pyogenes* by two sortase gene homologs. J Bacteriol **184:**2181–2191. http://dx.doi.org/10.1128/JB.184.8.2181-2191.2002.
12. **Madden JC, Ruiz N, Caparon M.** 2001. Cytolysin-mediated translocation (CMT): a functional equivalent of type III secretion in Gram-positive bacteria. Cell **104:**143–152. http://dx.doi.org/10.1016/S0092-8674(01)00198-2.
13. **Caparon MG, Scott JR.** 1991. Genetic manipulation of pathogenic streptococci. Methods Enzymol **204:**556–586.
14. **Port GC, Paluscio E, Caparon MG.** 2013. Complete genome sequence of *emm* type 14 *Streptococcus pyogenes* strain HSC5. Genome Announc **1**(4): e00612-13. http://dx.doi.org/10.1128/genomeA.00612-13.
15. **Ferretti JJ, McShan WM, Ajdic D, Savic DJ, Savic G, Lyon K, Primeaux C, Sezate S, Suvorov AN, Kenton S, Lai HS, Lin SP, Qian Y, Jia HG, Najar FZ, Ren Q, Zhu H, Song L, White J, Yuan X, Clifton SW, Roe BA, McLaughlin R.** 2001. Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. Proc Natl Acad Sci U S A **98:**4658–4663. http://dx.doi.org/10.1073/pnas.071559398.
16. **Port GC, Vega LA, Nylander AB, Caparon MG.** 2014. *Streptococcus pyogenes* polymyxin B-resistant mutants display enhanced ExPortal integrity. J Bacteriol **196:**2563–2577. http://dx.doi.org/10.1128/JB.01596-14.
17. **Lu YJ, Rock CO.** 2006. Transcriptional regulation of fatty acid biosynthesis in *Streptococcus pneumoniae*. Mol Microbiol **59:**551–566. http://dx.doi.org/10.1111/j.1365-2958.2005.04951.x.