Medical Research Archives

ESMED
European Society of Medicine

OPEN ACCESS

Published: September, 30, 2022

Citation: Christensen JJ, Jensen CS, et al., 2022. Genomic identification of streptococcal strains and relation to clinical characteristics. A substudy to The Partial Oral Treatment of Endocarditis (POET) Trial, Medical Research Archives, [online] 10(9). https://doi.org/10.18103/mra.v10i9.3037

DOI
https://doi.org/10.18103/mra.v10i9.3037

ISSN: 2375-1924

RESEARCH ARTICLE

Genomic identification of streptococcal strains and relation to clinical characteristics. A substudy to The Partial Oral Treatment of Endocarditis (POET) Trial.

Jens J. Christensen[1,2], Christian S. Jensen[1,3], Rimtas Dargis[1], Xiaohui C. Nielsen[1], Mia M. Pries-Heje[4], Christoffer Wiingaard[4], Nikolaj Ihlemann[4], Sabine Gill[5], Niels E. Bruun[2,6,7], Hanne Elming[6], Jonas A. Povlsen[8], Trine Madsen[9], Kaare T. Jensen[8], Kurt Fuursted[10], Lauge Østergaard[4], Ulrik Christiansen[9], Flemming S. Rosenvinge[12], Jannik Helweg-Larsen[13], Emil L. Fosbøl[4], Lars Køber[2,4], Christian Torp-Pedersen[2,7,14], Niels Tønder[14], Claus Moser[3,15], Kasper Iversen*[2,11], Henning Bundgaard*[2,4]

[1] The regional department of clinical microbiology, Zealand University Hospital, Køge, Denmark
[2] Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
[3] Department of Clinical Microbiology, Rigshospitalet, Copenhagen University Hospital, Denmark
[4] Department of Cardiology, The Heart Center, Rigshospitalet, Copenhagen University Hospital, University of Copenhagen, Denmark
[5] Department of Cardiology, Odense University Hospital, Denmark
[6] Department of Cardiology, Zealand University Hospital, Roskilde, Denmark
[7] Department of Clinical Medicine, Aalborg University, Denmark
[8] Department of Cardiology, Aarhus University Hospital, Denmark
[9] Department of Cardiology, Aalborg University Hospital, Denmark
[10] Department of Bacteria, parasites and fungy, Statens Serum Institut, Denmark
[11] Department of Cardiology and Emergency Department, Herlev-Gentofte University Hospital, University of Copenhagen, Denmark
[12] Department of Clinical Microbiology, Odense University Hospital, Odense University, Denmark
[13], Department of Infectious Diseases, Rigshospitalet, University of Copenhagen, Denmark
[14]Department of Cardiology, Hillerød Hospital, University of Copenhagen, Denmark
[15].Costerton Biofilm Center, Department for Immunology and Microbiology, University of Copenhagen

*shared last authorship

Correspondence: jejc@regionsjaelland.dk

ABSTRACT
The Danish National Partial Oral Treatment of Endocarditis Trial (POET) demonstrated non-inferiority of partly peroral compared to intravenous antibiotic therapy for infective endocarditis (IE) caused by *Streptococcus* spp, *Enterococcus faecalis, Staphylococcus aureus*, or coagulase-negative staphylococci. Identifications by whole genome sequencing (WGS) of available streptococcal strains were related to clinical data. Sequences were obtained using Illumina technology (Miseq®) followed by core genome analysis and single-nucleotide polymorphism phylogeny examinations. Average nucleotide identification (ANI) calculated using the tool fastANI. Informations on +/- preexisting valve prosthesis, valve surgery and outcome related to obtained identifications. Streptococcal strains (n=123) from 117 patients were WGS examined. Twelve percent were pyogenic group strains and 88% belonged to viridans groups, mainly mitis and bovis groups. Phylogenetic trees were in accordance regarding species and subspecies identifications. High ANI percentages to type strains were found. Respectively 39, 60 and 16 IE cases involved mitral, aortic or both valves. IE caused by pyogenic group or mitis plus bovis group streptococci most frequent affected, respectively, mitral and aortic valves. Thirty-one patients (26%) had a

preexisting prosthesis; notably, in 50% of bovis group IE cases. Fifty-six patients had valve surgery done during the current disease; 8% and 93% of patients having, respectively, pyogenic group and mitis group strains as causative agents. Of patients allocated to intravenous or intravenous followed by peroral antibiotic treatment, respectively 26 and 30 had valve surgery done during the current disease. Composite outcome (all-cause mortality, unplanned cardiac surgery, embolic events, or relapse of bacteremia with the primary pathogen) at five-year follow-up comprised in total 39 events. In conclusion, molecular examinations adds on substantially by detailing species and subspecies affiliations. A broad spectrum of streptococcal species and subspecies causing IE were identified with mitis- and bovis group strains dominating. Relating strain identifications to clinical data can assist in planning and treating confirmed/suspected IE patients. Adding WGS identification of streptococci in selected patients groups (e.g. IE) in order to expand number of cases characterized in detail seems ideal and advocates for centralized registration of results to reveal important clinical relations.

Keywords: whole-genome sequencing, infective endocarditis, streptococci, partial oral antibiotic treatment, clinical outcome

INTRODUCTION

Infective endocarditis (IE) is a life-threatening disease with a short-term mortality stubbornly exceeding 20% [1,2,3]. In the randomized controlled trial study "Partial Oral versus Intravenous Antibiotic Treatment of Endocarditis" (POET), 400 patients with left-sided IE fulfilling the modified Duke criteria for definite endocarditis were enrolled [4,5]. Half of the patients randomly assigned to continued conventional intravenous treatment, and the other half of the patients to a shift to oral treatment after fulfillment of criteria for stabilization. Included were patients with IE with an etiology of streptococci (49%), *Enterococcus faecalis* (24%), *Staphylococcus aureus* (22%), or coagulase-negative staphylococci (6%). Non-inferiority of partly oral treatment was demonstrated [5,6]. Demographic and basic clinical data were also given. The largest group of IE involved strains thus belonged to the streptococci. Streptococcal strain identifications were performed by phenotypic tests or Matrix Assisted Laser Desorption/ionization - Time of Flight Mass Spectrometry (MALDI-ToF MS). MALDI-ToF MS being the method that have revolutionized identification of bacteria within the last decade [7]. Several studies have investigated the distribution of streptococcal species in IE. It has been found that the prevalence of IE in streptococcal blood stream infections is species dependent with *Streptococcus mutans*, *Streptococcus gordonii*, *Streptococcus sanguinis*, *Streptococcus gallolyticus*, and *Streptococcus mitis/oralis* having the highest IE prevalence and the highest associated IE risk after adjusting for IE risk factors [8]. However, sometimes the results may be difficult to compare due to the different methods for species identification used [9]. Recent developments in molecular methods, especially whole genome sequencing (WGS) in combination with bioinformatics, have opened up for strongly improved identification and characterization of bacterial strains. Genome sequences are now available for most bacteria associated with human infections, making whole genome comparisons the ultimate tool for taxonomic studies of bacteria [10]. WGS identification of streptococci from IE episodes in order to expand number of cases characterized in detail may in the future add to clinical important recognitions.

Streptococci are a heterogeneous group of bacteria, which has undergone major taxonomic changes within the last decades, now consisting of at least 135 species. They consist of as well species showing beta-hemolytic actvity when grown on blood agar, the pyogenic group of streptococci, and a larger group of species without beta-hemolytic growth, the viridans group of streptococci, though exceptions with respect to hemolytic activity among strains of the different groups do occur [11]. In addition to the most often pyogenic group species (*Streptococcus pyoges* (group A), *Streptococcus agalactiae* (group B) and *Streptococcus dysgalactiae* (group G or C), several more seldom recognized species exists (at least six additional species). Viridans group strains are divided into six major groups; these being the mutans, salivarius, anginosus, mitis, sanguinis, and bovis groups [11]. In 1997, only 15 different streptococcal species were included in the viridans groups strains indicating the rapid taxonomic development within this group of bacteria. Identification by phenotypic methods, 16S rRNA gene sequence analyses and the, within the last

decade increasingly used method for bacterial identification, MALDI-ToF MS have been challenged when applied on streptococci. However, the latter being increasingly optimized and securely allocates strains to pyogenic and viridans groups [9,12,13]. The group of streptococci, especially viridans groups strains, are known for being taxonomically challenging [10,14], wherefore the application of detailed characterizations may open up for new recognitions. Detailed strain identification may give information on suspected pathogenesis, where to look for primary infection foci, characteristic disease manifestations, clinical course/complications, prognosis – all of importance for planning investigation- and treatment strategy [8,9].

We therefore found it of interest further characterizing etiological streptococcal strains from the POET trial with up-to-date molecular methods. One hundred and twenty-three strains from 117 patients (69% of all POET streptococcal strains from a circumscribed geographical area) were whole genome sequenced with the purposes of extended species identification, genome characterization, gain insight into their relative frequency and relate to clinical characteristics. Data on general demography, valve involvement, preexisting prosthesis, valve surgery during current disease related to treatment group (intravenous or partly peroral antibiotic treatment) and composite outcome (all-cause mortality, unplanned cardiac surgery, embolic events, or relapse of bacteremia with the primary pathogen outcome) were included.

MATERIALS AND METHODS

Microbiological examinations

Strains. One hundred and twenty-three streptococcal strains isolated from 117 patients enrolled in the POET study were examined [4]. All available streptococcal strains from patients from three of five regions of Denmark (The Capital Region, Region of Southern Denmark and Region Zealand) were included.

Briefly about the POET study: The trial (2011 to 2017) was a nationwide investigator-initiated, multicenter, randomized, un-blinded, non-inferiority trial performed at cardiac centers in Denmark. Only stable patients with a satisfactory clinical response to initial intravenous treatment, no signs of disease progression were eligible and where an oral antibiotic regime could be provided. A minimum of 10 days of intravenous antibiotics treatment (and for patients who had undergone valve surgery at least 7 days of intravenous antibiotics after surgery,) was required prior to randomization. Orally treated patients were offered out-patient treatment. In addition, transesophageal echocardiography performed before randomization had to show no signs of abscess formation or valve abnormalities that would require surgery. The trial found partial-oral outpatient antibiotic treatment to be non-inferior to conventional hospitalized intravenous antibiotic treatment on hard-endpoints [4].

Strain examinations. At inclusion in the POET trial streptococcal strains were identified with routine MALDI-ToF MS, either MALDI Biotyper (r) (Bruker Daltonik) or VITEK(r) MS (bioMerieux) setup, at the respective regional departments of clinical microbiology. Supplementary phenotypic tests were applied according to local guidelines.

WGS data for the following strains were downloaded from NCBI for comparative phylogenetic analysis.

Streptococcus equinus ATCC 9812^T, *Streptococcus vestibularis* ATCC 49124^T, *Streptococcus intermedius* NCTC 11324 ^{T}, *Streptococcus constellatus subsp constellatus* ATCC 27823 ^{T}, *Streptococcus agalactiae* NCTC 8181 ^{T}, *Streptococcus oralis subsp tigurinus* az 3a ^{T}, *Streptococcus oralis subsp dentisani* CSISP 7747^T, *Streptococcus gallolyticus subsp pasteurianus* NCTC 13784 ^{T}, *Streptococcus infantarius subsp infantarius* CCUG 43820 ^{T}, *Streptococcus sanguinis* NCTC7863^T, *Streptococcus salivarius* NCTC 8618 ^{T}, *Streptococcus pyogenes* NCTC 8198 ^{T}, *Streptococcus mutans* NCTC 10449 ^{T}, *Streptococcus lutetiensis* NCTC 13774 ^{T}, *Streptococcus gordonii* ATCC 10558 ^{T}, *Streptococcus gallolyticus subsp gallolyticus* ATCC 43143 ^{T}, *Streptococcus dysgalactiae* NCTC 8543 ^{T}, *Streptococcus anginosus* NCTC10713 ^{T}, *Streptococcus cristatus* ATCC 51100 ^{T}, *Streptococcus mitis* NCTC 12261^T, *Streptococcus oralis* CCUG13229^T, *Streptococcus pneumoniae* NCTC 7465^T, *Streptococcus gallolyticus subsp macedonicus* aca-dc198, *Streptococcus alactolyticus* bl-178-wt-3a, *Abiotrophia defectiva* ATCC 49176 ^{T}.

DNA extraction, sequencing and assembley. DNA was extracted using the Nextera™ DNA Flex Microbial Colony Extraction protocol (Illumina, San Diego, USA). Library preparation was made using the Nextera™ DNA Flex Library Preparation kit (Illumina, San Diego, USA), they were sequenced on an Illumina Miseq (Illumina, San Diego, USA) generating 150 bp. paired end reads and quality assured using the CLC Genomics Workbench (Qiagen, Denmark). The raw data were assembled using SPAdes v. 3.13.0. and assembly metrics were calculated using QUAST 5.0.2 [15].

Core genome analysis and phylogenetic tree construction. For core genome analysis genes

were predicted and translated into amino acid sequences using prodigal v. 2.6.2. The Bacterial Pan Genome Analysis tool (BPGA) version 1.3 (Using USEARCH with sequence identity cut-off 0.5) was used to find the genes present in all strains (the core genome) used in this study. Genes were aligned using MUSCLE v. 3.8.425. To ensure homology all genes with less than 35% conserved sites were excluded from the analysis. All the remaining core genes were concatenated for further analysis. A tree based on the amino acid sequences was built using PhyML [16]. The tree was visualized using the online tool Interactive Tree of Life [17].

CSI phylogeny. The Single-nucleotide polymorphism (SNP) based CSI phylogeny web application (https://cge.cbs.dtu.dk/services/CSIPhylogeny) with default settings was applied for species identification. In short, CSI phylogeny maps each contig against a chosen reference genome using BWA v. 0.7.2. Afterwards SNPs are called using SAMTools v. 0.1.18, leaving out all SNPs in a ten-base vicinity of each other. CSI phylogeny returns a newick file format for tree visualization. All strains tested using their assembled genomes and sequences were uploaded in a fasta format. We used the *S. pneumoniae* NCTC 7465 as the reference genome.

Average Nucleotide Identity (ANI). ANI was calculated using the tool fastANI v. 1.32 [18]. All strain pairs were tested using the "many to many" method in fastANI and by using the "--matrix" option results were obtained as a phylip-formatted lower triangular matrix".

Table 1. Demographic, clinical and paraclinical data on IE patients included in this study (n = 116*)

Peroral treatment (n (%)):	60 (51.7)
Age in years (mean (SD)):	63.91 (12.47)
Male sex (n (%):	94 (81.0)
BMI (mean (SD)):	26.16 (4.57)
Coexisting condition or risk factor (no. (%)):	
Diabetes	15 (12.9)
Renal failure	6 (5.2),
Dialysis	1 (0.9)
COPD	7 (6.0)
Liver disease	2 (1.7)
Cancer	6 (5.2)
Intravenous drug use	0 (0)
Laboratory results at randomization:	
Hemoglobin — mmol/liter (mean (SD))	6.30 (5.70, 7.00)
Leukocytes — ×10−9/liter (mean (SD))	7.26 (5.97, 8.42)
C-reactive protein — mg/liter (mean (SD))	20.50 (10.00, 34.25)
Creatinine — µmol/liter (mean (SD))	83.00 (67.00, 98.00)
Preexisting prosthesis, implant, or cardiac disease — no. (%):	
Prosthetic heart valve	30 (26.1)
Pacemaker	8 (6.9)
other known valve disease	63 (54.3)
Cardiac involvement at randomization — no. (%):	
Mitral-valve endocarditis	39 (33.6)
Aortic-valve endocarditis	59 (50.9)
Mitral-valve and aortic-valve endocarditis	16 (13.8)
Endocarditis in other locations	2 (1.6)*
Pacemaker endocarditis	3 (2.4)
Valve surgery during current disease	56 (48.3)
6 months mortality - n (%):	4 (3.4)
5-year mortality – n (%):	25 (21.6)
6 months primary outcome** - n (%)	10 (8.6)
Long term primary outcome – n (%)	39 (33.6)

* One patient infected with as well *S. aureus* is not included. One patient had an infected ventricular septal defect, and one patient had an infected myxoma in the left atrium.
**The primary composite outcome was all-cause mortality, unplanned cardiac surgery, embolic events, or relapse of bacteremia with the primary pathogen.

Data availability. The sequences of the 123 sequenced strains during this study deposited in the NCBI database Bioproject number PRJNA808899 with accession numbers SAMN26134978 - SAMN26135100.

Demographic and clinical data on patients. Information's were retrieved from the data obtained in relation to the original publication [4] and the long term (5.4 year) follow up publication [6].

RESULTS

Overall demographic data. In Table 1 some overall demographic, clinical and paraclinical data on the 116 patients (one patient having as well *S. aureus* left out) included in this study are given. In supplementary Table S1 a comparison to the additional 80 streptococcal IE cases included in the POET Trial from whom no strains were available is given with only statistic difference being regarding to age)

Strains. One hundred and twenty three streptococcal strains (one of which was an *Abiotrophia defectiva* strain) were examined. Six of the strains were from 4 patients experiencing more IE episodes (one patient experiencing 3 additional IE episodes (with *S. mitis*, *S. salivarius*, *S. sanguinis*) and 3 patients experiencing 1 additional IE episode (*S. dysgalactiae, S. gordonii*, *S. mitis*). One hundred and seventeen strains were used for the initial inclusion in the POET study from the three regions (see M&M) out of 172 possible (the remaining strains missing caused by a freezer breakdown). In the POET trial 196 of 400 IE causative strains were grouped as streptococcal strains. Thus, 68% of strains from the 3 regions and 59% of all streptococcal strains from the trial were included.

Overall strain characterizations. For distribution of strains included in main and subgroups of *Streptococcus* species see Table 2. As well identifications present at initial inclusion in the POET trial based on MALDI-ToF MS and phenotypic characteristics as identifications obtained with genomic examinations based on whole genome sequences are given. Twelve percent of strains were pyogenic group strains and 88% were viridans groups strains. Of the viridans groups strains 50% belonged to the mitis group, 20% to the bovis group, 8% to the mutans group and, 5% belonged to the salivarius- or the anginosus groups. A single included strain was identified as *Abiotrophia defectiva*. At entry in the POET trial pyogenic group strains were mostly grouped according to serogroup reactivity whereas by genomic examination definite species identifications were obtained. Of the viridans groups strains, approximately one third of the mitis- and 75% of the bovis group strains were identified to the group level at trial entry and for the remaining viridans groups strains the percentages of group level identifications were even higher. When identified based on genomic examinations all viridans groups strains achieved definitive identifications. Mitis group and bovis group streptococci dominated among the viridans groups strains. Among the mitis group strains *Streptococcus oralis*, *Streptococcus sanguinis* and *Streptococcus gordonii* occurred most often; the *S. oralis* strains could be sub-grouped of which *ssp. tigurinus* was most frequent though also strains belonging to *ssp oralis* and *ssp dentisani* were identified. Among bovis group streptococci the species *Streptococcus gallolyticus* was by far the most often recognized species with *ssp. gallolyticus* as the dominating subspecies. Of the mutans group and anginosus group strains all belonged respectively to *S. mutans* and *S. anginosus*.. Among the six salivarius group strains, respectively four and two were identified as *S. salivarius* and *S. vestibularis*. One strain was identified as *A. defectiva* and was not included in the core-genome-phylogeny, CSI phylogeny and ANI examinations. The use of genomic characterization thus improved the species assignment.

Table 2. Initial identifications of included streptococcal strains (n=123*) and after detailed molecular characterization based on whole genome sequences. In the two following rows Average Nucleotide Identity (ANI) of strains to their species type/representative strain are given.

	No. of strains/% of all strains	No. of IDs** at POET entrance	No. of IDs by molecular identification	ANI% to type or reference strain: Mean % (range %)	ANI % to related species (most related species, other)*****
Data for all strains included	123/100		123		
Streptococcal groups					
Pyogenic group	15/12		15		
S. pyogenes (group A)		2	2	98.5	<90 (S. dys), <84
S. agalactiae (group B)		5	5	98,3 (98.1-98.5)	<84
S. dysgalactiae (group C or G)		8	8	99 (98.9-99.5)	<90 (S. pyo), <83
Non-hemolytic streptococci		8			
S. mitis group	61/50	24	61		
S. mitis		0	4	94.0 (92.7-95.0)	<91.7 (S. pn), ≤87
S. oralis		7	29	93.3 (91.2-95.2)	<87
ssp. oralis			(10)	94.8 (94.3-95.2)	≤93.5
ssp. dentisani			(6)	94.1 (91.1-96.6)	≤87
ssp. tigurinus			(13)	94.0 (93.0-94.7)	≤87
S. sanguinis		11	12	94.9 (94.0-96.6)	≤86
S. parasanguinis		0	1	96.7	≤86
S. cristatus		0	2	91.5 (88.0-95.1)	≤86
S. gordonii		11	10	95.8 (95.7-96.2)***	≤86
S. pneumoniae		3	3	98.6 (98.6-98.7)	<92 (*S. mitis*) ,≤86
S. bovis group	24/19	18	24		
S. gallolyticus		4	21		
ssp. gallolyticus			(18)	98.7 (97.2-99.9)	<95.9 (SGSP), <86
ssp. pasteurianus			(3)	99.1 (99.0-99.2)	<96.2 (SGSG), <86
S. infantarius		0	2	98.4 (98.3-98.4)	<94 (S. lut), <86
S. lutetiensis		1	1	98.7	<94.0 (SISI), <86
S. mutans group	10/8	10	10		
S. mutans			10	98.9 (98.7-99.2)	<79
S. salivarius group	6/5	6	6		
S. salivarius			4	95.8 (95.5-96.2)	<92.2
S. vestibularis			2	98.6 (98.3-99.0)	<92.1
S. anginosus group	6/5	5	6		
S. anginosus			6	95.1 (94.9-95.2)****	≤91 (S. con), ≤88 (S. int)
Miscellaneous	1/<1		1		
Abiotrophia defectiva		1	1	96.1	ND

*123 strains from 117 patients; respectively one and 3 patients had 4 and two episodes with new taxons.
ID: identification; * *S. gordonii*: POET strains had ANI`s to each other >96% except for 2 strains.; **** *S. anginosus*: POET strains had ANI`s to each other >97%.
***** Taxon abbreviations: S. dys : *S. dysgalactiae*; S.pyo*: S.pyogenes*; S. pn: *S. pneumoniae*; SGSP: *S. gallolyticus ssp pasteurianus*¸ SGSG*: S. gallolyticus ssp gallolyticus;* S.lut*: S. lutetiensis; SISI: S. infantarius ssp infantarius*; S. con: *S. constellatus;* S. int: *S. intermedius*

In Supplementary Table S2 data on no. of genomes, no. of contigs, guanine plus cytosine percentage content (GC%), N50, genome size (average (range)) and no. of genes recognized are given. Overall, the mean no. of contigs varied from 14 to 137 with varying no. within species and groups. The mean GC% varied between 35.39 to 43.2 (the single strain of *A. defectiva* having a GC% of 46.9. The mean N50 varied between 44097 and 998524, the genome size between 1.82 and 2.35 million basepair and no. of genes between 1771 and 2181.

Core-genome and CSI phylogeny. Core-genome phylogeny of the 122 clinical *Streptococcus* strains and 25 downloaded type/collection strain WGS sequences was based on 280 translated common core-genes (corresponding to 87565 amino acids). CSI phylogeny was based on 22530 positions found in all analyzed genomes. Fig. 1A, 1B shows the phylogenetic tree obtained for the core genome analysis. In Fig. 1A the mitis group of strains formed two major branches. One including strains of *S. oralis, S. mitis* and *S. pneumoniae* and one including *S. sanguinis*, *S. cristatus* and *S. gordonii*. Inclusion of the WGS sequences of the typestrains of *S. oralis subspecies oralis*, *ssp dentisani* and *ssp tigurinus* resulted in quite clear separation of the strains into subspecies clusters. The same trend was seen when doing CSI phylogeny (Fig S1), though the *S. pneumoniae* branch being more separate than when doing core-genome phylogeny (data not shown). By CSI phylogeny the bovis group of strains grouped into subspecies groups, whereas subspecies grouping into *ssp gallolyticus* and *ssp pasteurianus* when using core-genome phylogeny was less

distinct. However, when doing core-genome phylogeny on only bovis group strains (1076 genes, alignment length of 348.506 amino acids) strains were separated correctly into groups (Fig. 2). Both analysis methods separated mutans, salivarius and anginosus group strains. In the pyogenic streptococcal group *S. dysgalactiae subspecies equisimilis* strains grouped closest to the *S. pyogenes* strains with more distance to *S. agalactiae* strains.

Figure 1 (top and bottom) Also available here

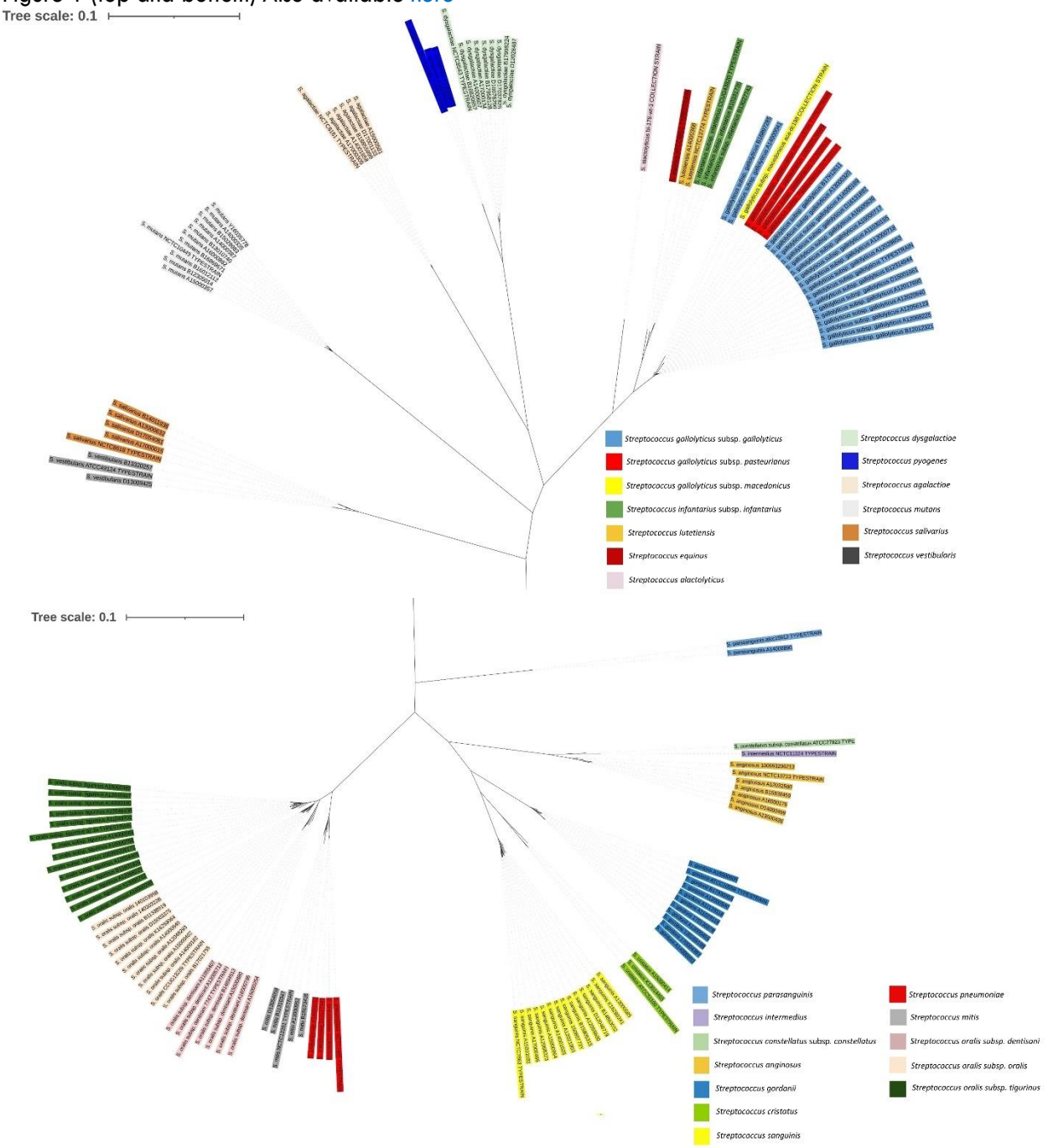

Figure 1. Core genome analysis of 122 *Streptococcus* strains obtained from patients included in the POET study [7] and 25 downloaded type/collection strain whole-genome sequences based on 280 translated common core-genes (corresponding to 87565 amino acids). In the bottom part of the tree, the mitis group and anginosus group strains are presented. In the top part of the tree, bovis, mutans and salivarius group strains together with the pyogenic streptococcal species *Streptococcus pyogenes*, *Streptococcus, agalactiae* and *S. dysgalactiae* are presented.

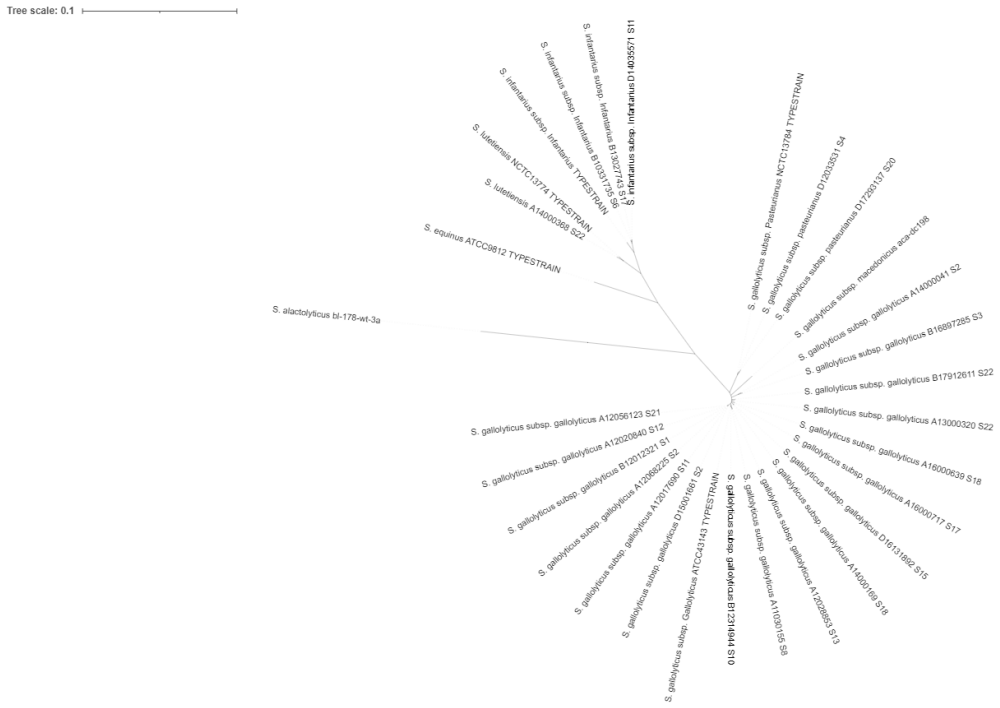

Figure 2. Also available here: Core genome analysis of 24 bovis group strains obtained from patients included in the POET study (ref) and four downloaded type/collection strain whole-genome sequences based on 1076 genes, alignment length of 348.506 amino acids.

ANI: Mean and range of ANI values of strains to type/reference strains of the different species are listed in Table 2. The ANI percentages for the pyogenic group streptococci were all >98% and *S. pyogenes* (group A) and *S. dysgalactiae* (group C or G, all belonging to *subsp. equisimilis*) were the closest related among these. In the mitis group of streptococci only the three *S. pneumoniae* strains exhibited ANI percentages to their own species type-strain above 98%. The remaining species showed ANI percentages between 91.5% and 95.8% with, importantly, no overlap in ANI percentages to other mitis group species. The bovis group strains, dominated by *S. gallolyticus subspecies gallolyticus* strains, had ANI`s above 98% with closest relationship of *S. gallolyticus subspecies gallolyticus* to *S. pasteurianus* and *S. lutetiensis* to *S. infantarius*, but without ANI percentage overlaps between species. For mutans group strains ANI`s above 98% to the typestrain of *S. mutans* were recorded. Of the salivarius group strains, three each had highest ANI percentages to respectively the typestrain of *S. salivarius* (ANI percentages above 95%) and *S. vestibularis* (ANI percentages above 98%) without ANI percentage overlaps. For anginosus group strains ANI percentages were 94.9-95.2; ANI percentages to the type strains of respectively *S. constellatus* and *S. intermedius* were lower than to the type strain of *S. anginosus*.

Clinical data related to streptococcal strain IE cases.

In Table 3 patient information on preexisting prosthesis and IE valve involvement are given in relation to left-sided IE causing streptococcal species. Respectively 39, 60 and 16 IE cases involved mitral, aortic or both valves (in one patient as well mitral and aorta as the tricuspid valves were involved). Among the IE cases caused by pyogenic group strains, mitral valves most frequent affected while when caused by mitis or bovis groups streptococci aortic valves were most frequent affected though as well mitral as aortic valves were found affected. Taxons most often related to native valve IE (n=86) were *S. oralis* (25 of 28 strains), *S. gordonii* (9 of 9 strains), *S. sanguinis* (9 of 11 strains) and *S. mutans* (8 of 10 strains). Thirty-one patients (26%) had a preexisting prosthesis; 4 patients having pyogenic group strains as causative agents and 27 patients having viridans groups strains as causative agents. Among the viridans groups strain conditioned IE cases 8 of 57 (14%) with mitis group strains (evenly distributed among mitis group species), 12 of 24 (50%) with bovis group strains (10/21 (48%) with *S. gallolyticus* strains and 8 of 18 (44%) being *subspecies gallolyticus* strains) and 2 of

10 (20%) with mutans group strains had a preexisting prosthesis.

Fifty-six patients had valve surgery done during the current disease; respectively 4 and 52 patients having pyogenic group strains and viridans groups strains as causative agents. Among the viridans groups strains conditioned IE cases 37 of 52 (63%) were with mitis group strains (evenly distributed among mitis group species), 6 of 24 (25%) with bovis group strains (all with *S. gallolyticus* strains), 3 of 10 (33%) with mutans group strains and 3 of 6 (50%) with anginosus group strains. Of those patients allocated to intravenous or intravenous followed by peroral antibiotic treatment, respectively 26 of 56 and 30 of 61 had valve surgery done during the current disease.

The long term (after 5.4 year follow up) composite outcome (all-cause mortality, unplanned cardiac surgery, embolic events, or relapse of bacteremia with the primary pathogen) in the only intravenous and the intravenous plus peroral treated groups were respectively 22 and 17 events.

Table 3. Some clinical data on 117 IE patients included in the POET trial related to identification of causative streptococcal strains. Data on infected left-sided heart valve(s), preexisting prosthesis, valve surgery during current disease (and treatment group) and primary outcome (see M&M) are given.

	No. of IDs by molecular identification	IE: Mitral/aortic/both	Preexisting prosthesis	Valve surgery during current disease: Mitral/aorta/both	Valve surgery Treatment group: Intravenous/peroral	Outcome: Intravenous/peroral
Data for all strains/patients included	117	39/60/16*	31	56 (16/30/8)*	26/30	22/17
Streptococcal groups						
Pyogenic group	14	9/4/1	4	4 (2/2/0)	3/2	4/1
S. pyogenes (serogroup A)	2	1/1/0	1	1/0/0	1/0	0
S. agalactiae (serogroup B)	5	3/2/0	1	0/1/0	0/1	1/1
S. dysgalactiae (serogroup C or G)	7	5/1/1	2	1/1/0	2/1	3/0
Viridans groups streptococci	103	30/56/15	27	52 (14/28/8)*	23/28	18/16
S. mitis group	57	17/30/6**	8	10/23/4	17/19	8/10
S. mitis	2	0/2/0	1	0/1/0	0/1	0/2
S. oralis	29	12/13/3**	3	5/8/1	6/8	5/5
ssp. oralis	(10)	(4/5/1)	(1)	(1/2/0)	(1/3)	(0/4)
ssp. dentisani	(6)	(3/3/0)	(1)	(2/2/0)	(3/1)	(2/0)
ssp. Tigurinus	(13)	(4/6/2)**	(1)	(2/4/0)*	(3/4)	(3/1)
S. sanguinis	11	2/5/4	2	1/5/3*	4/4	1/0
S. parasanguinis	1	0/1/0	1	0/1/0	1/0	1/0
S. cristatus	2	0/2/0	0	0/2/0	2/0	0
S. gordonii	9	4/5/0	0	3/4/0	3/4	1/1
S. pneumoniae	3	1/2/0	1	1/2/0	1/2	0/2
S. bovis group	24	6/16/2	12	1/3/2	2/4	8/4
S. gallolyticus	21	6/13/2	10	1/3/2	-	6/3
ssp. gallolyticus	(18)	(5/11/2**)	(8)	(1/2/2)	(2/3)	(5/3)
ssp. pasteurianus	(3)	(1/2/0)	(2)	(0/1/0)	(0/1)	(1/0)
S. infantarius	2	0/2/0	1	0	-	1/1
S. lutetiensis	1	0/1/0	1	0	-	1/0
S. mutans group	10	4/3/3	2	1/1/1	2/1	0
S. mutans	10	4/3/3	2	(1/1/1)	2/1	
S. salivarius group	5	1/2/2	3	0/0/2	1/1	1/1
S. salivarius	3	0/1/2	1	(0/0/2)	1/1	0/0
S. vestibularis	2	1/1/0	2	0	-	1/1
S. anginosus group	6	2/3/1	2	2/1/0	1/2	1/1
S. anginosus	6	2/3/1	2	(2/1/0)	1/2	1/1
Miscellaneous	1	0/0/1	0	0/0/1	0/1	0/0
Abiotrophia defectiva	1	0/0/1	0	(0/0/1)	0/1	0/0

* One IE involving 3 valves (mitral, aortic and tricuspid)

** Endocarditis in other locations: One patient had an infected ventricular septal defect, and one patient had an infected myxoma in the left atrium.

DISCUSSION

In a recent Danish study in the Capital Region of Denmark on streptococcal blood stream infections from 2008 to 2017, the IE prevalence was 7.1% among 6506 cases [8]. The variation in IE prevalence within streptococcal groups was substantial and concluded that the risk of IE being evaluated on species level. All species identifications based on standard microbiological methods (phenotypic methods) and MALDI-ToF MS from 2010 as the primary method. Blood stream infections with *S. mutans*, *S. gordonii*, *S. sanguinis*, *S. gallolyticus*, and *S. mitis/oralis* having the highest IE

prevalence and the highest associated IE risk after adjusting for IE risk factors [9,19].

Our study focused on benefits in species designation by use of whole genome sequence analysis on a collection of streptococcal strains obtained from patients with left-sided IE, i.e. from the previously published POET study [4,5]. Frequently streptococcal strains are not species identified, but grouped as streptococci or grouped in few categories [8]. Previous, greatly being a consequence of difficulties in making correct species identification of the frequent recognized mitis group streptococci. To the best of our knowledge this is the largest collection of strains from IE episodes characterized by WGS analysis. In recent decade(s), 16S rRNA gene sequence analysis and especially in the last decade MALDI-ToF MS have revolutionized taxonomy and identification opportunities of bacteria. MALDI-ToF MS identifications have been steadily improved in the last 10 years and securely allocates viridans groups strains to the group level and not seldom beyond [10,13,19,20]. Looking, with identification based on sequence determination of the *rnpB* gene and MALDI-ToF MS, at 63 blood culture strains of viridans groups strains from patients with IE, Isaksson et al [13] found agreement for all 36 strains identified in the anginosus, bovis, and mutans groups or identified as *S. cristatus*, *S. gordonii*, or *S. sanguinis*. Twenty-three strains belonging to the species *S. mitis*, *S. oralis*, or *S. tigurinus* were not reliably identified, much in agreement with the findings in this study where mitis and bovis group strains at inclusion often were identified to the group level (Table 2), but not beyond.

The development in molecular techniques and bioinformatics allowing examinations based on WGS have greatly enhanced the possibilities for taxonomic development and detailed taxonomic allocation of strains [10]. Identification of pyogenic group strains works well both by serogrouping and MALDI-ToF MS analysis as also seen at study entrance in accordance with core genome analysis and CSI phylogeny results (Table 1, Fig. S1). Especially within the mitis and bovis group of streptococci WGS analysis greatly increased no. of species/subspecies identifications. The mitis group of the genus *Streptococcus* currently comprises more than 20 species with validly published names [10]. In a comprehensive study by Jensen et al. [10] the taxonomy of this group of streptococci was re-evaluated using 195 available genomes in a comparative phylogenetic analysis based on core genome sequences, MLSA and 16S rRNA gene sequences, combined with estimations of ANI and in silico and in vitro analyses of specific phenotypic characteristics. Core genomic phylogenetic analyses revealed distinct clades or branches that, to some extent, and from the clustering of type strains represented known species. *S. oralis* strains formed subclusters within a coherent phylogenetic clade and they proposed that the species *S. oralis* consists of three subspecies: *S. oralis subsp. oralis subsp. nov.*, *S. oralis subsp. tigurinus comb. nov.*, and *S. oralis subsp. dentisani comb. nov*. This subgrouping of *S. oralis* strains from IE cases was also noticed previous by Rasmussen et al. [21,22]. Previous studies have shown that ANI values of between 94–96 % correspond to DNA–DNA hybridization values of between 60–70 %, which has been generally accepted as a species boundary [23]. In this study core-genome analysis and CSI phylogeny, in agreement with Jensen et al. [10] made a clear separation into subgroups of *S. oralis* strains (Fig. 1), whereas ANI data did not definitively separate strains into subgroups (Table 1). However, strains among the different MGS species not reaching ≥96% sequence identity with their type strain had to all related species ANI values clearly separating them from these.

The Lancefield group D non-enterococcal streptococci comprise the *Streptococcus bovis/equinus* complex [24]. Previous phenotypic characterization allowed the distinction of three biotypes (mainly according to mannitol fermentation). More than a decade ago a newer classification of the former biotypes was suggested based on single gene analysis [25,26]. Biotype I, biotype II/2, and *S. macedonicus* (or *S. waius*)—a closely related phylogenetic cluster—were reclassified into *S. gallolyticus subsp. gallolyticus* (SGSG), *S. gallolyticus subsp. pasteurianus* (SGSP), and *S. gallolyticus subsp. macedonicus* (SGSM). Biotype II/1 consisted of two taxonomic units, *S. infantarius subsp. infantarius* and *S. infantarius subsp. coli* (also designated as *S. lutetiensis*). Biotype I (SGSG) has been strongly associated with IE. Examining 40 patients with *S. bovis/equinus* complex bacteremia Ben-Chetrit et al. classified eight and 14 as having, respectively, definite and probable IE [24]. Combining these data with four other studies resulted in 320 bacteremia episodes [24]. Sixty-six (21 %) had IE. Twenty-eight persent had colon pathology (pre-malignant or malignant lesions). Of these 66 strains, 31 were SGSP (16% of all SGSP strains), 29 SGSG (37% of all SGSG strains) and five *S. infantarius* (11% of all *S. infantarius* strains), respectively. In the study by Chamat-Hedemand et al. [8] looking on approximately 200 *S. gallolyticus* blood stream infections 30% presented with IE. Most strains in our

study were SGSG strains, though the other bovis/equinus taxons were also included. Among our patients with bovis group IE episodes only two patients had known cancer (unpublished data). The importance of amount of data included when making phylogeny was illustrated for the bovis/equinus strains in this study. Core genome analysis phylogeny based on translated genomes common for all strains (n = 280) did not clearly separate *subspecies gallolyticus* strains from the type strain of *subspecies macedonicus* and *subspecies pasteurianus* strains (Fig 1). Based on only genes from bovis/equinus strains (n = 1076) separation was evident as it was using CSI analysis based on single-nucleotide polymorphism (Fig. 2). Strains from the mutans, salivarius and anginosus groups were characterized to their species level using core genome analysis and CSI phylogeny. Within these groups taxonomic developments also points on creation of subspecies [27].

Interestingly, a very broad spectrum of different streptococci were identified and streptococci giving close to exclusively rise to left-sided IE cases [28,29]. Although the inclusion criteria for entering the POET trial was a supposed well-controlled benign IE course, the distribution of streptococcal species involved seems in accordance with other studies with approximately 50-60% belonging to the mitis group streptococci, 25-30% to the bovis group streptococci and 10-15% being pyogenic streptococci [29,30].

Results from the POET study have been described in publications after 6 month, 3.5 and most recently 5.4 years [4,5,6]. The conclusion in the 5.4-year follow up report was that there was no indications of long-term inferiority in the oral step-down treated patients. Looking strictly on streptococcal strains support the conclusion as no difference with respect to valve surgery during current disease and composite outcome was observed between patients treated only with intravenous treatment and the partly peroral treated group. These reassuring findings further support current considerations of implementing oral step-down antibiotic therapy in selected patients with left-sided IE.

Two recent publications from Sweden have looked at cases of IE caused by beta- and alpha streptococci or *S. bovis* (reported, respectively 2008-2016 and 2008-2014) retrieved from The Swedish Registry of Infective Endocarditis [9,31]. Of 104 episodes of IE with pyogenic group A, C or G, strains from 66 of the cases were available for MALDI-ToF MS examination; 13 strains belonged to *S. pyogenes* and 50 to *S. dysgalactiae* and all except 6 strains were related to left-sided IE [31]. The distribution between mitral and aortic involvement was rather equal somewhat in contrast to our data though the included number was rather small. Comparing clinical data between pyogenic group and viridans groups strains IE episodes showed a shorter onset to hospitalization period and a more pronounced tendency to embolization for the IE episodes caused by pyogenic group strains. Of 774 episodes of IE with viridans groups strains or *S. bovis*, 201 strains were retrieved and analysed [9]. Only 5% of cases were on native right-sided valves. Aortic and mitral valves were close to equally affected, though for the bovis group of strains aortic valve affection was twice as frequent as mitral valve involvement; the same trend was seen in this study.

Most mitis group streptococcal strains from IE patients belonged to *S. oralis*, *S. sanguinis* and *S. gordonii* being in line with other reports [8]. Nilson et al. [9] demonstrates that mutans and mitis group isolates, as identified by MALDI-ToF MS, were overrepresented in patients with IE whereas the anginosus group is more common in all-cause bacteremia than in IE. This finding is in line with previous studies [8] and underlines the fact that different streptococci have different propensities to cause IE. This may relate to the fact that mitis and mutans groups are members of the mouth flora rather than the intestinal flora, though differences in molecular virulence mechanisms may also play a role [20]. The association of certain streptococcal groups with IE may help clinicians to determine which patients with streptococcal bacteremia should be referred to transesophageal echocardiography. Species identifications may also give information on which source of infection to suspect related to their natural residence though not seldom no clinical suspected focal source is recognized. *S. anginosus* is known for its ability to cause abscess formation and it has to be remembered that *S. pneumoniae*, though more related to pneumonia, septicemia and meningitis are recognized as a IE causing species as also in this study [32].

In our study patients with bovis group IE episodes were characterized by often having preexisting prosthetic valve at onset. This being in agreement with Nilson et al 2016 who found that patients infected with bovis group isolates were older, had more cardiac devices, and had more commonly prosthetic valve IE compared to IE caused by streptococci of the other groups.

Strength, limitations and conclusion.

Identifications based on deep molecular characterization revealed presence of great variety of different streptococci including recently

defined subspecies especially within mitis and bovis groups. The strain distribution was influenced by the clinical criteria for inclusion into the POET trial and that not 100% but 68% of strains from the well-defined geographical area (3 regions in Denmark) were available. Detailed species identification may be helpful with respect to investigation planning. Though being the study with largest number of WGS examined streptococcal strains from IE patients, the number of found species/subspecies calls for inclusion of more data on strains and patients in order to reveal important clinical relations.

In conclusion: Routine identification of streptococci by MALDI-ToF MS allocates strains to helpful streptococcal species/groups. Molecular examinations substantially adds on by detailing species and subspecies affiliations. A broad spectrum of streptococcal species and subspecies causing IE were identified with mitis- and bovis group strains dominating. Allocation of streptococcal strains to species/groups is of importance for planning investigation- and treatment strategy. Clinical presentation and microbiological identification may support each other in handling of IE confirmed/suspected patients. Our results suggests for adding WGS diagnosis of streptococci in selected patient groups (e.g. IE) in order to expand number of cases characterized in detail and advocates for centralized registration of results to reveal important clinical relations.

Conflicts of Interest Statement: "The authors have no conflicts of interest to declare."

Funding Statement: This study is a substudy of the POET trial, which was supported by unrestricted grants from the Danish Heart Foundation, the Capital Region Research Council, the Hartmann's Foundation, Svend Aage Andersens Foundation, and the Novo Nordisk Foundation (Borregaard Clinical Scientist Fellowship in translational research; grant no. NNF17OC0025074).

Acknowledgments: None

REFERENCES

1. Allen CJ, Klein JL, Prendergast BD. Streptococcal Infective Endocarditis: 'On the Origin of Species'. *Circulation*. 2020;142(8):731-733. doi:10.1161/CIRCULATIONAHA.120.049055
2. Pant S, Patel NJ, Deshmukh A, et al. Trends in infective endocarditis incidence, microbiology, and valve replacement in the United States from 2000 to 2011. *J Am Coll Cardiol*. 2015;65(19):2070-2076. doi:10.1016/j.jacc.2015.03.518
3. Cresti A, Chiavarelli M, Scalese M, et al. Epidemiological and mortality trends in infective endocarditis, a 17-year population-based prospective study. *Cardiovasc Diagn Ther*. 2017;7(1):27-35. doi:10.21037/cdt.2016.08.09
4. Iversen K, Ihlemann N, Gill SU, et al. Partial oral versus intravenous antibiotic treatment of endocarditis. *N Engl J Med*. 2019;380(5). doi:10.1056/NEJMoa1808312
5. Bundgaard H, Ihlemann N, Gill SU, et al. Long-term outcomes of partial oral treatment of endocarditis. *N Engl J Med*. 2019;380(14). doi:10.1056/NEJMc1902096
6. Pries-Heje MM, Wiingaard C, Ihlemann N, et al. Five-Year Outcomes of the Partial Oral Treatment of Endocarditis (POET) Trial. *N Engl J Med*. 2022;386(6):601-602. doi:10.1056/NEJMc2114046
7. Tsuchida S, Umemura H, Nakayama T. Current Status of Matrix-Assisted Laser Desorption/Ionization–Time-of-Flight Mass Spectrometry (MALDI-TOF MS) in Clinical Diagnostic Microbiology. *Molecules*. 2020;25(20). doi:10.3390/molecules25204775
8. Chamat-Hedemand S, Dahl A, Østergaard L, et al. Prevalence of Infective Endocarditis in Streptococcal Bloodstream Infections Is Dependent on Streptococcal Species. *Circulation*. 2020;142(8):720-730. doi:10.1161/CIRCULATIONAHA.120.046723
9. Nilson B, Olaison L, Rasmussen M. Clinical presentation of infective endocarditis caused by different groups of non-beta haemolytic streptococci. *Eur J Clin Microbiol Infect Dis Off Publ Eur Soc Clin Microbiol*. 2016;35(2):215-218. doi:10.1007/s10096-015-2532-5
10. Jensen A, Scholz CFP, Kilian M. Re-evaluation of the taxonomy of the Mitis group of the genus Streptococcus based on whole genome phylogenetic analyses, and proposed reclassification of Streptococcus dentisani as Streptococcus oralis subsp. dentisani comb. nov., Streptococcus tigurinus. *Int J Syst Evol Microbiol*. 2016;66(11):4803-4820. doi:10.1099/ijsem.0.001433
11. Facklam R. What Happened to the Streptococci : Overview of Taxonomic and Nomenclature Changes. *Clin Microbiol Rev*. 2002;15(4):616-630. doi:10.1128/CMR.15.4.613
12. Harju I, Lange C, Kostrzewa M, Maier T, Rantakokko-Jalava K, Haanperä M. Improved Differentiation of Streptococcus pneumoniae and Other S. mitis Group Streptococci by MALDI Biotyper Using an Improved MALDI Biotyper Database Content and a Novel Result Interpretation Algorithm. *J Clin Microbiol*. 2017;55(3):914-922. doi:10.1128/JCM.01990-16
13. Isaksson J, Rasmussen M, Nilson B, et al. Comparison of species identification of endocarditis associated viridans streptococci using rnpB genotyping and 2 MALDI-TOF systems. *Diagn Microbiol Infect Dis*. 2015;81(4):240-245. doi:10.1016/j.diagmicrobio.2014.12.007
14. Kilian M, Riley DR, Jensen A, Brüggemann H, Tettelin H. Parallel evolution of Streptococcus pneumoniae and Streptococcus mitis to pathogenic and mutualistic lifestyles. *MBio*. 2014;5(4):e01490-14. doi:10.1128/mBio.01490-14
15. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072-1075. doi:10.1093/bioinformatics/btt086
16. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59(3):307-321. doi:10.1093/sysbio/syq010
17. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. 2019;47(W1):W256-W259.

doi:10.1093/nar/gkz239

18. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*. 2018;9(1):5114. doi:10.1038/s41467-018-07641-9
19. Su T-Y, Lee M-H, Huang C-T, Liu T-P, Lu J-J. The clinical impact of patients with bloodstream infection with different groups of Viridans group streptococci by using matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS). *Medicine (Baltimore)*. 2018;97(50):e13607. doi:10.1097/MD.0000000000013607
20. Jensen CS, Dargis R, Shewmaker P, Nielsen XC, Christensen JJ. Identification of Streptococcus pseudopneumoniae and other mitis group streptococci using matrix assisted laser desorption/ionization - time of flight mass spectrometry. *Diagn Microbiol Infect Dis*. 2021;101(3):115487. doi:10.1016/j.diagmicrobio.2021.115487
21. Rasmussen LH, Højholt K, Dargis R, et al. In silico assessment of virulence factors in strains of streptococcus oralis and Streptococcus mitis isolated from patients with infective endocarditis. *J Med Microbiol*. 2017;66(9). doi:10.1099/jmm.0.000573
22. Rasmussen LH, Dargis R, H?jholt K, et al. Whole genome sequencing as a tool for phylogenetic analysis of clinical strains of Mitis group streptococci. *Eur J Clin Microbiol Infect Dis*. 2016;35(10). doi:10.1007/s10096-016-2700-2
23. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A*. 2009;106(45):19126-19131. doi:10.1073/pnas.0906412106
24. Ben-Chetrit E, Wiener-Well Y, Kashat L, Yinnon AM, Assous M V. Streptococcus bovis new taxonomy: does subspecies distinction matter? *Eur J Clin Microbiol Infect Dis Off Publ Eur Soc Clin Microbiol*. 2017;36(2):387-393. doi:10.1007/s10096-016-2814-6
25. Schlegel L, Grimont F, Ageron E, Grimont PAD, Bouvet A. Reappraisal of the taxonomy of the Streptococcus bovis/Streptococcus equinus complex and related species: description of Streptococcus gallolyticus subsp. gallolyticus subsp. nov., S. gallolyticus subsp. macedonicus subsp. nov. and S. gallolyticus subsp. *Int J Syst Evol Microbiol*. 2003;53(Pt 3):631-645. doi:10.1099/ijs.0.02361-0
26. Poyart C, Quesne G, Trieu-Cuot P. Taxonomic dissection of the Streptococcus bovis group by analysis of manganese-dependent superoxide dismutase gene (sodA) sequences: reclassification of 'Streptococcus infantarius subsp. coli' as Streptococcus lutetiensis sp. nov. and of Streptococcus bo. *Int J Syst Evol Microbiol*. 2002;52(Pt 4):1247-1255. doi:10.1099/00207713-52-4-1247
27. Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol*. 2020;70(11):5607-5612. doi:10.1099/ijsem.0.004332
28. Habib G, Lancellotti P, Erba P-A, et al. The ESC-EORP EURO-ENDO (European Infective Endocarditis) registry. *Eur Hear journal Qual care Clin outcomes*. 2019;5(3):202-207. doi:10.1093/ehjqcco/qcz018
29. Muñoz P, Kestler M, De Alarcon A, et al. Current Epidemiology and Outcome of Infective Endocarditis: A Multicenter, Prospective, Cohort Study. *Medicine (Baltimore)*. 2015;94(43):e1816. doi:10.1097/MD.0000000000001816
30. Habib G, Erba PA, Iung B, et al. Clinical presentation, aetiology and outcome of infective endocarditis. Results of the ESC-EORP EURO-ENDO (European infective endocarditis) registry: a prospective cohort study. *Eur Heart J*. 2019;40(39):3222-3232. doi:10.1093/eurheartj/ehz620
31. Bläckberg A, Nilson B, Özenci V, Olaison L, Rasmussen M. Infective endocarditis due to Streptococcus dysgalactiae: clinical presentation and microbiological features. *Eur J Clin Microbiol Infect Dis Off Publ Eur Soc Clin Microbiol*. 2018;37(12):2261-2272. doi:10.1007/s10096-018-3367-7
32. de Egea V, Muñoz P, Valerio M, et al. Characteristics and Outcome of Streptococcus pneumoniae Endocarditis in the XXI Century: A Systematic Review of 111 Cases (2000-2013). *Medicine (Baltimore)*. 2015;94(39):e1562. doi:10.1097/MD.0000000000001562

Supplementary Material

Table S1: Comparison of basic data to the additional 80 streptococcal IE cases included in the POET Trial where streptococcal strains were not available.

No of streptococcal IE episodes	80	116	p-valu
Age (mean (SD))	70.80 (11.41)	63.91 (12.47)	<0.001
Per oral treatment (%)	32 (40.0)	60 (51.7)	0.141
Male (%)	51 (63.7)	94 (81.0)	0.011
Diabetes (%)	8 (10.0)	15 (12.9)	0.689
Renal failure (%)	6 (7.5)	6 (5.2)	0.715
Dialysis (%)	4 (5.0)	1 (0.9)	0.179
COP (%)	4 (5.0)	7 (6.0)	1.000
Liver disease (%)	1 (1.2)	2 (1.7)	1.000
Cancer (%)	9 (11.4)	6 (5.2)	0.191
IV drug abuse = 0 (%)	80 (100.0)	116 (100.0)	NA
CRP (median [IQR])	16.00 [7.45, 30.25]	20.50 [10.00, 34.25]	0.164
Hemoglobin (median [IQR])	6.65 [5.88, 7.40]	6.30 [5.70, 7.00]	0.109
Creatinin (median [IQR])	89.50 [69.50, 104.75]	83.00 [67.00, 98.00]	0.215
Leucocytes (median [IQR])	6.90 [5.50, 8.53]	7.26 [5.97, 8.42]	0.423
Prosthetic valve (%)	18 (22.5)	30 (26.1)	0.687
Pacemaker (%)	7 (8.8)	8 (6.9)	0.836
Other known valve disease (%)	34 (42.5)	63 (54.3)	0.139
Mitral endocarditis (%)	39 (48.8)	39 (33.6)	0.048
Aortic endocarditis (%)	35 (43.8)	59 (50.9)	0.404
Aortic and Mitral Endocarditis (%)	6 (7.5)	16 (13.8)	0.254
Valve Surgery (%)	21 (26.2)	56 (48.3)	0.003
6 months Mortality (%)	6 (7.5)	4 (3.4)	0.349
5-year mortality (%)	25 (31.2)	25 (21.6)	0.173
6 months primary outcome (%)	8 (10.0)	10 (8.6)	0.939
5-year primay outcome (%)	34 (42.5)	39 (33.6)	0.266

Table S2: Data on no. of genomes, no. of contigs, GC%, N50, genome size (average (range)) and no. of genes recognized for 123 streptococcal strains from patients included in the POET study [7].

	no. of genomes	no. of contigs	GC (%)	N50	Total length	no. of genes
S. pyogenes	2	40.0 (35.0-45.0)	38.33 (38.3-38.36)	140662.0 (133284.0-148040.0)	1.82 (1.79-1.85)	1698 (1669-1728)
S. dysgalactiae	8	54.5 (46.0-72.0)	39.25 (39.13-39.47)	66896.0 (42385.0-83521.0)	2.1 (2.07-2.17)	1968 (1938-2023)
S. agalactaiae	5	30.2 (17.0-42.0)	35.39 (35.24-35.47)	218096.6 (102157.0-412840.0)	2.1 (2.06-2.16)	2001 (1964-2066)
S. mitis	4	160.75 (11.0-527.0)	40.05 (39.83-40.22)	145648.5 (4109.0-295621.0)	1.88 (1.78-1.99)	1814 (1689-1921)
S. oralis ssp. oralis	10	35.5 (9.0-129.0)	41.11 (40.91-41.29)	541080.5 (255257.0-1323251.0)	2.02 (1.95-2.06)	1876 (1800-1961)
S. oralis ssp. tigurinus	13	13.46 (5.0-25.0)	41.02 (40.7-41.29)	676751.77 (242515.0-1539503.0)	1.93 (1.87-1.99)	1808 (1739-1916)
S. oralis ssp. dentisani	6	16.17 (9.0-32.0)	41.09 (40.8-41.37)	459240.17 (125609.0-1154445.0)	1.9 (1.82-2.01)	1775 (1682-1869)
S. sanguinis	13	17.31 (6.0-68.0)	43.18 (42.89-43.43)	693404.38 (134581.0-1657740.0)	2.35 (2.29-2.49)	2177 (2105-2299)
S. parasanguinis	1	27.0 (27.0-27.0)	41.52 (41.52-41.52)	131928.0 (131928.0-131928.0)	2.21 (2.21-2.21)	1985 (1985-1985)
S. christatus	2	53.5 (19.0-88.0)	42.16 (42.15-42.18)	106396.5 (40333.0-172460.0)	2.07 (2.07-2.08)	1914 (1890-1937)
S. gordonii	10	85.7 (6.0-367.0)	40.39 (40.25-40.53)	485725.1 (260414.0-1115759.0)	2.22 (2.18-2.26)	2090 (2011-2255)
S. pneumoniae	3	55.33 (52.0-62.0)	39.57 (39.57-39.58)	58460.67 (44954.0-69599.0)	2.03 (2.02-2.04)	1891 (1877-1899)
S. gallolyticus ssp. Gallolyticus	18	15.44 (6.0-34.0)	37.55 (37.31-37.72)	956319.33 (112597.0-1803069.0)	2.28 (2.03-2.4)	2141 (1912-2282)
S. gallolyticus ssp. pasteurianus	3	29.0 (17.0-37.0)	37.19 (37.15-37.26)	146203.67 (87056.0-197898.0)	2.13 (2.04-2.26)	1995 (1898-2124)
S. infantarius ssp. infantarius	2	40.5 (32.0-49.0)	37.61 (37.55-37.67)	120911.5 (119111.0-122712.0)	1.9 (1.88-1.92)	1816 (1794-1839)
S. lutetiensis	1	70.0 (70.0-70.0)	37.51 (37.51-37.51)	52044.0 (52044.0-52044.0)	1.8 (1.8-1.8)	1690 (1690-1690)
S. mutans	10	18.5 (11.0-49.0)	36.71 (36.66-36.84)	256833.9 (78787.0-497011.0)	1.99 (1.95-2.04)	1824 (1797-1861)
S. salivarius	4	42.25 (32.0-56.0)	39.81 (39.74-39.87)	109832.75 (81932.0-132214.0)	2.15 (2.11-2.18)	1900 (1886-1915)
S. vestibularis	2	124.0 (65.0-183.0)	39.49 (39.37-39.62)	37391.0 (17341.0-57441.0)	1.92 (1.92-1.93)	1810 (1806-1815)
S. anginosus	6	25.33 (9.0-51.0)	38.69 (38.31-38.79)	243709.33 (122076.0-490230.0)	1.88 (1.8-2.01)	1763 (1704-1867)
S. Abiotrophia defectiva	1	14.0 (14.0-14.0)	46.9 (46.9-46.9)	473668.0 (473668.0-473668.0)	2.02 (2.02-2.02)	1763 (1763-1763)

Figure S1 is available here: CSI phylogeny analysis of whole-genome sequences from 122 *Streptococcus* strains obtained from patients included in the POET study [7] and 25 downloaded type/collection strain. Phylogeny was based on 22530 positions found in all analyzed genomes. Notice the distinct separation of the strains into species and subspecies clusters (including the bovis group of strains).