



# Whole-Genome Sequencing and *De Novo* Assembly of 67 *Staphylococcus pseudintermedius* Strains Isolated from the Skin of Healthy Dogs

Norma Fàbregas,<sup>a</sup> Daniel Pérez,<sup>b</sup> Joaquim Viñes,<sup>a</sup> Rocio Fonticoba,<sup>b</sup> Anna Cuscó,<sup>a</sup> Lourdes Migura-García,<sup>c</sup> Lluís Ferrer,<sup>b</sup> Olga Francino<sup>d</sup>

<sup>a</sup>Vetgenomics, Edifici EUREKA, PRUAB, Campus UAB, Barcelona, Spain

<sup>b</sup>Department of Animal Medicine and Surgery, Universitat Autònoma de Barcelona, Barcelona, Spain

<sup>c</sup>IRTA, Centre de Recerca en Sanitat Animal (CReSA, IRTA-UAB), Barcelona, Spain

<sup>d</sup>SVGM, Molecular Genetics Veterinary Service, Universitat Autònoma de Barcelona, Barcelona, Spain

Norma Fàbregas and Daniel Pérez contributed equally to this work. Author order was determined by seniority.

**ABSTRACT** We have *de novo* assembled 67 *Staphylococcus pseudintermedius* genomes, with median values of 2.6 Mbp size and 99.43% completeness, 2,386 coding sequences, 19 complete rRNAs, 59 tRNAs, and 4 noncoding RNAs. We released 51 single-contig complete genomes and 16 genomes with a circular main contig using Nanopore sequencing.

**S**taphylococcus pseudintermedius is a commensal of the skin of dogs and the main causative agent of canine pyoderma (1), but populations inhabiting the skin of healthy dogs are largely unknown. The current paradigm indicates that infection arises when the skin barrier is altered by predisposing factors (2). We previously demonstrated that Nanopore sequencing allows *de novo* assembly of the entire genome of *S. pseudintermedius* (3). Here, we retrieved 67 *S. pseudintermedius* genomes from isolates from the skin of healthy dogs using long-read Nanopore sequencing.

Samples were obtained by rubbing sterile swabs on four skin sites from nine healthy dogs—perinasal, perioral, inguinal, and perianal. After culture in blood agar at 37°C for 24 h, colonies grown with the morphology of *S. pseudintermedius* (small silver colonies) were seeded in 3 mL of brain heart infusion (BHI) at 37°C for 16 h.

DNA was extracted with a ZymoBIOMICS DNA miniprep kit (Zymo Research). DNA quality and quantity were determined using a NanoDrop 2000 instrument and a Qubit double-stranded DNA (dsDNA) broad range (BR) assay kit (Fisher Scientific). The sequencing libraries were prepared with the rapid barcoding sequencing kit (SQK-RBK004; Oxford Nanopore Technologies [ONT]). Up to 12 barcoded samples were loaded in a MinION FLO-MIN106 9.4.1 flow cell and sequenced in a MinION Mk1B or Mk1C device (ONT). The fast5 files were base-called and demultiplexed, and adapters were trimmed with Guppy 5.0.11 (4) (ONT) (--dna\_r9.4.1\_450bps\_sup.cfg) (--config configuration.cfg --barcode\_kits SQK-RBK004 --trim\_barcodes; min\_score threshold default 60). Reads with a quality score lower than 10 were discarded. Run summary statistics were obtained with Nanoplot 1.38.1 (5) (--N50 --fastq).

Isolates were confirmed as *S. pseudintermedius* by EPI2ME WIMP workflow (6). Genomes were *de novo* assembled using Flye 2.8.3 (7) (-nano-raw -plasmids -trestle), except HSP279 and HSP281, which were assembled with Flye 2.9 (-nano-hq). Contigs were polished with medaka 1.4.3 (8) (medaka\_consensus; -m r941\_min\_sup\_g507). Genome completeness and contamination were assessed with CheckM 1.1.3 (lineage\_wf) (9). Circlator 1.5.5 was used to rotate the genomes, fixing the start with the  *dnaA* gene (10) (fixstart -min\_id 70). Genomes

**Editor** Steven R. Gill, University of Rochester School of Medicine and Dentistry

**Copyright** © 2022 Fàbregas et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Olga Francino, olga.francino@uab.cat, or Norma Fàbregas, norma.fabregas@vetgenomics.com.

The authors declare a conflict of interest.

**Received** 15 January 2022

**Accepted** 15 February 2022

**Published** 2 March 2022

**TABLE 1** Characteristics and accession numbers of high-quality genome assemblies from 67 *Staphylococcus pseudintermedius* isolates from 4 different skin sites of 9 healthy dogs<sup>a</sup>

ID	MLST	Source	Yr of isolation	Country of isolation	Assembly level	Genome accession no.	Genome assembly no.	No. of contigs	N <sub>50</sub> contigs (bp)	Genome size (bp)	No. of contigs (total)	CDS (X)	Comp (%)	Cont (%)	GC content (%)
HSP147	45 <sup>b</sup>	Cf, I, H	2020	Spain	Comp	GCA_019997105.1	CP083230	1	2,596,738	2,379	199	97.96	0.57	37.6	
HSP149	45 <sup>b</sup>	Cf, I, H	2020	Spain	Comp	GCA_019997905.1	CP083229	1	2,592,238	2,384	90	97.82	0.57	37.6	
HSP151	45 <sup>b</sup>	Cf, Pa, H	2020	Spain	Comp	GCA_019997185.1	CP083231	1	2,596,528	2,380	134	97.96	0.57	37.6	
HSP152	45 <sup>b</sup>	Cf, Pa, H	2020	Spain	Comp	GCA_019997145.1	CP083228	1	2,596,499	2,386	78	96.92	0.57	37.6	
HSP153	402 <sup>b</sup>	Cf, Pa, H	2020	Spain	Comp	GCA_019997125.1	CP083227	1	2,596,563	2,385	113	97.68	0.57	37.6	
HSP154	45 <sup>b</sup>	Cf, Pa, H	2020	Spain	Comp	GCA_019997265.1	CP083226	1	2,596,624	2,376	53	97.96	0.57	37.6	
HSP155	45 <sup>b</sup>	Cf, Pa, H	2020	Spain	Comp	GCA_019998705.1	CP083225	1	2,596,628	2,371	158	98.77	0.57	37.6	
HSP156	45 <sup>b</sup>	Cf, P, H	2020	Spain	Comp	GCA_019997565.1	CP083224	1	2,596,695	2,374	132	98.11	0.57	37.6	
HSP157	45	Cf, P, H	2020	Spain	Comp	GCA_019998065.1	CP083223	1	2,596,646	2,381	151	98.25	0.57	37.6	
HSP158	45 <sup>b</sup>	Cf, P, H	2020	Spain	Comp	GCA_019997985.1	CP083222	1	2,596,630	2,391	120	97.85	0.57	37.6	
HSP159	45 <sup>b</sup>	Cf, P, H	2020	Spain	Comp	GCA_019998395.1	CP083221	1	2,596,316	2,377	90	97.66	0.57	37.6	
HSP160	45 <sup>b</sup>	Cf, P, H	2020	Spain	Comp	GCA_019998665.1	CP083220	1	2,596,481	2,388	38	97.57	0.57	37.6	
HSP166	45 <sup>b</sup>	Cf, Ph, H	2020	Spain	Comp	GCA_019997225.1	CP083219	1	2,596,789	2,375	148	97.82	0.57	37.6	
HSP199	2189	Cf, Pa, H	2020	Spain	Comp	GCA_019998005.1	CP083218	1	2,574,155	2,380	33	99.29	0.00	37.6	
HSP204	2190	Cf, Pa, H	2020	Spain	Comp	GCA_019997445.1	CP083217	1	2,470,602	2,260	33	98.30	0.00	37.8	
HSP210	2191	Cf, P, H	2020	Spain	Comp	GCA_019998825.1	CP083216	1	2,472,661	2,249	191	99.43	0.00	37.8	
HSP211	2175	Cf, P, H	2020	Spain	Comp	GCA_019998805.1	CP083215	1	2,472,653	2,249	136	99.43	0.00	37.8	
HSP212	2175	Cf, P, H	2020	Spain	Comp	GCA_019998765.1	CP083214	1	2,472,754	2,257	158	99.43	0.00	37.8	
HSP213	2175	Cf, P, H	2020	Spain	Comp	GCA_019998525.1	CP083213	1	2,472,662	2,247	398	99.43	0.00	37.8	
HSP214	2192	Cf, Pn, H	2020	Spain	Comp	GCA_019998185.1	CP083212	1	2,524,384	2,319	119	99.43	0.00	37.8	
HSP216	2176	Cf, Ph, H	2020	Spain	Comp	GCA_019998745.1	CP083211	1	2,611,897	2,403	97	99.43	0.57	37.6	
HSP224	2177	Cf, P, H	2020	Spain	Comp	GCA_019998685.1	CP083210	1	2,711,397	2,711,397	2,501	88	99.43	0.57	37.6
HSP225	2177	Cf, P, H	2020	Spain	Comp	GCA_019998785.1	CP083209	1	2,752,898	2,752,898	2,571	289	99.43	0.57	37.6
HSP226	2177	Cf, P, H	2020	Spain	Comp	GCA_019997165.1	CP083208	1	2,750,631	2,567	227	99.43	0.57	37.6	
HSP227	2177	Cf, P, H	2020	Spain	Comp	GCA_019996845.1	CP083207	1	2,750,563	2,568	212	99.43	0.57	37.6	
HSP228	2177	Cf, P, H	2020	Spain	Comp	GCA_019997225.1	CP083206	1	2,750,560	2,568	195	99.43	0.57	37.6	
HSP232	2178	Cf, Pa, H	2020	Spain	Contig	JAIEKF000000000	CP083205	2	2,598,204	2,607,192	2,403	220	99.43	0.00	37.6
HSP235	551	Cf, Pn, H	2020	Spain	Contig	JAIEKT000000000	CP083204	2	2,870,862	2,873,604	2,791	312	99.43	1.14	37.2
HSP236	551	Cf, Pn, H	2020	Spain	Contig	JAIEKS000000000	CP083203	2	2,876,047	2,878,792	2,790	126	99.43	1.14	37.3
HSP237	551	Cf, Ph, H	2020	Spain	Contig	JAIEKR000000000	CP083202	2	2,829,605	2,835,087	2,727	157	99.43	1.14	37.3
HSP238	45 <sup>b</sup>	Cf, Ph, H	2020	Spain	Contig	GCA_02068525.1	CP083201	2	2,646,644	2,453	272	96.88	0.57	37.5	
HSP239	45 <sup>b</sup>	Cf, Ph, H	2020	Spain	Contig	GCA_019997805.1	CP083200	1	2,647,982	2,450	265	98.06	0.57	37.5	
HSP240	551	Cf, P, H	2020	Spain	Contig	GCA_020685905.1	CP083203	1	2,870,858	2,870,858	2,760	169	99.43	1.14	37.2
HSP241	551	Cf, P, H	2020	Spain	Contig	GCA_020685685.1	CP083202	2	2,870,853	2,873,594	2,784	181	99.43	1.14	37.2
HSP242	551	Cf, P, H	2020	Spain	Contig	GCA_019997285.1	CP083201	2	2,870,853	2,876,339	2,794	182	99.43	1.14	37.2
HSP243	551	Cf, P, H	2020	Spain	Contig	GCA_019998365.1	CP083200	2	2,870,749	2,876,235	2,779	266	99.43	1.14	37.2
HSP244	551	Cf, P, H	2020	Spain	Contig	GCA_020685805.1	CP083201	2	2,870,863	2,876,349	2,805	204	99.43	1.14	37.2
HSP245	2178	Cf, Pa, H	2020	Spain	Contig	JAIEKM000000000	CP083200	3	2,598,196	2,609,478	2,399	198	99.43	0.00	37.6
HSP246	2178	Cf, Pa, H	2020	Spain	Contig	JAIEKL000000000	CP083199	2	2,597,938	2,600,981	2,397	137	99.43	0.00	37.6
HSP249	2178	Cf, Pa, H	2020	Spain	Contig	JAIEKR000000000	CP083198	2	2,597,934	2,600,977	2,396	148	99.43	0.00	37.6
HSP250	2178	Cf, I, H	2020	Spain	Contig	JAIEKU000000000	CP083197	2	2,598,182	2,601,232	2,405	175	99.43	0.00	37.6
HSP251	2178	Cf, I, H	2020	Spain	Contig	GCA_020685785.1	CP083196	2	2,598,180	2,602,649	2,400	158	99.43	0.00	37.6
HSP252	2178	Cf, I, H	2020	Spain	Contig	GCA_019997885.1	CP083195	1	2,598,184	2,380	176	99.43	0.00	37.6	
HSP253	2178	Cf, I, H	2020	Spain	Contig	GCA_020685765.1	CP083194	2	2,598,192	2,604,278	2,398	206	99.43	0.00	37.6
HSP255	1026	Cf, Ph, H	2020	Spain	Contig	GCA_019997765.1	CP083201	1	2,599,010	2,599,010	2,377	131	98.86	0.00	37.6

(Continued on next page)

**TABLE 1** (Continued)

ID	MLST	Source	Yr of isolation	Country of isolation	Assembly ID (accession no.)	Assembly level	Genome accession no.	Genome assembly	No. of contigs	N <sub>50</sub> contigs (bp)	Genome size (bp)	No. of all contigs	CDS (total)	Cov (%)	Comp (%)	Cont (%)	GC content (%)
HSP258	1026	Cf, P, H	2020	Spain	GCA_019997205.1	Compl	CP083200	Circular	1	2,599,012	2,370	120	98.30	0.00	37.6		
HSP259	1026	Cf, P, H	2020	Spain	GCA_019997005.1	Compl	CP083199	Circular	1	2,599,031	2,383	155	98.86	0.00	37.6		
HSP260	1026	Cf, P, H	2020	Spain	GCA_019997245.1	Compl	CP083198	Circular	1	2,599,024	2,375	71	99.24	0.00	37.6		
HSP261	1026	Cf, P, H	2020	Spain	GCA_019997785.1	Compl	CP083197	Circular	1	2,599,041	2,372	145	99.43	0.00	37.6		
HSP262	1026	Cf, P, H	2020	Spain	GCA_019997825.1	Compl	CP083196	Circular	1	2,599,036	2,382	147	99.43	0.00	37.6		
HSP263	1026	Cf, I, H	2020	Spain	GCA_019997065.1	Compl	CP083195	Circular	1	2,599,019	2,371	125	99.43	0.00	37.6		
HSP264	1026	Cf, I, H	2020	Spain	GCA_019997705.1	Compl	CP083194	Circular	1	2,599,017	2,384	155	98.86	0.00	37.6		
HSP265	1026	Cf, I, H	2020	Spain	GCA_019997085.1	Compl	CP083193	Circular	1	2,599,016	2,369	109	99.43	0.00	37.6		
HSP266	1026	Cf, I, H	2020	Spain	GCA_019997845.1	Compl	CP083192	Circular	1	2,599,010	2,371	204	99.43	0.00	37.6		
HSP267	1026	Cf, I, H	2020	Spain	GCA_019997625.1	Compl	CP083191	Circular	1	2,598,988	2,374	91	99.43	0.00	37.6		
HSP274	2179	Cf, P, H	2020	Spain	GCA_019998045.1	Compl	CP083190	Circular	1	2,587,513	2,394	230	97.82	0.00	37.6		
HSP276	2179	Cf, P, H	2020	Spain	GCA_019998025.1	Compl	CP083189	Circular	1	2,587,526	2,386	219	98.67	0.00	37.6		
HSP277	2179	Cf, P, H	2020	Spain	GCA_019997945.1	Compl	CP083188	Circular	1	2,587,535	2,391	316	98.30	0.00	37.6		
HSP278	2179	Cf, Pa, H	2020	Spain	GCA_019997225.1	Compl	CP083187	Circular	1	2,587,540	2,386	196	98.30	0.00	37.6		
HSP279	Unknown <sup>b</sup>	Cf, Pa, H	2020	Spain	GCA_020693965.1	Compl	CP085724	Circular <sup>c</sup>	1	2,575,504	2,436	139	98.56	0.00	37.8		
HSP280	Unknown <sup>b</sup>	Cf, Pa, H	2020	Spain	GCA_019997045.1	Compl	CP083186	Circular <sup>c</sup>	1	2,575,827	2,374	137	99.41	0.00	37.8		
HSP281	Unknown <sup>b</sup>	Cf, Pa, H	2020	Spain	GCA_020693985.1	Compl	CP085723	Circular	1	2,575,390	2,441	274	99.13	0.00	37.8		
HSP282	Unknown <sup>b</sup>	Cf, Pa, H	2020	Spain	GCA_019998285.1	Compl	CP083185	Circular <sup>c</sup>	1	2,575,851	2,381	170	99.15	0.00	37.8		
HSP283	294	Cf, I, H	2020	Spain	GCA_019997965.1	Compl	CP083184	Circular	1	2,607,096	2,399	154	99.43	0.00	37.6		
HSP284	2180	Cf, I, H	2020	Spain	GCA_019996945.1	Compl	CP083183	Circular	1	2,568,734	2,361	113	99.43	1.14	37.6		
HSP285	2181	Cf, Pn, H	2020	Spain	GCA_020685605.1	Contig	JAEKF000000000	Circular	2	2,642,410	2,467	127	99.43	0.00	37.5		
HSP286	2181	Cf, Pn, H	2020	Spain	GCA_020686685.1	Contig	JAEKF000000000	Circular	4	2,642,415	2,463	144	99.43	0.00	37.5		

<sup>a</sup>ID, identifier; MLST, multilocus sequence type; Cf, *Canis lupus familiaris*; Pn, pernasal; I, inguinal; Pa, perianal; H, healthy; Compl, complete genome; CDS (total), total number of coding sequences; Cov, coverage; Comp, completeness; Cont, contamination.

<sup>b</sup> MLST alleles with mutations that have not been reported.

<sup>c</sup> Genomes where the *dnaA* gene is not detected start with a random ATG gene.

were annotated with the NCBI Prokaryotic Genome Annotation Pipeline 5.2 and 5.3 (11). Multilocus sequence types (MLSTs) were assigned with PubMLST (<https://pubmlst.org/>; accessed November 2021) (12).

Nanopore sequencing allowed successful *de novo* assembly and polishing of 67 *S. pseudintermedius* genomes (Table 1). The average read  $N_{50}$  value was 5,270.77 bp with 133,226.86 reads per sample. The median values were 154 $\times$  coverage (33 to 398 $\times$ ), 99.43% completeness (96.9 to 99.4%) as in previous hybrid assemblies (13), 2.6 Mbp for contig  $N_{50}$  value and genome size (2.5 to 2.9 Mbp), 37.6% GC content (37.2 to 37.8%), 2,386 coding DNA sequences (CDS) (2,247 to 2,805), 19 complete rRNAs (6 to 8 5S, 5 to 7 26S, and 5 to 8 23S), 59 tRNAs (58 to 61), and 4 noncoding RNAs (ncRNAs).

A total of 51 bacterial isolates were assembled into single-circular-contig complete genomes; 16 isolates were assembled into a main circular contig corresponding to the bacterial chromosome and smaller contigs containing replicon proteins supporting their plasmid origin. Together with our previous study (3), we have announced 95 *S. pseudintermedius* genomes from the skin of healthy dogs, contributing to a better understanding of *S. pseudintermedius* pathogenesis.

**Data availability.** The standardized strain descriptions and accession numbers are presented in Table 1; the genome assemblies, genomic data, and raw data are publicly available in GenBank under BioProject [PRJNA685966](#) and genome accession numbers [CP083183](#) to [CP083231](#), [CP085723](#), [CP085724](#), and [JAJEKF0000000000](#) to [JAJEKU0000000000](#).

## ACKNOWLEDGMENTS

Funding sources were as follows: Research Project RTI2018-101991-B-I00 ("From Whole Genome Sequencing to Clinical Metagenomics: Investigations on the Pathogenesis of *Staphylococcus pseudintermedius* Pyoderma in the Dog"), Spanish Ministry of Science, Innovation and Universities; Torres Quevedo Project, PTQ2018-009961, cofinanced by the European Social Fund and the Spanish Ministry of Science and Innovation; and Industrial Doctorate Program grant 2017DI037 AGAUR, Generalitat de Catalunya, Spain.

L.F. has received unrelated honoraria for lecturing from Zoetis, Bayer, LETI, and Affinity Petcare. O.F. and A.C. have received unrelated honoraria for lecturing from Oxford Nanopore Technologies.

## REFERENCES

- Bannoehr J, Guardabassi L. 2012. *Staphylococcus pseudintermedius* in the dog: taxonomy, diagnostics, ecology, epidemiology and pathogenicity. *Vet Dermatol* 23:253–266. <https://doi.org/10.1111/j.1365-3164.2012.01046.x>.
- Pinchbeck LR, Cole LK, Hillier A, Kowalski JJ, Rajala-Schultz PJ, Bannerman TL, York S. 2006. Genotypic relatedness of staphylococcal strains isolated from pustules and carriage sites in dogs with superficial bacterial folliculitis. *Am J Vet Res* 67:1337–1346. <https://doi.org/10.2460/ajvr.67.8.1337>.
- Francino O, Pérez D, Viñes J, Fontcoba R, Madroñero S, Meroni G, Martino P, Martínez S, Cusco A, Fàbregas N, Migura-García L, Ferrer L. 2021. Whole-genome sequencing and *de novo* assembly of 61 *Staphylococcus pseudintermedius* isolates from healthy dogs and dogs with pyoderma. *Microbiol Resour Announc* 10:16. <https://doi.org/10.1128/MRA.00152-21>.
- 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>.
- De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34:2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>.
- Juul S, Izquierdo F, Hurst A, Dai X, Wright A, Kulesha E, Pettett R, Turner DJ. 2015. What's in my pot? Real-time species identification on the Mi-ION. *bioRxiv* <https://doi.org/10.1101/030742>.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–549. <https://doi.org/10.1038/s41587-019-0072-8>.
- Oxford Nanopore Technologies. 2018. medaka: sequence correction provided by ONT Research. <https://github.com/nanoporetech/medaka>.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
- Hunt M, De Silva N, Otto TD, Parkhill J, Keane J, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 16:294. <https://doi.org/10.1186/s13059-015-0849-0>.
- 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>.
- Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 3:124. <https://doi.org/10.12688/wellcomeopenres.14826.1>.
- Viñes J, Cuscó A, Francino O. 2020. Hybrid assembly from a pathogenic methicillin- and multidrug-resistant *Staphylococcus pseudintermedius* strain isolated from a case of canine otitis in Spain. *Microbiol Resour Announc* 9:e01121-19. <https://doi.org/10.1128/MRA.01121-19>.