Complete Genome Sequences of Three Invasive Strains of *Streptococcus pyogenes* Subtype *emm*5.23 Isolated in Scotland

Davide Pagnossin,a,b Andrew Smith,c,d William Weir,b Chiara Crestani,a Diane Lindsay,c Roisin Ure,c Katarina Oravcovaa

aInstitute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, United Kingdom
bSchool of Veterinary Medicine, University of Glasgow, Glasgow, United Kingdom
cBacterial Respiratory Infection Service, Scottish Microbiology Reference Laboratory, Glasgow Royal Infirmary, Glasgow, United Kingdom
dCollege of Medical, Veterinary & Life Sciences, Glasgow Dental Hospital & School, University of Glasgow, Glasgow, United Kingdom

ABSTRACT *Streptococcus pyogenes emm*5.23 is uncommon; however, it has recently been involved in a relatively high proportion of cases of invasive disease in Scotland. Here, we report the complete genome sequences of three *emm*5.23 isolates, which may be used as a reference for investigating the virulence and epidemiology of this strain.

*S*treptococcus pyogenes* is a Gram-positive human pathogen that can cause superficial or invasive infections (1). *S. pyogenes* strains are classified as *emm* types and subtypes based on the sequence identity of the first 180 bp of the *emm* gene (https://www.cdc .gov/streplab/groupa-strep/emm-background.html). Here, we report the closed genomes of three invasive isolates of *S. pyogenes emm*5.23 collected in Scotland in 2015, 2018, and 2019. This *emm* subtype has been circulating in Scotland since 2015, causing a relatively high proportion of invasive disease cases. In 2018 and 2019, *emm*5.23 was, respectively, the sixth and the second most common strain involved in *S. pyogenes* invasive disease in Scotland (4.7% and 9.8% of all cases). To our knowledge, *emm*5.23 strains are not commonly isolated worldwide (2) but have been reported in one outbreak of invasive disease (3) and in association with increased mortality in the United Kingdom (4). The three closed and annotated genomes we present here may be utilized as references for mapping-based sequencing studies, which may in the future elucidate why this *emm* subtype is disproportionately responsible for invasive infection in Scotland.

Invasive *S. pyogenes* strains isolated in Scotland are routinely collected and *emm* typed by the Bacterial Respiratory Infection Service of the Scottish Microbiology Reference Laboratory (SMiRL) in Glasgow, United Kingdom. In order to produce complete and closed whole-genome sequences (WGS) for our three isolates, we used a hybrid assembly approach, merging short Illumina reads and long Oxford Nanopore reads. Illumina sequencing was performed at the SMiRL, while Oxford Nanopore Technologies (ONT) (Oxford, UK) MinION sequencing was carried out at the University of Glasgow. Prior to DNA extraction, a single colony of each strain was inoculated in Todd-Hewitt broth (THB) (Thermo Fisher Scientific, Waltham, MA) for overnight growth at 37°C. For Illumina sequencing, bacterial cells were lysed using mutanolysin, lysozyme, and proteinase K. Genomic DNA was extracted with the DNeasy 96 blood and tissue kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Extracted DNA was quantified using a Qubit 3 fluorometer (Thermo Fisher, UK), and sequencing libraries were generated using the NexteraXT DNA sample preparation kit and index kit v2 (Illumina, San Diego, CA). Next-generation paired-end WGS was then performed using Illumina MiSeq technology. Paired-end raw reads in FASTQ format were trimmed using ConDeTri v3.11.1 (5) with default settings. The script FilterPCRdupl (5) was then used to remove redundant read copies that might have emerged in the PCR step. For Oxford Nanopore MinION sequencing, bacterial cells were lysed using lysozyme and proteinase K, and genomic

TABLE 1 Isolation, sequencing data, and genome characteristics of the three *S. pyogenes emm*5.23 isolates

Characteristic[a]	Data for isolate[b]:		
	iGAS376	iGAS391	iGAS426
Yr of isolation	2018	2019	2015
Origin	Edinburgh, Scotland	Glasgow, Scotland	Glasgow, Scotland
Source	Blood	Blood	Blood
No. of Illumina reads (pairs)	731,618	572,775	647,143
No. of sequenced bp, Illumina	190,745,659 forward, 188,292,105 reverse	145,710,885 forward, 148,445,698 reverse	161,008,620 forward, 163,918,397 reverse
Genome coverage, Illumina ($\times$)	100	78	86
No. of ONT reads	28,201	95,719	90,232
ONT read length N_{50} (bp)	14,576	9,633	8,828
No. of sequenced bp, ONT	192,673,926	451,467,050	379,063,977
Genome coverage, ONT ($\times$)	102	238	198
Genome size (bp)	1,897,124	1,897,129	1,897,111
GC content (%)	38.6	38.6	38.6
No. of genes	1,913	1,913	1,913
No. of CDSs	1,824	1,824	1,824
Coding genes	1,775	1,775	1,775
rRNAs (5S, 16S, 23S)	6, 6, 6	6, 6, 6	6, 6, 6
tRNAs	67	67	67
MLST	ST99	ST99	ST99
GenBank accession no.	CP067010	CP067009	CP067008

[a] CDS, coding DNA sequences; MLST, multilocus sequence type.
[b] ST99, sequence type 99.

DNA was extracted using the Wizard genomic DNA purification kit (Promega, Madison, WI, USA). Genomic DNA was quantified with both a Qubit 3 fluorometer and a Nanodrop spectrophotometer, and the MinION input concentration was adjusted to approximately 57 ng/μl. Libraries were prepared using the rapid barcoding kit SQK-RBK004 vRBK_9054_v2_revQ_14Aug2019 (Oxford Nanopore Technologies) following the manufacturer's instructions. The genomic libraries were sequenced using a MinION flow cell v9.4.1 on a MinION sequencer (Oxford Nanopore Technologies). The raw read files were base called and demultiplexed with Guppy v3.6.0 (6). The genomes were *de novo* assembled using Unicycler v0.4.8 (7) in the hybrid assembly mode, merging the Illumina and MinION reads. Unicycler was used to confirm that complete circular genomes had been generated and to rotate each genome so that it commenced with the *dnaA* gene encoded on the forward strand. Annotations were produced using the NCBI Prokaryotic Genome Annotation Pipeline v5.0 (8).

The sequence characteristics of the three *emm*5.23 genomes are presented in Table 1.

Data availability. The raw reads and whole-genome assemblies are available under the NCBI BioProject accession number PRJNA685009, and annotation files are available under the GenBank accession numbers CP067008, CP067009, and CP067010.

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. https://doi.org/10.1128/CMR.00101-13.
2. Steer AC, Law I, Matatolu L, Beall BW, Carapetis JR. 2009. Global emm type distribution of group A streptococci: systematic review and implications for vaccine development. Lancet Infect Dis 9:611–616. https://doi.org/10.1016/S1473-3099(09)70178-1.
3. Gossain S, Chalker V, Kapatai G, Coelho J, Mohamed H, Martin K, Tahir M. 2016. A protracted outbreak of invasive group A streptococcal infection at a UK long-term facility investigated using whole genome sequencing. Open Forum Infect Dis 3(Suppl 1):1411. https://doi.org/10.1093/ofid/ofw172.1114.
4. Blagden S, Watts V, Verlander NQ, Pegorie M. 2020. Invasive group A streptococcal infections in North West England: epidemiology, risk factors and fatal infection. Public Health 186:63–70. https://doi.org/10.1016/j.puhe.2020.06.007.
5. Smeds L, Künstner A. 2011. ConDeTri: a content dependent read trimmer

for Illumina data. PLoS One 6:e26314. https://doi.org/10.1371/journal.pone.0026314.

6. Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol 20:129. https://doi.org/10.1186/s13059-019-1727-y.

7. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.

8. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.