Veterinary Dermatology

Vet Dermatol 2021; **32**: 654–663 DOI: 10.1111/vde.13040

Whole genome sequencing and *de novo* assembly of *Staphylococcus pseudintermedius*: a pangenome approach to unravelling pathogenesis of canine pyoderma

Lluís Ferrer*, Rocío García-Fonticoba*, Daniel Pérez†, Joaquim Viñes†‡, Norma Fàbregas‡, Sergi Madroñero†, Gabriele Meroni§, Piera A. Martino§, Sofía Martínez¶, M. Laura Maté, Sergio Sánchez-Bruni**, Anna Cuscó‡, Lourdes Migura-García†† and Olga Francino†**

*Departament de Medicina i Cirurgia Animals, Universitat Autònoma de Barcelona, Barcelona, Spain
†Servei Veterinari de Genètica Molecular, Universitat Autònoma de Barcelona, Barcelona, Spain
‡Vetgenomics, PRUAB, Bellaterra, Barcelona, Spain
§Department of Biomedical, Surgical and Dental Sciences – One Health Unit, Via Pascal 36, Milano, 20133, Italy
¶Hospital Escuela de Pequeños Animales (HEPA), Centro de Investigación Veterinaria de Tandil (CIVETAN), UNCPBA-CONICET-CICPBA, Grupo de Medicina Veterinaria Traslacional (MEVET), Facultad de Ciencias Veterinarias, Universidad Nacional del Centro de la Provincia de Buenos Aires (FCV-UNCPBA), Tandil, Argentina
**Laboratorio de Farmacología, Centro de Investigación Veterinaria de Tandil (CIVETAN), UNCPBA-CONICET-CICPBA, Facultad de Ciencias Veterinarias, Universidad Nacional del Centro (FCV-UNCPBA), Tandil, Argentina
††Centre de Recerca en Sanitat Animal, IRTA, Bellaterra, Barcelona, Spain

Correspondence: Lluís Ferrer, Department de Medicina I Cirurgia Animals, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain.
E-mail: lluis.ferrer@uab.cat

Background – *Staphylococcus pseudintermedius* is the main aetiological agent of canine pyoderma. Whole genome sequencing is the most comprehensive way of obtaining relevant genomic information about microorganisms.

Hypothesis/Objectives – Oxford Nanopore technology enables quality sequencing and *de novo* assembly of the whole genome of *S. pseudintermedius*. Whole genome analysis of *S. pseudintermedius* may help to better understand the pathogenesis of canine pyodermas.

Methods and materials – Twenty-two strains of *S. pseudintermedius* isolated from the skin of five healthy dogs and 33 strains isolated from skin of 33 dogs with pyoderma were analysed. DNA was extracted and sequenced using Oxford Nanopore MinION, a new technology that delivers longer reads in a hand-held device. The pangenome was analysed and visualised with Anvi'o 6.1.

Results – Nanopore technology allowed the sequencing and *de novo* assembly of the genomes of 55 *S. pseudintermedius* strains isolated from healthy dogs and from dogs with pyoderma. The average genome size of *S. pseudintermedius* was 2.62 Mbp, with 48% being core genome. Pyoderma isolates contained a higher number of antimicrobial resistance genes, yet the total number of virulence factors genes did not change between isolates from healthy dogs and from dogs with pyoderma. Genomes of meticillin-resistant *S. pseudintermedius* (MRSP) strains were larger than those of meticillin-susceptible (MSSP) strains (2.80 Mbp versus 2.59 Mbp), as a consequence of a greater presence of antimicrobial resistance genes, phages and prophages.

Conclusions and clinical importance – This technique allows much more precise and easier characterisation of canine *S. pseudintermedius* populations and may lead to a better understanding of the pathogenesis of canine pyodermas.

Accepted 23 September 2021

Sources of Funding: Research grant RTI2018-101991-B-100 from the Ministerio de Ciencia e Innovación, Spain; Industrial Doctorate Program grant 2017DI037 AGAUR, Generalitat de Catalunya, Spain. Ministerio de Ciencia e Innovación, Spain: Torres Quevedo Project PTQ2018-009961.

Conflicts of Interest: 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.

This study was presented as a supporting original study at the 9th World Congress of Veterinary Dermatology, 20–24 October 2020, Sydney, Australia.

Introduction

Staphylococcus pseudintermedius is a component of the dog skin microbiota[1,2] and the main causative agent of pyoderma in this species.[3,4] A study using pulse-field gel electrophoresis (PFGE) of dogs with pyoderma concluded that *S. pseudintermedius* isolated from skin lesions (pustules) were identical to the *S. pseudintermedius* isolated from nonlesional sites of the same dog.[5] This finding indicated that the *S. pseudintermedius* causing skin infections very likely originated from commensal *S. pseudintermedius* populations in the canine skin. Thus, the current paradigm indicates that infection usually arises when the skin and mucosal barriers are altered by predisposing factors such as atopic dermatitis (AD), medical and surgical procedures, and/or immunosuppressive disorders.[3] Therefore, *S. pseudintermedius* is considered to be an opportunistic pathogen. Nevertheless, the mechanisms by which a commensal micro-organism is transformed into a pathogen are poorly understood.

The increasing development of antibiotic resistance in *S. pseudintermedius* populations and especially meticillin resistance (MR), which confers resistance to all beta-lactam antibiotics, is a very serious problem worldwide.[4] The reduction and rational use of antibiotics in veterinary medicine is one of the main strategies to reduce bacterial resistance and one that is advocated by health organisations. Understanding the pathogenic mechanisms of pyoderma could help prevent the development of pyoderma and help develop new therapies, which could reduce the use of antibiotics. Nevertheless, to date, it has been challenging to find a research strategy to learn more about how pathogenic *S. pseudintermedius* populations originate and the differences – if any – between commensal and pathogenic *S. pseudintermedius* populations.

The recent development of next-generation sequencing (NGS) techniques that allow relatively easy and economical mass sequencing of genomes has opened up a new pathway for studying infectious diseases. In particular, the sequencing of the complete genome of micro-organisms allows a very precise genotypic characterisation, identifying, for example, factors of virulence or antimicrobial resistance and allowing the comparison of different strains and isolates. The release of Oxford Nanopore Technologies' MinION in 2014 generated much excitement in the genomics community by offering portability (it measures approximately 3 cm x 10 cm), speed and the capability of producing reads of virtually any length.[6] The fact that it can work with longer reads facilitates the assembly of genomes and the characterisation, location and genomic context of virulence and resistance genes.[7] So even though it has a higher error rate than the Illumina system, it has quickly become very useful for rapid clinical responses and sequencing in the field.[8,9]

The objective of this study was two-fold. First, to determine whether the Nanopore technology allows a sequencing and *de novo* assembly of the whole genome of *S. pseudintermedius*. Second, in the event that the result was positive, to analyse and compare the genome of different *S. pseudintermedius* strains isolated from the skin of healthy dogs and from skin lesions of dogs with pyoderma. We aimed to reach a better understanding of the genome of *S. pseudintermedius* and detect differences that could explain the change from a commensal micro-organism to a pathogen.

Materials and methods

Bacterial cultures

The present study was carried out on 22 cultures of *S. pseudintermedius* isolated from the skin of six healthy dogs and 33 cultures isolated from 33 dogs diagnosed with pyoderma. In the case of the healthy dogs, the samples were obtained by rubbing a sterile swab on the perioral or abdominal skin for 15 s. In the dogs with pyoderma, samples were obtained, whenever possible, from the content of the pustules. In three cases where no obvious pustules were present, the samples were obtained from epidermal collarettes, as described in the literature.[10] The samples were cultured in blood agar at 37°C and incubated in aerobiosis for 24 h or 48 h, depending on the visual growth on the plate. Colonies with characteristic morphology of *S. pseudintermedius* were subcultured and stored in brain heart infusion (BHI) broth with 20% glycerol at −80°C for further studies. Before the DNA extraction procedure, cultures were plated in blood agar and seeded in 3 mL BHI broth at 37°C for 24 h in aerobiosis.

DNA extraction and sequencing

Staphylococcus pseudintermedius DNA from the BHI broth cultures was extracted with a Zymo BIOMICS DNA Miniprep Kit (Zymo Research; Irvine, CA, USA). DNA quality and quantity were determined using a Nanodrop 2000 Spectrophotometer and Qubit dsDNA BR Assay Kit (Fisher Scientific S.L.; Madrid, Spain). The sequencing libraries were prepared using 200 400 ng DNA which were subjected to transposase fragmentation using the Rapid Barcoding Sequencing kit (SQK-RBK004, Oxford Nanopore Technologies; Oxford Science Park, UK), and 12 barcoded samples were loaded in a MinION FLO-MIN106 v9.4.1 flow cell (Oxford Nanopore Technologies) and sequenced in a MinION Mk1B. The fast5 files were basecalled with Guppy 4.0.11 (Oxford Nanopore Technologies) with high accuracy base-calling mode, demultiplexed and with adapters trimmed. Reads with a quality score <7 were discarded.

Assembly and visualisation of the genomes

Nanoplot 1.27 (https://github.com/wdecoster/NanoPlot) was used to obtain the run summary statistics.[11] Sequences corresponding to *S. pseudintermedius* after taxonomy assignment using What's in my pot (WIMP) workflow from the EPI2ME platform[12] were *de novo* assembled using Flye 2.7.1.[13] Minimap 2.17 was used to align the assembled sequences to the raw data files.[14] The resulting contigs then were first polished with the graph-based correction method in Racon 1.4.13 (https://github.com/lbcb-sci/racon), followed by the neural-network based correction method used by Medaka 1.0.3 (https://nanoporetech.github.io/medaka/).

Genome completeness was assessed with CheckM 1.1.1.[15] Circlator 1.5.5 was used to identify the origin and the reverse complementary sequences.[16] Genomes were annotated with NCBI Prokaryotic Genome Annotation Pipeline (PGAP)[17] (after uploading to NCBI; Bioproject PRJNA685966), and the total numbers of coding sequences, rRNA and tRNA were determined. Comparative genome identity was assessed with FASTANI.[18]

Anvi'o 6.2 (https://merenlab.org/2016/11/08/pangenomics-v2/) allows comparison of shared genes.[19] It was used to determine the core genome and accessory genome of *S. pseudintermedius* isolates.

A custom phylogenetic tree was built with PATRIC[20] from the 55 *de novo* assembled genomes and 31 complete genomes available in NCBI as of May 2021, with 100 single-copy PATRIC PGFams selected by the codon tree method as homology groups (Fig. 1). The aligned proteins and coding DNA from single-copy genes were analysed with RAxML,[21] as included in PATRIC. The resulting Newick file was viewed in FigTree v1.4.4 (https://github.com/rambaut/figtree).

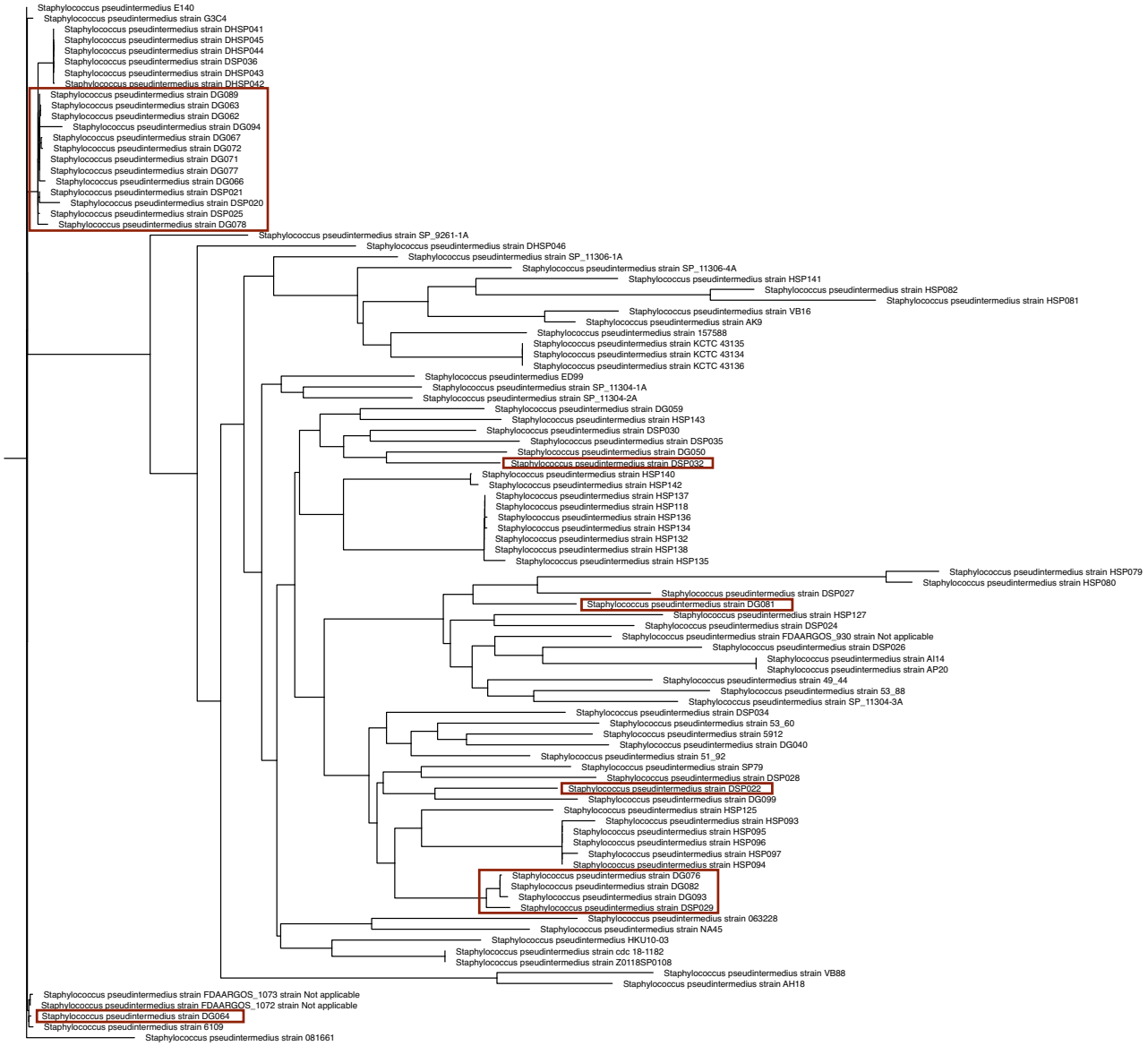

Figure 1. Phylogenetic tree built with PATRIC from 31 complete genomes available in NCBI and the 55 *de novo* assembled genomes from this study.
Thirty-three strains were isolated from dogs with pyoderma (strains with codes DG and DSP) and 22 from the skin of healthy dogs (strains with code HSP)(full data for each strain can be seen in Table 1). Red boxes indicate meticillin-resistant *Staphylococcus pseudintermedius* (MRSP) clusters.

Multilocus sequence type, antibiotic-resistance genes, virulence factors and bacteriophages

Multilocus sequence types (MLST) were assigned with MLST 2.0 software and its database 2.0.0.[22] Antibiotic resistance genes were identified with ABRICATE 0.8.13[23] (https://github.com/tseemann/abricate) with the CARD database.[24] Plasmids (replicons) were identified with PLASMIDFINDER 2.1.[25] A custom database also was created to analyse the virulence factors (SPVFDB), containing 58 genes encoding for virulence factors that include exfoliative toxins, enterotoxins, leukocidins, pore-forming proteins and intercellular adhesion proteins. Subsequently, the results were filtered by genes with identity and coverage ≥90%. PHIGARO 2.2.6[26] and VIRSORTER 1.0.6[27] were used to identify phage and bacteriophage sequences within the genomes.

Data availability

The whole-genome assemblies were deposited at GenBank repository under BIOPROJECT PRJNA685966 and with the accession numbers CP066702–066718, CP066884, CP066885 and JAENBQ00000000–JAENDF000000000. The version described in this paper is version 1. Full data from the whole-genome assemblies also have been published elsewhere.[28]

Results

DNA extraction and purification from BHI broth cultures developed as expected. Sequencing libraries were prepared with 12 isolates per library with the Rapid Barcoding kit (12-plex barcode libraries) and sequenced in the Nanopore MinION Mk1B for 24 h. The mean size of the reads was 2,600 bp. After quality assessment of the reads, all genomes were assembled, achieving completeness >96.7% and considered high-quality (completeness >90%, contamination <5%) for further analyses. Mean sequencing coverage was ×249.18. Table 1 presents the most relevant genomic data of the 55 isolates of *S. pseudintermedius* according to the origin of the

Table 1. Results of nanopore sequencing and *de novo* assembly of the 55 *Staphylococcus pseudintermedius* isolates, including the following parameters: health status, isolate, country of origin of isolate, genome size, %GC, multilocus sequence types (MLST), meticillin-resistance (MRSP) and SSC-mec type cassette, number of phages and prophages, virulence factor and antimicrobial resistance genes identified, and multidrug-resistance genotype.

Healthy skin/ pyoderma	Isolate	Origin of sample	Genome size (bp)	GC %	MLST	Meticillin resistance (mecA)	SCCmec Type	Phages & prophages	Virulence factor genes	Antimicrobial resistance genes	Multidrug resistant genotype
Pyoderma	D_G062	Italy	2991046	37.4	71	MRSP	SCCmec_type_II-III	9	41	12	yes
Pyoderma	D_G063	Italy	2981523	37.4	71	MRSP	SCCmec_type_II-III	9	41	11	yes
Pyoderma	D_G064	Italy	2895060	37.5	71	MRSP	SCCmec_type_II-III	7	40	10	yes
Pyoderma	D_G066	Italy	2808032	37.5	71	MRSP	SCCmec_type_II-III	5	41	11	yes
Pyoderma	D_G067	Italy	2896399	37.5	71	MRSP	SCCmec_type_II-III	8	40	10	yes
Pyoderma	D_G071	Italy	2807986	37.5	71	MRSP	SCCmec_type_II-III	5	42	11	yes
Pyoderma	D_G072	Italy	2837133	37.4	71	MRSP	SCCmec_type_II-III	6	41	10	yes
Pyoderma	D_G077	Italy	2791496	37.5	71	MRSP	SCCmec_type_II-III	6	42	10	yes
Pyoderma	D_G078	Italy	2799396	37.4	71	MRSP	SCCmec_type_II-III	5	41	11	yes
Pyoderma	D_G089	Italy	2797937	37.5	71	MRSP	SCCmec_type_II-III	5	41	11	yes
Pyoderma	D_G094	Italy	2849103	37.4	71	MRSP	SCCmec_type_II-III	5	39	11	yes
Pyoderma	D_SP020	Spain	2793830	37.5	71	MRSP	SCCmec_type_II-III	5	42	10	yes
Pyoderma	D_SP021	Spain	2795724	37.5	71	MRSP	SCCmec_type_II-III	5	41	10	yes
Pyoderma	D_SP025	Spain	2805515	37.5	71	MRSP	SCCmec_type_II-III	6	41	10	yes
Pyoderma	D_SP036	Spain	2927015	37.3	71	MRSP	SCCmec_type_II-III	5	41	10	yes
Pyoderma	D_G076	Italy	2780144	37.2	258	MRSP	SCCmec_type_IVg	1	39	8	yes
Pyoderma	D_G082	Italy	2626557	37.6	258	MRSP	SCCmec_type_IVg	1	39	7	yes
Pyoderma	D_G093	Italy	2648891	37.6	258	MRSP	SCCmec_type_IVg	2	39	9	yes
Pyoderma	D_SP029	Spain	2723805	37.4	258	MRSP	SCCmec_type_IVg	1	40	8	yes
Pyoderma	D_G081	Italy	2694747	37.6	301	MRSP	SCCmec_type_IVg	2	38	9	yes
Pyoderma	D_SP032	Argentina	2670199	37.5	1631	MRSP	none	1	43	9	yes
Pyoderma	D_SP022	Spain	2767901	37.2	Unknown	MRSP	none	3	41	9	yes
Pyoderma	D_G050	Italy	2637713	37.6	Unknown	MSSP		2	42	8	yes
Pyoderma	D_G099	Italy	2623014	37.6	Unknown	MSSP		1	45	8	yes
Pyoderma	D_SP035	Argentina	2642865	37.6	Unknown	MSSP		2	43	6	yes
Pyoderma	D_SP028	Spain	2575420	37.6	Unknown	MSSP		1	42	6	yes
Pyoderma	D_SP024	Spain	2762026	37.5	611	MSSP		3	42	4	yes
Pyoderma	D_SP026	Spain	2567628	37.6	503	MSSP		1	42	1	no
Pyoderma	D_SP034	Argentina	2550368	37.6	1827	MSSP		0	44	1	no
Pyoderma	D_SP027	Spain	2717194	37.3	Unknown	MSSP		2	44	1	no
Pyoderma	D_SP030	Argentina	2612059	37.6	Unknown	MSSP		1	41	1	no
Pyoderma	D_G059	Italy	2573568	37.6	Unknown	MSSP		1	40	1	no
Pyoderma	D_G040	Italy	2564892	37.7	Unknown	MSSP		1	43	1	no
Healthy skin	H_SP141	Spain	2622529	37.6	257	MSSP		2	42	2	no
Healthy skin	H_SP125	Spain	2551473	37.6	1061	MSSP		1	46	1	no
Healthy skin	H_SP118	Spain	2512855	37.7	1248	MSSP		0	42	1	no
Healthy skin	H_SP079	Spain	2585691	37.5	Unknown	MSSP		0	43	3	yes
Healthy skin	H_SP080	Spain	2585570	37.5	Unknown	MSSP		0	43	3	yes
Healthy skin	H_SP081	Spain	2621254	37.7	Unknown	MSSP		3	40	1	no
Healthy skin	H_SP082	Spain	2620663	37.7	Unknown	MSSP		3	40	1	no
Healthy skin	H_SP093	Spain	2590335	37.5	Unknown	MSSP		2	42	1	no
Healthy skin	H_SP094	Spain	2570595	37.6	Unknown	MSSP		2	42	1	no
Healthy skin	H_SP095	Spain	2575879	37.6	Unknown	MSSP		2	42	1	no
Healthy skin	H_SP096	Spain	2578330	37.6	Unknown	MSSP		2	42	1	no
Healthy skin	H_SP097	Spain	2575223	37.6	Unknown	MSSP		2	42	1	no
Healthy skin	H_SP127	Spain	2690618	37.5	Unknown	MSSP		2	43	2	no
Healthy skin	H_SP132	Spain	2515164	37.7	Unknown	MSSP		0	42	1	no
Healthy skin	H_SP134	Spain	2514594	37.7	Unknown	MSSP		0	42	1	no
Healthy skin	H_SP135	Spain	2512727	37.7	Unknown	MSSP		0	42	1	no
Healthy skin	H_SP136	Spain	2512726	37.7	Unknown	MSSP		0	42	1	no
Healthy skin	H_SP137	Spain	2512757	37.7	Unknown	MSSP		0	42	1	no
Healthy skin	H_SP138	Spain	2512830	37.7	Unknown	MSSP		0	42	1	no
Healthy skin	H_SP140	Spain	2597272	37.6	Unknown	MSSP		1	39	1	no
Healthy skin	H_SP142	Spain	2597098	37.6	Unknown	MSSP		1	39	1	no
Healthy skin	H_SP143	Spain	2779670	37.6	Unknown	MSSP		5	40	1	no

sample and MLST. The full data for all isolates can be seen in Table S1 of the Supporting information. Pairwise average nucleotide identity (ANI) values among the 55 *S. pseudintermedius* isolates are shown in Table S2. The average nucleotide identity among the 55 isolates was 99.27%.

In the case of isolates from dogs with pyoderma, ST71 and ST257 were predominant. However, among isolates from healthy dogs, these two STs were absent, and a long list of previously unreported STs was identified (19 of 22), even with 100% coverage and identity values for the alleles of the seven genes analysed in the MLST.

Considering all 55 samples together, the average genome size of *S. pseudintermedius* was 2.62 Mbp. Comparing the genomes of the isolates from dogs with pyoderma (n = 33) with those from healthy dogs (n = 22), several differences were detected. The genomes of the pathogenic strains were, on average, larger than those of the strains from healthy dogs (2.74 Mbp versus 2.58 Mbp; $P = 4.883e\text{-}07$; Figure 2a). However, when we analysed

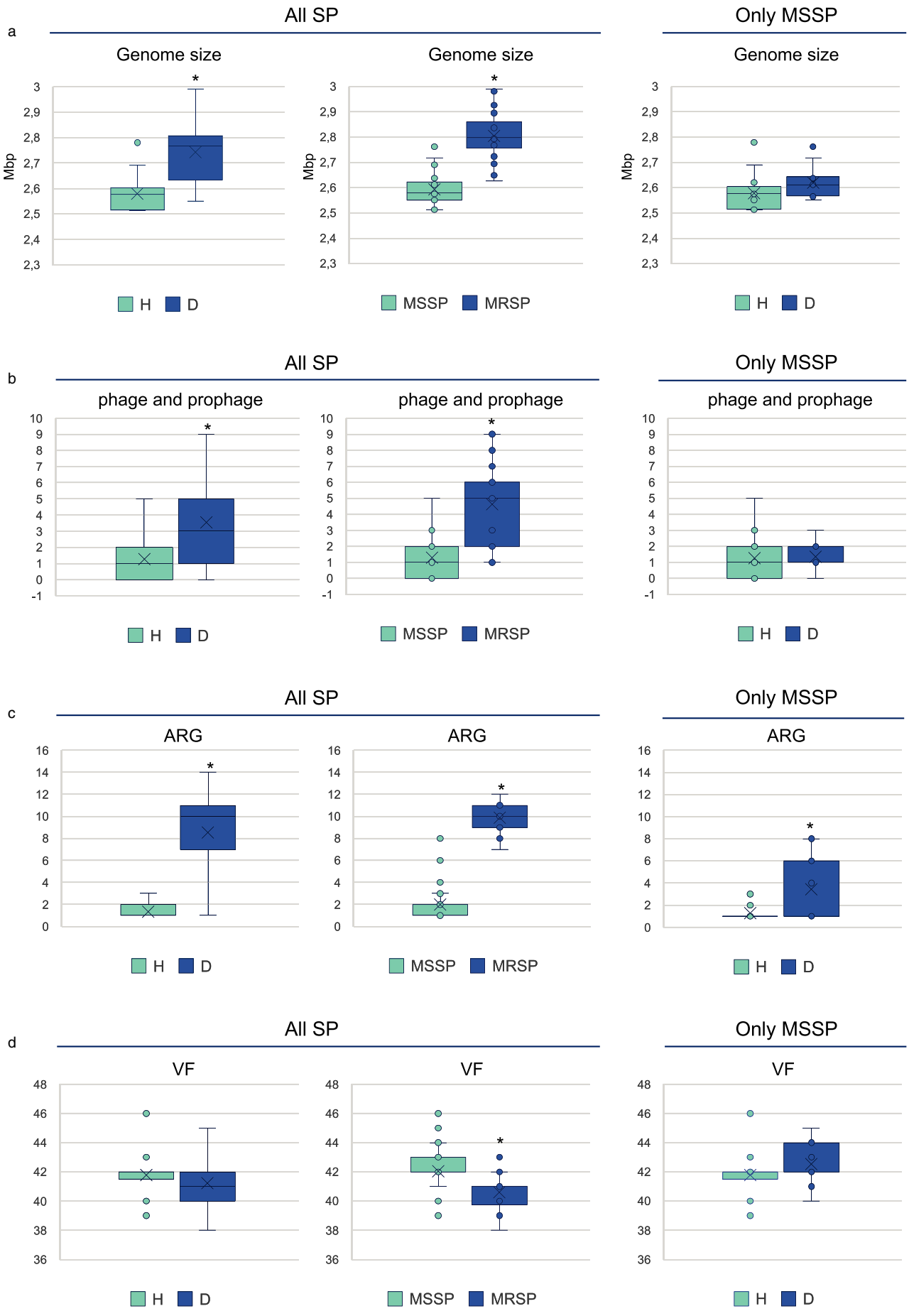

Figure 2. Meticillin-resistant *Staphylococcus pseudintermedius* (MRSP) are larger and contain more phages, prophages and antimicrobial resistant genes than meticillin-susceptible *S. pseudintermedius* (MSSP).
Box plots show the distribution of (a) genome size, (b) phage and prophage number (c) antimicrobial resistant gene number and (d) virulence factor number of the *S. pseudintermedius* strains isolated and sequenced in this study. Left column shows results of all *S. pseudintermedius* (all SP, n = 55), comparing the strains isolated from healthy dogs (turquoise) and those isolated from dogs with pyoderma (blue). Middle column shows results of all *S. pseudintermedius* (all SP, n = 55), comparing the MSSP strains (turquoise) and the MRSP strains (blue). Right column shows results of the MSSP strains (only MSSP, n = 33), comparing the strains isolated from healthy dogs (turquoise) and those isolated from dogs with pyoderma (blue). Shapiro–Wilk normality tests revealed that the data were not normally distributed. A Wilcoxon rank sum exact test revealed significant differences in those cases marked by asterisks, denoting significant statistical differences ($P < 0.05$).

the data by comparing the meticillin-resistant *S. pseudintermedius* (MRSP) isolates, which contained the *mecA* gene (n = 22) to the meticillin-susceptible (MSSP) isolates (n = 33) (Table 1), MRSP showed a significant larger genome size than the MSSP strains (2.80 Mbp versus 2.59 Mbp; $P = 4.173e\text{-}12$; Figure 2a). In agreement,

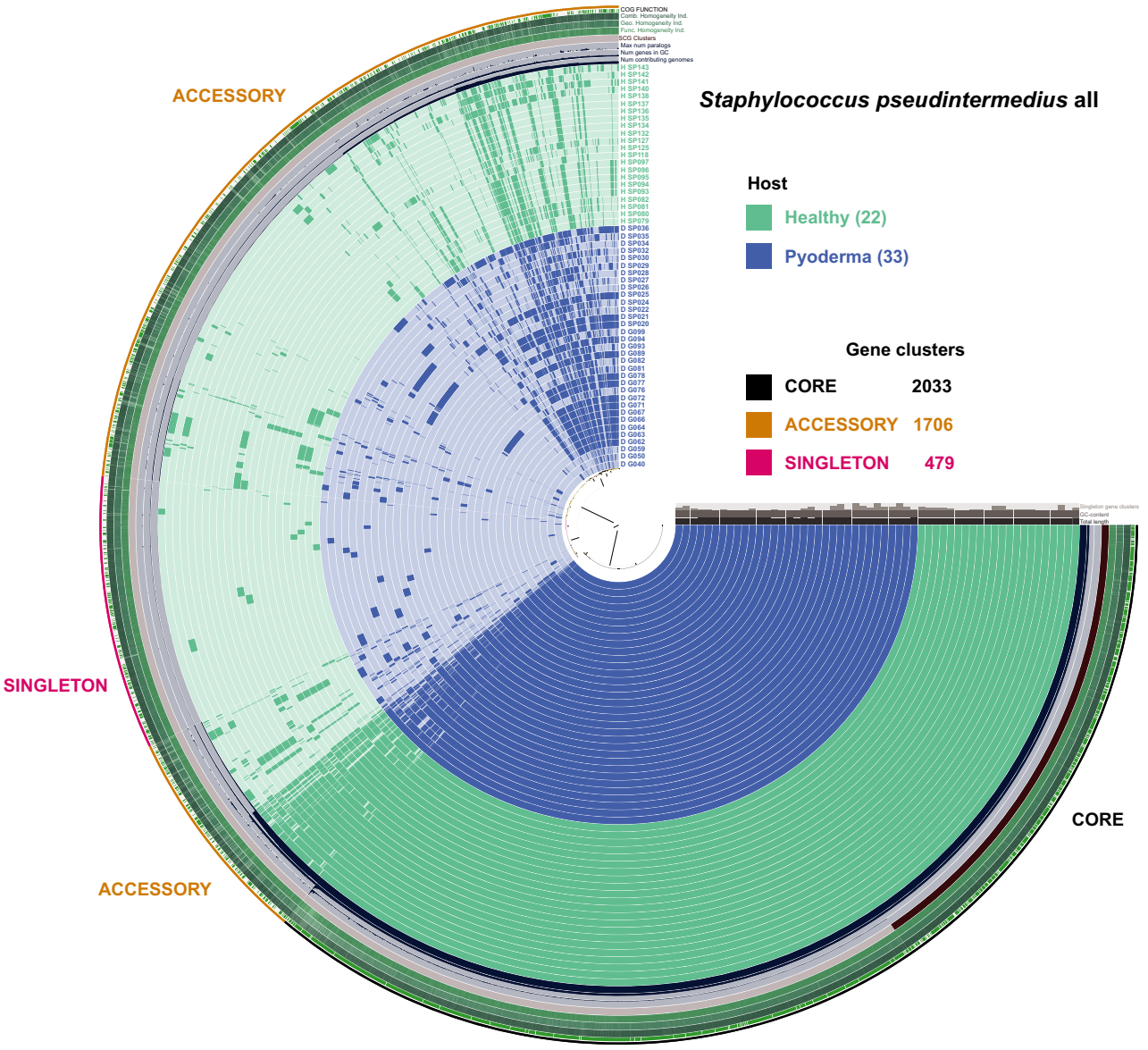

Figure 3. Global pangenome visualization of all *Staphylococcus pseudintermedius* strains shows 48% of core genome and 52% of accessory genome.
Pangenome results of all *S. pseudintermedius* strains (n = 55), comparing the strains isolated from healthy dogs (turquoise) and the strains isolated from dogs with pyoderma (blue). Core genome is by definition the part of the pangenome that is present and shared by all the genomes within the pangenome. Accessory genome is specific for a group of strains within the pangenome and singletons are strain specific genome sequences. Visualization of pangenome analyses carried by Anvi'o. Central dendrogram clustering of samples is ordered by gene cluster presence/absence. Items order: presence absence (D, Euclidean; L, Ward).

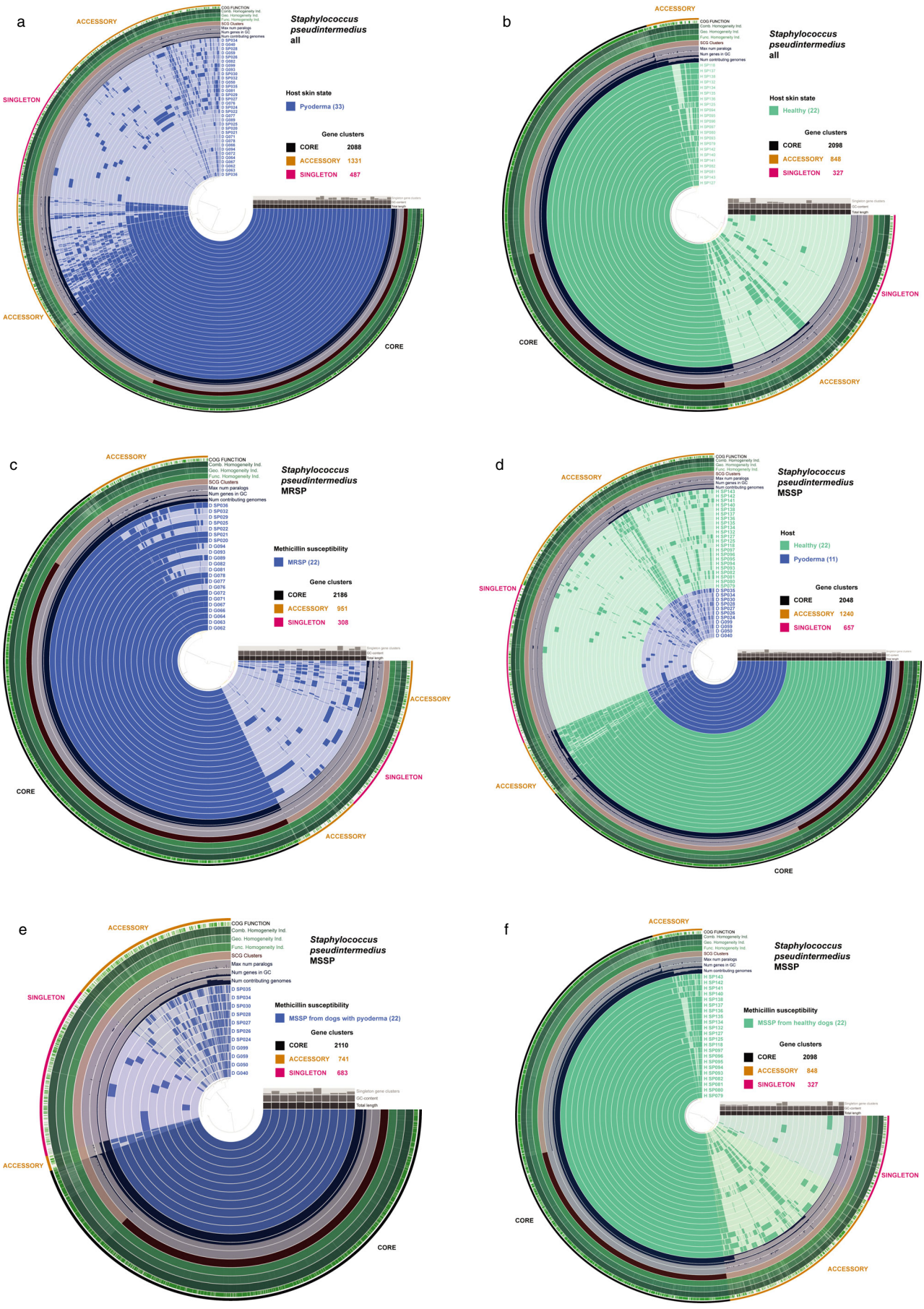

Figure 4. Pangenome analyses show that meticillin-susceptible *Staphylococcus pseudintermedius* (MSSP) strains are more diverse than methicillin-resistant (MRSP) strains.
(a) Split pangenome results of all *S. pseudintermedius* strains isolated from dogs with pyoderma (blue; n = 33). (b) Split pangenome results of all *S. pseudintermedius* strains isolated from healthy dogs (turquoise; n = 22). (c) Split pangenome results of all MRSP strains (blue; n = 22). (d) Split pangenome results of all MSSP strains (turquoise; n = 33). (e) Split pangenome results of the MSSP strains (only MSSP, n = 33) isolated from dogs with pyoderma (blue, n = 11). (F) Split pangenome results of the MSSP strains (only MSSP, n = 33) isolated from healthy dogs (turquoise, n = 22). Core genome is by definition the part of the pangenome that is present and shared by all the genomes within the pangenome. Accessory genome is specific for a group of strains within the pangenome and singletons are strain-specific genome sequences. Visualisation of pangenome analyses was carried out using Anvi'o. Central dendrogram clustering of samples is ordered by gene cluster presence/absence. Item order: presence absence (D, Euclidean; L, Ward).

MRSP showed, on average, a significantly higher number of phage and prophages (five versus one; $P = 3.924e\text{-}06$; Figure 2b) and antibiotic resistance genes ($P = 1.12e\text{-}10$; Figure 2c) than MSSP strains. By contrast, the average number of virulence factors was significantly lower in MRSP than in MSSP isolates (41 versus 42; $P = 0.0002$; Figure 2d). Altogether, these results indicate that the differences observed previously between isolates from healthy dogs and dogs with pyoderma (H and D; Figure 2) resulted mainly from the differences between MRSP and MSSP strains. Supporting this, when comparing the healthy (n = 22) and pyoderma isolates (n = 11; Table 1) within the MSSP isolates only (n = 33; Table 1), we did not observe any differences in either the genome size (2.58 Mbp versus 2.59 Mbp; $P = 0.133$; Figure 2a) or the number of phages and prophages (one versus one; $P = 0.6195$; Figure 2b) or the virulence factors number (42 versus 43; $P = 0.1491$; Figure 2d). However, a significantly higher number of antibiotic resistance genes was identified in MSSP isolates from dogs with pyoderma when compared to MSSP isolates from healthy dogs (three versus one; $P = 0.03848$; Figure 2c).

Global pangenome analyses of all the 55 *S. pseudintermedius* genomes revealed 48% of core genome and 52% of accessory genome (Figure 3). When pangenome analyses were carried out comparing the 22 MRSP isolates against the 33 MSSP isolates, we observed a larger core genome within the MRSP than within the MSSP pangenome (63% versus 52%; Figure 4c, d). These results suggest that the MSSP pangenome is more diverse than the MRSP pangenome. In agreement with this hypothesis, the specific singleton genome within the MSSP pangenome was larger than in the MRSP pangenome (17% versus 9%; Figure 4c, d). Supporting these results, 70% of the MRSP isolates were identified as ST71 MLST, while the MSSP comprised a mixture of different MLSTs, most as yet unidentified and thus classified as unknown (Table S1). Comparison of the pyoderma and healthy cases within the MSSP isolates only, showed similar core and accessory genome sizes (60% versus 64%; Figure 4e, f). Altogether, the pangenome analyses indicate that differences detected between isolates from healthy dogs and from dogs with pyoderma are mainly the result of differences between MRSP and MSSP strains.

Fifty virulence genes were identified, including accessory gene regulators (agr A, B, C, D), adhesins and biofilm formation genes (sps A-H,ICA A-D,ebps), toxins (expA, expB, siet, speta) and invasins (Luk F, S; Hlb, Coa). Thirty-two (58.18%) of the virulence factors were present in all 55 isolates. Even though there are no differences in virulence factor content between pyoderma isolates and healthy isolates (Figure 2d), it is noteworthy that four genes coding for surface proteins (spsD, spsF, spsP and spsQ) involved in colonisation by binding to the host's extracellular matrix were present only on pyoderma isolates. The remaining genes for virulence factors were present in some isolates only, without being specific for isolates from healthy dogs or dogs with pyoderma.

The current analysis also provided detailed information about antimicrobial resistance genes (Table S1). Twenty-two of 33 of the isolates from dogs with pyoderma (66.6%) carried the *mecA* gene and were characterised genotypically as MRSP. However, none of the isolates from healthy dogs carried the *mecA* gene and, therefore, all were considered MSSP. Furthermore, *mecI* and *mecR1* genes, two genes involved in *mecA* gene expression, were present in all ST71 isolates, the ST1631 isolate and in one unknown ST. However, the five isolates that belonged to clonal complex 258 (four ST258, one ST301) possessed only the *mecA* gene. Table S1 presents all of the data related to the presence of AMR genes in the 55 strains; those cases where phenotypic sensitivity profiling was available also are indicated.

Considering that multidrug resistance (MDR) is defined when the micro-organism is resistant to at least three families of antibiotics, we observed an MDR gene profile in 81.82% (27 of 33) of the *S. pseudintermedius* isolates from lesional skin versus 9.09% (two of 22) of the isolates from healthy dogs. Among the 27 MDR strains isolated from dogs with pyoderma, 22 held the *mecA* gene (Table 1).

Discussion

In the present study, we were able to sequence and *de novo* assemble the complete genome of multiple isolates of *S. pseudintermedius* using Nanopore technology. The method is notably faster, simpler and cheaper than other sequencing systems, such as the traditional Illumina system. In addition, the use of Anvi'o allowed for easy comparison of the genomes.

The complete genome analysis of *S. pseudintermedius* is undoubtedly a very useful tool to characterise the different isolates. In particular, it allowed the identification and location of virulence and antimicrobial resistance factor genes as well as the presence of phages, and is very likely to help to better understand the pathogenesis of canine pyodermas.

The *S. pseudintermedius* isolates from our study have an average genome size of 2.62 Mbp. This size is similar

to that reported by other authors[29-31] and the NCBI *S. pseudintermedius* reference genome (NC_014925.1). This is an example of an open genome, with a core genome of 48% and an accessory genome of 52%.

The *S. pseudintermedius* isolates from pyoderma cases had a significantly larger genome than isolates from healthy skin. However, when further comparative analyses were performed, it was found that the size difference was to the result of the greater presence of MRSP in the group of isolates from dogs with pyoderma. This larger genome size in MRSP compared to MSSP (2.80 Mbp versus 2.59 Mbp) was undoubtedly a consequence of the increased presence of phages, prophages and antimicrobial resistance factors in MRSP.

Interestingly, hardly any differences were found in the presence of virulence factor genes among the different strains. Only a few genes encoding surface proteins (sps BF) were found, exclusively or much more frequently in samples from pyoderma cases. The surface proteins of *S. pseudintermedius* express binding activity to components of the host's extracellular matrix (ECM), including fibronectin, fibrinogen and cytokeratin10, and they are considered to be virulence factors associated with bacterial survival, immune evasion and biofilm formation.[32,33] Our results support previous data indicating that these proteins may play a role in staphylococcal colonisation and/or infection.[33-35] Interestingly, however, most virulence factors are present also in strains isolated from healthy dogs, indicating that commensal staphylococcal populations already have full pathogenic potential. In any case, it is worth noting that this genomic approach investigates only the presence of certain genetic elements in the bacterial genome and not their expression or functionality. In the case of both antimicrobial resistance and virulence factors, it is necessary to complement genetic studies with functionality tests.

The isolates from lesional skin of dogs with pyoderma had a much higher number of antimicrobial resistance genes than isolates from skin of healthy dogs (Figure 2c), an expected finding that probably reflects prior antibiotic exposure. This finding, again, resulted from the presence of MRSP in this group (22 of 33 isolates were MRSP). Figure 2c shows the greater presence of AMRG in MRSPs compared to MSSPs.

The present approach allowed not only the detection of antimicrobial resistance genes, but also its location in the genome. For instance, three ST71 isolates and one ST258 isolate harboured the *tetK* gene in another contig rather than the chromosome, which was probably a plasmid, as these isolates harboured the replicon rep7_1_repC too (Table S1). Three ST258 and one ST301 isolates harboured the *tetM* gene in the chromosome (Table S1). Five ST71 and two unknown ST isolates harboured the chloramphenicol resistance gene in the same contig as the replicon rep7_7_rep(pKH7), suggesting that this gene is linked to a plasmid in those isolates (Table S1).

An unexpected result was the difference in the MLSTs identified in animals with pyoderma and in healthy animals. In the strains from dogs with pyoderma, the STs identified were those most frequently reported in Europe (ST71, ST258). However, none of these STs were detected in samples from healthy dogs, which seems to go against the paradigm constructed from the study by Pinchberk and colleagues stating that the *S. pseudintermedius* causing skin infections very likely originate from commensal *S. pseudintermedius* populations.[5] The difference also could be a consequence of the fact that all ST71 and ST258 strains isolated so far (https://pubmlst.org) are MRSP and in the group of healthy dogs all isolates were MSSP. In any case, our results should be interpreted with caution because the number of samples was small, especially from healthy dogs. Extensive research and sequencing of *S. pseudintermedius* from healthy dogs are needed to clarify whether there are typically pathogenic STs and others that are associated only with commensal behaviour. Additionally, to identify genetic changes associated with the transformation of a commensal microorganism into a pathogen we should study isolates obtained from the same dog, from both lesional and nonlesional areas.

In summary, our approach has allowed the sequencing and *de novo* assembly of the complete genome of 55 *S. pseudintermedius* strains isolated from healthy dogs and from dogs with pyoderma, and their analysis and comparison. This technique allows much more precise characterisation of *S. pseudintermedius* populations in dog skin than techniques usedpreviously , such as PFGE. Our hope is that this new approach will lead to a better understanding of the pathogenesis of canine pyodermas.

References

1. Paul NC, Bärgman SC, Moodley A, *et al*. Staphylococcus pseudintermedius colonization patterns and strain diversity in healthy dogs: a cross-sectional and longitudinal study. *Vet Microbiol* 2012; 160: 420–427.
2. Rubin JE, Chirino-Trejo M. Prevalence, sites of colonization, anantimicrobial resistance among Staphylococcus pseudintermedius isolated from healthy dogs in Saskatoon, Canada. *J Vet Diagn Invest* 2011; 23: 351–354.
3. Bannoehr J, Guardabassi L. *Staphylococcus pseudintermedius* in the dog: taxonomy, diagnostics, ecology, epidemiology and pathogenicity. *Vet Dermatol* 2012; 23: 253-66, e51-2.
4. Loeffler A, Lloyd DH. What has changed in canine pyoderma? A narrative review. *Vet J* 2018; 235: 73–82.
5. Pinchbeck LR, Cole LK, Hillier A, *et al*. Genotypic relatedness of staphylococcal strains isolated from pustules and carriage sites in dogs with superficial bacterial folliculitis. *Am J Vet Res* 2006; 67: 1,337–1,346.
6. Jain M, Olsen HE, Paten B, *et al*. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* 2016; 17: 239.
7. Viñes J, Cuscó A, Francino O. Hybrid assembly from a pathogenic methicillin- and multidrug-resistant *Staphylococcus pseudintermedius* strain Isolated from a case of canine otitis in Spain. *Microbiol Resour Announc* 2020; 9: e01121–19.
8. Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet* 2018; 19: 9–20.
9. Lu H, Giordano F, Ning Z. Oxford nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics* 2016; 14: 265–279.
10. Hillier A, Lloyd DH, Weese JS, *et al*. Guidelines for the diagnosis and antimicrobial therapy of canine superficial bacterial folliculitis (Antimicrobial Guidelines Working Group of the International Society for Companion Animal Infectious Diseases). *Vet Dermatol* 2014; 25: 163-e43.

11. De Coster W, D'Hert S, Schultz DT, *et al*. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 2018; 34: 2666–2669.
12. Juul S, Izquierdo F, Hurst A, *et al*. What's in my pot? Real-time species identification on the MinION™. BioRxiv 2015. https://doi.org/10.1101/030742.
13. Kolmogorov M, Yuan J, Lin YU, *et al*. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019; 37: 540–549.
14. Li H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 2018; 34: 3,094–3,100.
15. Parks DH, Imelfort M, Skennerton CT, *et al*. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015; 25: 1,043–1,055.
16. Hunt M, Silva ND, Otto TD, *et al*. Circlator: Automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 2015; 16: 294.
17. Tatusova T, DiCuccio M, Badretdin A, *et al*. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 2016; 44: 6,614–6,624.
18. Jain C, Rodriguez-R LM, Phillippy AM, *et al*. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018; 9: 5,114.
19. Eren AM, Esen ÖC, Quince C, *et al*. Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ* 2015; 3: e1319.
20. Davis JJ, Wattam AR, Aziz RK, *et al*. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Res* 2020; 48(D1), D606–D612.
21. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014; 30: 1312–1313.
22. Larsen MV, Cosentino S, Rasmussen S, *et al*. Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* 2012; 50: 1,355–1,361.
23. Seemann T. ABRicate. GitHub 2017. Available at: https://github.com/tseemann/abricate Accessed Oct 6, 2021.
24. Jia B, Raphenya AR, Alcock B, *et al*. CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2017; 45(D1): D566–D573.
25. Carattoli A, Zankari E, García-Fernández A, *et al*. *In silico* detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014; 58: 3895–3903.
26. Starikova EV, Tikhonova PO, Prianichnikov NA, *et al*. Phigaro: high-throughput prophage sequence annotation. *Bioinformatics* 2020; 36: 3,882–3,884.
27. Roux S, Enault F, Hurwitz BL, *et al*. VirSorter: Mining viral signal from microbial genomic data. *PeerJ* 2015; 3: e985.
28. Francino O, Pérez D, Viñes J, *et al*. Whole-genome sequencing and *de novo* assembly of 61 *Staphylococcus pseudintermedius* isolates from healthy dogs and dogs with pyoderma. *Microbiol Res Announc* 2021; 10: e00152–21.
29. Tse H, Tsoi HW, Leung SP, *et al*. Complete genome sequence of the veterinary pathogen *Staphylococcus pseudintermedius* strain HKU10-03, isolated in a case of canine pyoderma. *J Bacteriol* 2011; 193: 1,783–1,784.
30. Eichhorn I, Lübke-Becker A, Lapschies A-M, *et al*. Draft genome sequence of *Staphylococcus pseudintermedius* strain 13–13613, isolated from a case of canine pyoderma. *Microbiol Resour Announc* 2020; 9: e00027–20.
31. Cao W, Hicks K, White A, *et al*. Draft genome assemblies of two *Staphylococcus pseudintermedius* strains isolated from canine skin biopsy specimens. *Microbiol Resour Announc* 2020; 9: e00369–20.
32. Foster TJ. Colonization and infection of the human host by staphylococci: adhesion, survival and immune evasion. *Vet Dermatol* 2009; 20: 456–470.
33. Sung JM-L, Lloyd DH, Lindsay JA. Staphylococcus aureus host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. *Microbiology* 2008; 154: 1,949–1,959.
34. Bannoehr J, Ben Zakour NL, Reglinski M, *et al*. Genomic and surface proteomic analysis of the canine pathogen *Staphylococcus pseudintermedius* reveals proteins that mediate adherence to the extracellular matrix. *Infect Immun* 2011; 79: 3074–3086.
35. Phumthanakorn N, Chanchaithong P, Prapasarakul N. Development of a set of multiplex PCRs for detection of genes encoding cell wall-associated proteins in Staphylococcus pseudintermedius isolates from dogs, humans and the environment. *J Microbiol Methods* 2017; 142: 90–95.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Table S1. Data for the 55 sequenced genomes. For each genome the following parameters are indicated: MLST (multilocus site typing); location (country of origin of sample); genome size (bp); number of contigs; longest contig (bp); number of phages and prophages detected; presence of different antimicrobial resistance genes and presence of plasmids.

Table S2. Pairwise average nucleotide identity (ANI) values among the 55 *S. pseudintermedius* isolates. Values >99.995 are indicated in bold.