

Complete reference genomes of *Haemophilus ducreyi* isolates ITG3105, CIP76118, and TMA3542, using Oxford Nanopore and Illumina sequencing

Daniel Golparian,¹ Daniel Schröder,¹ Magnus Unemo^{1,2}

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT We report the complete reference genome sequences of three *Haemophilus ducreyi* isolates (ITG3105, CIP76118, and TMA3542), generated using Oxford Nanopore and Illumina sequencing, and hybrid assemblies. Each genome has a closed circular chromosome; ITG3105 and TMA3542 also have plasmids. These reference genomes support genomic, epidemiological, and antimicrobial resistance surveillance of *H. ducreyi*.

KEYWORDS *Haemophilus ducreyi*, chancroid, complete reference genomes

Haemophilus ducreyi causes chancroid, which has mainly disappeared as a sexually transmitted infection in many countries where it was previously endemic (1–4). However, in the recent decade, *H. ducreyi* was also shown to cause cutaneous ulceration in children, especially in yaws-endemic regions (3, 4). The global epidemiology of *H. ducreyi* infections remains unknown because syndromic management is mostly used, and quality-assured, sensitive diagnostic assays are lacking, especially in resource-limited settings (1–6). Due to the recently described high prevalence in, for example, Malawi (7), and the rising importance of genomic and antimicrobial resistance (AMR) surveillance for sexually transmitted and skin pathogens, complete reference genomes are needed for *H. ducreyi*. We report high-quality, complete reference genomes for three *H. ducreyi* isolates (ITG3105, CIP76118, and TMA3542) that will support genomic, epidemiological, and AMR surveillance.

The *H. ducreyi* isolates (Table 1) were grown on chocolate agar supplemented with IsoVitalEx (BD Diagnostics) at 33°C ± 1°C in a 5% CO₂-enriched atmosphere for 48–72 h, and species verified using MALDI-TOF MS (Bruker). Genomic DNA was extracted using the Nanobind HMW DNA extraction kit (PacBio). Illumina DNA Library Prep (Illumina) and Ligation Sequencing Kit V14 SQK-LSK114 (Oxford Nanopore Technologies [ONTs]), without DNA shearing or size selection, were used for short- and long-read library preparations, respectively. Sequencing was performed on a NextSeq 550 platform (2 × 150 bp reads) and an Mk1D with an R10.4 flow cell (ONT). DNA and libraries were quality controlled using Qubit (Thermo Fisher) and TapeStation (Agilent). Basecalling was performed using guppy with default parameters (<https://nanoporetech.com/software/other/guppy>; Table 1). Adapters were trimmed using Q-score trimming (Q30) and Porechop (v0.2.4; <https://github.com/rrwick/Porechop>), and all sequencing reads were further quality controlled and species confirmed (e.g., to exclude contamination) using Kraken 2 (8). Hybrid assembly using paired-end Illumina and long-read ONT reads was performed and circularized using Unicycler v0.4.8 (9), which additionally rotated the single contig to start with *dnaA*. The nullarbor pipeline (<https://github.com/tseemann/nullarbor>) was used for additional quality control and antimicrobial resistance (ResFinder v4.7.2) analysis. Default parameters were used for all software unless otherwise specified. The Prokaryotic Genome Annotation Pipeline v6.1 (10) provided by

Editor Zhenjiang Zech Xu, Nanchang University, Nanchang, Jiangxi, China

Address correspondence to Magnus Unemo, magnus.unemo@regionorebrolan.se.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 12 December 2025

Accepted 29 April 2026

Published 18 May 2026

Copyright © 2026 Golparian 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/).

TABLE 1 Characteristics of *Haemophilus ducreyi* isolates ITG3105, CIP76118, and TMA3542; raw sequencing data; and complete reference genome features, including antimicrobial resistance determinants^a

Features	ITG3105	CIP76118	TMA3542
Origin of the isolate	South Africa	France	Thailand
Year of isolation	1983	1978	1990
Source of the isolate	Chancroid	Chancroid	Chancroid
CCUG	14,233	7,470	27,022
Illumina short reads (N)	3,685,698	3,865,436	3,933,254
Average coverage depth	308×	326×	314×
Average Phred score	>Q30	>Q30	>Q30
Average read length (bp)	144	145	142
ONT long reads (N)	94,199	1,415,965	97,218
Average coverage depth	112×	4,626×	354×
Average Phred score	Q25	Q19	Q25
Average read length (bp)	5,126	4,684	3,100
Read N50	2,399	4,676	6,778
Maximum read length (bp)	116,716	243,471	102,388
Assembled sequences	3	1	4
Chromosome length (bp)	1,725,365	1,721,101	1,750,314
GC content (%)	38.3	38.3	38.0
CDS (N)	1,716	1,701	1,727
rRNA	19	19	19
tRNA	49	47	47
tmRNA	1	1	1
Plasmids	2	0	3
Antibiotic resistance determinants			
aph(3')-Ia	— ^b	—	Yes
aph(3'')-Ib	—	—	Yes
blaTEM-1B	Yes	—	Yes
blaROB-1	—	—	Yes
catS	—	—	Yes
sul2	Yes	—	Yes
tet(32)	—	Yes	Yes
tet(B)	Yes	—	Yes

^aN, number; ONT, Oxford Nanopore Technologies; bp, base pairs; CDS, coding sequence; CCUG, Culture Collection University of Gothenburg.

^b“—” indicates that the specific antimicrobial resistance determinants were not detected in the isolates.

NCBI was used to identify and annotate the coding sequences (CDS) and RNAs in the genomes.

Characteristics of the *H. ducreyi* isolates, sequencing data, and complete reference genomes (i.e., structural completeness: fully circularized, hybrid assemblies without gaps), including AMR determinants, are summarized in Table 1. Briefly, the ITG3105 genome consists of a single circular chromosome of 1,725,365 bp (encoding 1,716 CDS, 19 rRNAs, 49 tRNAs, and 1 tmRNA), and two plasmids: pDG001 (9,154 bp; 8 CDS) and pDG002 (7,584 bp; 9 CDS). The CIP76118 genome has a chromosome of 1,721,101 bp (1,701 CDS, 19 rRNAs, 47 tRNAs, and 1 tmRNA) and no plasmids. The TMA3542 genome has a chromosome of 1,750,314 bp (1,727 CDS, 19 rRNAs, 47 tRNAs, and 1 tmRNA) and three plasmids: pDG001 (9,154 bp; 8 CDS), pDG003 (5,685 bp; 6 CDS), and pDG004 (4,625 bp; 5 CDS). These complete reference genomes (chromosomes and plasmids) provide valuable resources for comparative genomics and molecular epidemiological surveillance of *H. ducreyi*, including monitoring of AMR determinants.

ACKNOWLEDGMENTS

The present work was funded by grants from the Örebro County Council Research Committee and the Foundation for Medical Research at Örebro University Hospital, Sweden.

AUTHOR AFFILIATIONS

¹World Health Organization Collaborating Centre for Gonorrhoea and other Sexually Transmitted Infections, Department of Laboratory Medicine, Microbiology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden

²Institute for Global Health, University College London (UCL), London, United Kingdom

AUTHOR ORCIDs

Magnus Unemo  <http://orcid.org/0009-0001-3445-6968>

FUNDING

Funder	Grant(s)	Author(s)
Örebro County Council Research Committee	2024	Magnus Unemo
Foundation for Medical Research at Örebro University Hospital	2022	Magnus Unemo

DATA AVAILABILITY

The *H. ducreyi* ITG3105 (CCUG 14233), CIP76118 (CCUG 7470), and TMA3542 (CCUG 27022) isolates are available at Culture Collection University of Gothenburg (CCUG; <https://www.ccug.se>). ITG3105 originated from a human chancroid infection contracted in South Africa in 1983, CIP76118 from a chancroid case in France in 1978, and TMA3542 from a chancroid case in Bangkok, Thailand, in 1990.

The genome assemblies and raw reads have been deposited under the following accession numbers: [SAMN53162900](https://samn53162900) for ITG3105 chromosome, and plasmids pDG001 and pDG002; [SAMN53162901](https://samn53162901) for CIP76118 chromosome; [SAMN53162902](https://samn53162902) for TMA3542 chromosome and plasmids pDG001, pDG003, and pDG004. Short Read Archive: Illumina reads, ITG3105 ([SRR35993037](https://srr35993037)), CIP76118 ([SRR35993039](https://srr35993039)), and TMA3542 ([SRR35993041](https://srr35993041)); Nanopore reads, ITG3105 ([SRR35993036](https://srr35993036)), CIP76118 ([SRR35993038](https://srr35993038)), and TMA3542 ([SRR35993040](https://srr35993040)). The project summary is found under BioProject accession number [PRJNA1359372](https://prjna1359372).

ETHICS APPROVAL

The *H. ducreyi* isolates examined in this study were obtained from a public culture collection (Culture Collection University of Gothenburg, CCUG), and no personal identification data were available. Accordingly, no ethical approval was required.

REFERENCES

- Morse SA. 1989. Chancroid and *Haemophilus ducreyi*. Clin Microbiol Rev 2:137–157. <https://doi.org/10.1128/CMR.2.2.137>
- González-Beiras C, Marks M, Chen CY, Roberts S, Mitjà O. 2016. Epidemiology of *Haemophilus ducreyi* infections. Emerg Infect Dis 22:1–8. <https://doi.org/10.3201/eid2202.C12202>
- Lewis DA. 2014. Epidemiology, clinical features, diagnosis and treatment of *Haemophilus ducreyi* - a disappearing pathogen? Expert Rev Anti Infect Ther 12:687–696. <https://doi.org/10.1586/14787210.2014.892414>
- Mitjà O, Lukehart SA, Pokowas G, Moses P, Kapa A, Godornes C, Robson J, Cherian S, Houine W, Kazadi W, Siba P, de Lazzari E, Bassat Q. 2014. *Haemophilus ducreyi* as a cause of skin ulcers in children from a yaws-endemic area of Papua New Guinea: a prospective cohort study. Lancet Glob Health 2:e235–41. [https://doi.org/10.1016/S2214-109X\(14\)70019-1](https://doi.org/10.1016/S2214-109X(14)70019-1)
- Handsfield HH, Totten PA, Fennel CL, Falkow S, Holmes KK. 1981. Molecular epidemiology of *Haemophilus ducreyi* infections. Ann Intern Med 95:315–318. <https://doi.org/10.7326/0003-4819-95-3-315>
- Al-Tawfiq JA, Spinola SM. 2024. Infections caused by *Haemophilus ducreyi*: one organism, two stories. Clin Microbiol Rev 37:e0013524. <https://doi.org/10.1128/cmr.00135-24>
- Matoga MM, Chen JS, Seña AC, Ndalama B, Nyirenda N, Bonongwe N, Mathiya E, Kamtamba B, Jere E, Banda GK, Thengolose I, Coelho J, Hennelly CM, Schröder D, Unemo M, Salazar JC, Radolf JD, Parr JB, Hoffman IF. 2026. High prevalence of *Haemophilus ducreyi* among patients with suspected primary syphilis in Malawi, 2019–2022. Clin Infect Dis 82:e604–e611. <https://doi.org/10.1093/cid/ciaf114>

8. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol* 20:257. <https://doi.org/10.1186/s13059-019-1891-0>
9. 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>
10. 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>