



Aalborg Universitet

AALBORG UNIVERSITY  
DENMARK

## One Day in Denmark

*Nationwide point-prevalence survey of human bacterial isolates and comparison of classical and whole-genome sequence-based species identification methods*

Rebelo, Ana Rita; Ibfelt, Tobias; Bortolaia, Valeria; Leekitcharoenphon, Pimlapas; Hansen, Dennis Schrøder; Nielsen, Hans Linde; Ellermann-Eriksen, Svend; Kemp, Michael; Røder, Bent Løwe; Frimodt-Møller, Niels; Søndergaard, Turid Snekløth; Coia, John Eugenio; Østergaard, Claus; Pedersen, Michael; Westh, Henrik; Aarestrup, Frank Møller

*Published in:*  
PLoS ONE

*DOI (link to publication from Publisher):*  
[10.1371/journal.pone.0261999](https://doi.org/10.1371/journal.pone.0261999)

*Creative Commons License*  
CC BY 4.0

*Publication date:*  
2022

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

### *Citation for published version (APA):*

Rebelo, A. R., Ibfelt, T., Bortolaia, V., Leekitcharoenphon, P., Hansen, D. S., Nielsen, H. L., Ellermann-Eriksen, S., Kemp, M., Røder, B. L., Frimodt-Møller, N., Søndergaard, T. S., Coia, J. E., Østergaard, C., Pedersen, M., Westh, H., & Aarestrup, F. M. (2022). One Day in Denmark: Nationwide point-prevalence survey of human bacterial isolates and comparison of classical and whole-genome sequence-based species identification methods. *PLoS ONE*, *17*(2), Article e0261999. <https://doi.org/10.1371/journal.pone.0261999>





### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

## RESEARCH ARTICLE

# One Day in Denmark: Nationwide point-prevalence survey of human bacterial isolates and comparison of classical and whole-genome sequence-based species identification methods

Ana Rita Rebelo <sup>1\*</sup>, Tobias Ibfelt<sup>2</sup>, Valeria Bortolaia<sup>1</sup>, Pimlapas Leekitcharoenphon<sup>1</sup>, Dennis Schröder Hansen<sup>3</sup>, Hans Linde Nielsen<sup>4</sup>, Svend Ellermann-Eriksen<sup>5</sup>, Michael Kemp <sup>6</sup>, Bent Løwe Røder<sup>7</sup>, Niels Frimodt-Møller<sup>8</sup>, Turid Snekløth Søndergaard<sup>9</sup>, John Eugenio Coia <sup>10</sup>, Claus Østergaard<sup>11</sup>, Michael Pedersen<sup>2</sup>, Henrik Westh <sup>2,12</sup>, Frank Møller Aarestrup<sup>1</sup>

**1** Technical University of Denmark, National Food Institute, Kongens Lyngby, Denmark, **2** Hvidovre Hospital, Department of Clinical Microbiology, Hvidovre, Denmark, **3** Herlev Hospital, Department of Clinical Microbiology, Herlev, Denmark, **4** Aalborg University Hospital, Department of Clinical Microbiology, Aalborg, Denmark, **5** Aarhus University Hospital, Department of Clinical Microbiology, Aarhus, Denmark, **6** Odense University Hospital, Department of Clinical Microbiology, Odense, Denmark, **7** Slagelse Hospital, Department of Clinical Microbiology, Slagelse, Denmark, **8** Rigshospitalet, Department of Clinical Microbiology, København, Denmark, **9** Hospital of Southern Jutland, Department of Clinical Microbiology, Sønderborg, Denmark, **10** Sydvestjysk Hospital, Department of Clinical Microbiology, Esbjerg, Denmark, **11** Vejle Hospital, Department of Clinical Microbiology, Vejle, Denmark, **12** Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

\* [anrire@food.dtu.dk](mailto:anrire@food.dtu.dk)



## OPEN ACCESS

**Citation:** Rebelo AR, Ibfelt T, Bortolaia V, Leekitcharoenphon P, Hansen DS, Nielsen HL, et al. (2022) One Day in Denmark: Nationwide point-prevalence survey of human bacterial isolates and comparison of classical and whole-genome sequence-based species identification methods. PLoS ONE 17(2): e0261999. <https://doi.org/10.1371/journal.pone.0261999>

**Editor:** Holger Rohde, Universitätsklinikum Hamburg-Eppendorf, GERMANY

**Received:** December 16, 2020

**Accepted:** December 16, 2021

**Published:** February 11, 2022

**Copyright:** © 2022 Rebelo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All genome files are available from the European Nucleotide Archive database (<http://www.ebi.ac.uk/ena>) under study accession number PRJEB37711.

**Funding:** This project was supported by The Novo Nordisk Foundation - Global Surveillance of Antimicrobial Resistance (NNF16OC0021856) attributed to FMA (<https://novonordiskfonden.dk/en/>), and the H2020 project VEO grant (874735)

## Abstract

### Objectives

Implementing whole-genome sequencing (WGS) technologies in clinical microbiology laboratories can increase the amount and quality of information available for healthcare practitioners. In this study, we analysed the applicability of this method and determined the distribution of bacterial species processed in clinical settings in Denmark.

### Methods

We performed a point-prevalence study of all bacterial isolates (n = 2,009) processed and reported in the Clinical Microbiology Laboratories in Denmark in one day in January 2018. We compared species identification as performed by classical methods (MALDI-TOF) and by bioinformatics analysis (KmerFinder and rMLST) of WGS (Illumina NextSeq) data. We compared the national point-prevalence of bacterial isolates observed in clinical settings with the research attention given to those same genera in scientific literature.

### Results

The most prevalent bacterium was *Escherichia coli* isolated from urine (n = 646), followed by *Staphylococcus* spp. from skin or soft tissues (n = 197). The distribution of bacterial

attributed to FMA (<https://www.veo-europe.eu/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

species throughout the country was not homogeneous. We observed concordance of species identification for all methods in 95.7% ( $n = 1,919$ ) of isolates, furthermore obtaining concordance for 99.7% ( $n = 1,999$ ) at genus level. The number of scientific publications in the country did not correlate with the number of bacterial isolates of each genera analysed in this study.

## Conclusions

WGS technologies have the potential to be applied in clinical settings for routine diagnostics purposes. This study also showed that bioinformatics databases should be continuously improved and results from local point-prevalence surveys should not be applied at national levels without previously determining possible regional variations.

## Introduction

Of the 57 million yearly deaths worldwide, 8.5 million (approximately 15%) are the direct result of infectious diseases [1]. This leads to major expenditures associated with diagnostics, treatment and infection control, as well as health impacts and societal costs associated with illness [2,3]. Out of the multiple potential bacterial infectious agents, current surveillance programs and control efforts are often focused on few pre-defined microorganisms and might not adequately reflect the true prevalence and burden of bacterial diseases [4,5]. In Denmark, large efforts go into tracking, reporting and controlling selected pathogens, namely methicillin-resistant *Staphylococcus aureus* (MRSA), Shiga toxin-producing *Escherichia coli* (STEC), ESBL-producing invasive *E. coli*, carbapenemase-producing organisms (CPO), vancomycin-resistant enterococci (VRE), *Neisseria gonorrhoeae* and *Neisseria meningitidis* isolates, amongst others ([www.ssi.dk](http://www.ssi.dk)). Similarly, for Europe, the European Centre for Disease Prevention and Control (ECDC) has in place surveillance systems for many of the same pathogens, and the World Health Organization (WHO) includes several antimicrobial-resistant bacterial species of concern in their “WHO priority pathogens list for R&D of new antibiotics” [6]. These systems are focused on bacteria causing infections considered most severe or those considered to have the highest potential to be transmitted or to be refractory to antimicrobial therapy. New surveys and research studies are regularly concerned with novel emerging issues often because of severity of disease, risk of spread of antimicrobial resistance mechanisms or even the emphasis put on novel risks by the media and politicians [7,8]. Globally, a very large number of studies has reported on prevalence and antimicrobial resistance in selected bacterial species and in specific clinical settings. Examples are the SENTRY program which analysed blood-stream infection isolates throughout a 20-year study incorporating over 200 medical centres worldwide [9], and the ECDC reports on the prevalence of healthcare-associated infections in point-prevalence surveys undertaken in 2011 and 2012 [10], and from 2016 to 2017 in the EU/EEA [11]. Surprisingly, very few studies have reported on the relative prevalence of all bacterial species observed nationally without pre-selection of settings or species. One such example corresponds to a five-year analysis of clinical samples at one Ethiopian hospital [12]. These factors might mask the true burden of human bacterial pathogens and we speculated that some bacteria frequently targeted by surveillance efforts can represent a lower healthcare burden than expected, while certain bacterial species might represent a high risk for human health while not being given sufficient attention. To our knowledge, the occurrence of all bacteria observed at the clinical microbiological laboratories in Denmark has not been yet elucidated, although

several reports and studies have focused on specific bacterial species, isolates from certain sample sources or on the prevalence of multi-drug resistant bacteria [13–16].

Currently, all clinical microbiology laboratories in Denmark have access to Matrix Assisted Laser Desorption Ionization-Time Of Flight (MALDI-TOF) technologies to identify bacterial pathogens in their daily routine work, although many urine isolates are identified directly from growth on chromogenic agars combined with an easy to perform confirmatory test like the indole spot test. However, the recent improvements and the decrease in cost of next generation sequencing technologies might allow for their application in clinical settings [17]. Implementing whole-genome sequencing (WGS) processes in clinical microbiology laboratories might lead to accurate species identification of bacterial pathogens, while providing information on antimicrobial resistance and virulence genes and allowing data to be processed in the exact same way throughout whole regions or countries. Furthermore, those data can be easily stored and transferred, and can be retrospectively screened if suspicion of an outbreak arises or if new mechanisms of antimicrobial resistance are discovered [18,19]. Although several publications have discussed the potential of employing WGS technologies in clinical settings, to date very few studies effectively analysed the applicability of WGS in a clinical setting, and a more limited number of studies did not focus on pre-selected species [20–23].

To obtain an unbiased overview of bacterial species distribution in clinical settings in Denmark, we conducted the “One Day in Denmark” point prevalence survey including all clinical microbiological laboratories in Denmark and covering all bacterial isolates processed on a single day, from a population of 5.8 million people. We describe our observations regarding the distribution of bacterial species and sample sources and we present a comparison of species identification performed through currently used classical diagnostics methods, in particular MALDI-TOF, and through WGS and different bioinformatics analyses to evaluate the applicability of WGS in a routine clinical context. Furthermore, we compare our findings with estimations of research attention attributed to each bacterial genus.

## Materials and methods

### Bacterial isolates

The study involved all 11 Departments of Clinical Microbiology (DCM) in Denmark including Herlev Hospital, Herlev; Hvidovre Hospital, Hvidovre; Nykøbing F. Sygehus, Nykøbing F; Odense Universitetshospital, Odense; Rigshospitalet, København; Slagelse Sygehus, Slagelse; Sydvestjysk Sygehus, Esbjerg; Sygehus Lillebælt, Vejle; Sygehus Sønderjylland, Sønderborg; Aalborg Universitetshospital, Aalborg; and Aarhus Universitetshospital, Skejby. Geographical provenience of samples was anonymized and each of the 11 DCM received a code (from F1 to F11) for the purpose of this study. The DCM provided all isolates from routine analyses ( $n = 2,073$ ) present on Wednesday, January 10, 2018, and for which species identification and antimicrobial resistance results had been reported by the responsible clinical microbiologist to the associated hospital or requesting general practitioners. These isolates corresponded to the clinically relevant bacteria recovered and analysed after culturing patients' biological samples in the DCM, and did not incorporate surveillance nor environmental samples. Metadata including isolate source and species identification were provided. All isolates were anonymized and no patient identifiers were transferred. Isolate source and species identification were normalized for analyses, including correction of species for 13 isolates due to very high probability of mislabelling, which was confirmed by re-analysis through MALDI-TOF in 12 cases (S1 and S2 Tables). The project was approved by local Danish Data Protection Agencies in each Region and Material Transfer Agreements were entered between DTU and all DCM.

The cultures were collected by car and transported to the Technical University of Denmark (DTU) within 24 hours of collection maintaining a refrigeration (agar plates) or freezing (glycerol stocks) temperature. After purity check, and sub-culture for purity when needed, the isolates were stored at  $-80^{\circ}\text{C}$  in glycerol stocks. According to the metadata provided, 2,024 isolates were bacteria and 49 were yeasts. Only the bacterial isolates were analysed in this study. Of these, 19 were excluded during further laboratory procedures due to: i) being correctly identified as yeasts (where the metadata reported *Streptococcus agalactiae* (n = 2) and *Streptococcus dysgalactiae* (n = 1)); ii) loss of viability (where the metadata reported *N. gonorrhoeae* (n = 5), *Streptococcus pneumoniae* (n = 3), *Campylobacter jejuni* (n = 2) and non-haemolytic *Streptococcus* sp. (n = 1)); and iii) mixed cultures impossible to purify to the level of single species mentioned in the metadata (*Kocuria* sp. (n = 1), *Moraxella catarrhalis* (n = 1), *Proteus hauseri* (n = 1), *Rothia dentocariosa* (n = 1) and *N. gonorrhoeae* (n = 1)). Four new isolates of clinical relevance were obtained from mixed cultures, yielding a final collection of 2,009 bacterial isolates with available clinical metadata which were analysed in this study.

### Classical microbiology methods for species identification

Species identification was performed at the DCM using routine methods including MALDI-TOF (Bruker Daltonik, Bremen, Germany) or Vitek MS (BioMerieux, Marcy l'Etoile, France) and/or analysis of specific colony morphology on selective culture plates or chromogenic plates according to each laboratory's standard operating procedures. Library versions vary between DCM and are continuously updated by manufacturers and supplemented internally and specifically for particular genera such as *Vibrio*, *Yersinia* and *Staphylococcus*.

### Whole genome sequence-based species identification

Genomic DNA was extracted from 2,009 bacterial isolates using the Easy-DNA<sup>TM</sup> Kit (Invitrogen, Carlsbad, CA, USA) and DNA concentrations were determined using the Qubit<sup>TM</sup> dsDNA high-sensitivity (HS) and/or broad-range (BR) assay kits (Invitrogen, Carlsbad, CA, USA). Genomic DNA was prepared for Illumina pair-end sequencing using the Illumina (Illumina, Inc., San Diego, CA, USA) NexteraXT<sup>®</sup> DNA Library Prep Reference Guide (Document #15031942, v03, February 2018) and NextSeq System Denature and Dilute Libraries Guide (Document #15048776, v03, April 2018). The libraries were sequenced using the Illumina NextSeq 500 platform. The raw reads were *de novo* assembled using the Center for Genomic Epidemiology pipeline for assembly and quality control. Quality thresholds were set at maximum 500 contigs per genome and maximum 0.5 million base-pairs of deviation from expected genome size.

Species identification was performed using bioinformatics tools KmerFinder [24,25] which performs k-mer alignment from WGS data (<https://cge.cbs.dtu.dk/services/KmerFinder/>) and Ribosomal Multilocus Sequence Typing (rMLST) [26] including 53 ribosomal genes using pubmlst database (<https://pubmlst.org/rmlst/>). Detection of the *mecA* gene was performed using ResFinder [27] 4.0 (<https://cge.cbs.dtu.dk/services/ResFinder/>) for 347 *S. aureus* or *Staphylococcus argenteus* isolates.

Raw sequence data have been submitted to the European Nucleotide Archive (<http://www.ebi.ac.uk/ena>) under study accession no.: PRJEB37711. A complete list of genomic sequence data is available in the [S1 Table](#).

### Estimation of research focus

To estimate research attention associated with all bacterial genera present in our collection we performed a search using the Web of Science platform (December 2019). We extracted the

total number of publications indexed in all databases for the pre-set timespans of i) all years and ii) last five years. We further refined the search criteria by the field “Countries/Regions” and extracted the corresponding number of publications from Denmark. The search was performed both at genera level and species level, and both with and without the inclusion of the keyword “clinical” (S3 Table). To compare the data we first calculated the ratio between the number of isolates in our collection and the number of publications found in the platform. When the resulting proportion was large, it indicated that there were more isolates present in our bacterial collection than what would be expected according to the number of publications. When the resulting proportion was low, the genus was prevalent in the scientific literature but we found few bacterial isolates in our collection. We calculated the mean of these proportions, and compared each of them with the obtained mean: a result above the mean corresponded to genera under-represented in literature and a result below the mean corresponded to genera that were over-represented. In order to obtain the results in a linear scale we converted them using the logarithm of the proportions compared to the mean proportion.

For analyses purposes we used the results obtained for the interval of all time, at national (Danish) level, including the search term “clinical”. To analyse the research focus of one resistance-specific surveillance system, screening of the *mecA* gene through ResFinder 4.0 was used for an estimation of prevalence of MRSA isolates in the clinical collection yielding 11 potential MRSA isolates (S1 Table).

## Results

### Bacterial isolates

The 2,009 isolates were distributed throughout 37 bacterial genera, according to the original data received from the DCM. The most prevalent genus was *Escherichia* (n = 707, 35.1%) followed by *Staphylococcus* (n = 414, 20.6%), *Streptococcus* (n = 232, 11.5%), *Klebsiella* (n = 143, 7.1%), *Enterococcus* (n = 132, 6.6%), *Haemophilus* (n = 62, 3.1%) and *Pseudomonas* (n = 61, 3.0%). Thirty additional genera included 251 (12.5%) isolates (Table 1). Seven isolates were included in the “Undetermined” category (0.3%), which encompassed two isolates identified either as *Aerococcus urinae* or *Escherichia coli*, one isolate belonging either to *Proteus vulgaris* or genus *Enterococcus*, one isolate identified either as *Proteus mirabilis* or Lancefield group B hemolytic *Streptococcus*, one isolate classified as “enterobacteria”, one isolate described as “Gram-positive cocci” and one isolate characterized as “not identified normal flora”. Although the species *Enterobacter aerogenes* was recently reclassified as *Klebsiella aerogenes* [28], and although there have been suggestions to re-evaluate the members of genus *Propionibacterium* (such as renaming *Propionibacterium acnes* as *Cutibacterium acnes*) [29], we maintained the original nomenclature both because the majority of the databases used in this project still contain the previous nomenclature and to be able to perform retrospective literature searches when estimating research attention attributed to each genera throughout “all-years” publications.

### Origin of samples

According to the information provided by the DCM, urine cultures was the most prevalent isolate source (n = 1,095, 54.5% of all isolates) followed by skin or soft tissue samples (n = 290, 14.4%), samples originating from the respiratory tract (n = 158, 7.9%) and blood (n = 123, 6.1%). 211 isolates (10.5%) were from undetermined sample sites, including 157 (7.8%) from swabs, 15 from pus (0.7%), 13 from tissues (0.6%), nine from abscess (0.4%), five from secretions (0.2%) and 12 from unknown sources (0.6%). The remaining 132 isolates (6.6%) were

Table 1. Distribution of the 2,009 clinical isolates according to bacterial genera and sample source according to the metadata provided by the DCM.

Sample source	Urine	Skin or soft tissue	Respiratory system	Blood	Stool or rectum	Eye	Ear	Reproductive system	Abdomen	Undetermined and other sources (1)	Total	Percentage of total
<i>Escherichia</i>	646	4	8	23	17				1	8	707	35.2
<i>Staphylococcus</i>	25	197	28	40	1	9	12	3	3	96	414	20.6
<i>Streptococcus</i>	38	57	23	18	2		1	19	2	72	232	11.5
<i>Klebsiella</i>	117	2	7	12			2		2	1	143	7.1
<i>Enterococcus</i>	101	4	3	16	1		1	1	3	2	132	6.6
<i>Haemophilus</i>			33			10	4			15	62	3.1
<i>Pseudomonas</i>	32	13	10			1	1			4	61	3
<i>Proteus</i>	48	5	2			1					56	2.8
<i>Enterobacter</i>	25		7	2			2		1	3	40	2
<i>Citrobacter</i>	28		3					1	3	1	36	1.8
<i>Aerococcus</i>	23										23	1.1
<i>Moraxella</i>			17			2	1			2	22	1.1
<i>Propionibacterium</i>		4		1		1			1	1	8	0.4
<i>Bacteroides</i>		1		4						2	7	0.3
<i>Serratia</i>			6							1	7	0.3
<i>Salmonella</i>	1			3	1						5	0.2
<i>Stenotrophomonas</i>			5								5	0.2
<i>Yersinia</i>				1	4						5	0.2
<i>Corynebacterium</i>		1	1			1				1	4	0.2
<i>Pasteurella</i>										4	4	0.2
<i>Actinomyces</i>						1				2	3	0.1
<i>Campylobacter</i>					3						3	0.1
<i>Neisseria</i>	1		1							1	3	0.1
<i>Raoultella</i>	3										3	0.1
<i>Acinetobacter</i>							1			1	2	0.1
<i>Morganella</i>	1	1									2	0.1
<i>Prevotella</i>			1							1	2	0.1
<i>Providencia</i>	2										2	0.1
Undetermined and other genera (2)	4	1	3	3	3					2	16	0.8
Total	1095	290	158	123	32	26	25	24	16	220	2009	100
Percentage of total	54.5	14.4	7.9	6.1	1.6	1.3	1.2	1.2	0.8	11	100	

(1) Samples of undetermined source (n = 211, 10.5% of the whole collection) or samples originating from sources with few representatives in the collection (Bone or joint (n = 4), Brain or nervous system (n = 3), and Milk (n = 2)).

(2) Isolates of undetermined species (n = 7, 0.3% of the whole collection) or isolates belonging to genera with only one representative in the collection (*Aeromonas*, *Anaerococcus*, *Clostridium*, *Finexgoldia*, *Fusobacterium*, *Micrococcus*, *Peptoniphilus*, *Rothia* and *Shigella*).

<https://doi.org/10.1371/journal.pone.0261999.t001>

distributed among other sources including stool or rectum, eye, ear, reproductive system, abdomen, brain or nervous system, milk, and bone or joint (Table 1).

The most prevalent genus-sample source combination observed in the collection corresponded to *Escherichia* isolates recovered from urine (n = 646), followed by *Staphylococcus* recovered from skin or soft tissue (n = 197), and *Klebsiella* and *Enterococcus* isolated from urine (n = 117 and n = 101, respectively) (Table 1).

The 11 DCM contributed between 41 (2%) and 392 (19.5%) isolates each (S2 Table). For the 12 genera that are represented by more than 10 isolates in the collection (n = 1,928 isolates)

their distribution according to the providing DCM presented a few unexpected patterns, with some large deviations observed for certain DCM-species combinations justified by laboratory-specific procedures.

### Whole-genome sequence-based species identification

Three isolates did not respect the quality control parameters' threshold and were excluded from the WGS analyses (S1 Table).

For 1,872 isolates (93.3%) both KmerFinder and rMLST-based identification were in concordance with the species identification provided by the DCM. This included 48 cases where DCM identification was only performed to genus level and 50 cases where it was performed at Lancefield group level, per DCM protocols. Furthermore, concordance was observed for 1,976 isolates (98.6%) at genus level. After obtaining these results, 91 isolates were reanalysed by MALDI-TOF because of discordance between original identification by the DCM and WGS results. The re-analyses were limited to those isolates for which none of the bioinformatics tool outputs were in concordance with the original species identification. After MALDI-TOF re-testing, the concordance between classical and WGS-based identification was 95.7% ( $n = 1,919$ ) at the species level and 99.7% ( $n = 1,999$ ) at the genus level.

Different types of discordance were identified (Table 2). For 70 isolates (3.5%) both bioinformatics tools provided a concordant species identification, which was different from the species identified by the DCM. Of these, discordance at the genus and the species level was observed in 14 (0.7%) and 56 isolates (2.8%), respectively. After re-analysis by MALDI-TOF, the remaining discordances were represented by 28 cases (1.4%), of which 27 (1.3%) corresponded to discordant species identification but concordant genus identification. For two of these 27 cases we were unable to obtain new MALDI-TOF results, as well as for the remaining discordant case which was the only for which a discordance at genus level persisted (Table 2).

For 40 isolates (2%), there was concordance between species identification obtained by one bioinformatics tool and the routine method at the DCM, but the other bioinformatics tool provided a different result. Of these, 32 (1.6%) isolates corresponded to discordance using rMLST which all occurred at species but none at genus level. Eight cases were attributed to discordance using KmerFinder (0.4%), with only two of those cases (0.1%) corresponding to misidentification at the genus level.

For 16 isolates (0.8%) it was not possible to obtain an agreement in species identification by any of the three methods, although for 10 (0.5%) isolates the methods were concordant at genus level. After MALDI-TOF re-testing, conventional method and KmerFinder provided concordant species identification for six of the 16 isolates (0.3%), and for these isolates rMLST results were also concordant at the genus level. Furthermore two (0.1%) cases revealed concordance at species level between MALDI-TOF results and rMLST, also agreeing at genus level with KmerFinder. Of the remaining eight discordant cases, six were discordant only at species level; new MALDI-TOF analyses were either unable to characterize those samples to species or even genus level ( $n = 1$  and  $n = 3$ , respectively) or the isolates were not re-tested ( $n = 2$ ). The other two cases correspond to a complete discordance between the three sets of data, of which only one was re-tested through MALDI-TOF maintaining the discrepancy.

Eight cases (0.4%) were classified as undetermined regarding species concordance. For five cases (0.2%) no definite species identification was received from the DCM but results from both bioinformatics tools were in agreement at species level, and MALDI-TOF re-testing results revealed complete species concordance across the species identification methods. In another case (0.05%), the bioinformatics tools provided identification of congeneric species whereas DCM identification was not known. New MALDI-TOF results revealed a different



Table 2. Discordance in species identification performed initially at the DCM, and then by MALDI-TOF re-testing, KmerFinder and rMLST.

Discordance type	Initial DCM species identification	Number of isolates	MALDI-TOF re-test results	KmerFinder species identification	rMLST species identification
Both bioinformatics tools provide the same species identification, which is different from initial DCM identification	<i>Acinetobacter baumannii</i>	1	<i>Acinetobacter pittii</i>	<i>Acinetobacter pittii</i>	<i>Acinetobacter pittii</i>
	<i>Aerococcus urinae</i>	4	<i>Aerococcus sanguinicola</i> (n = 2)	<i>Aerococcus sanguinicola</i>	<i>Aerococcus sanguinicola</i>
			<i>Aerococcus urinae</i> (n = 1)		
			No peaks found (n = 1)		
	<i>Campylobacter jejuni</i>	1	<i>Escherichia coli</i>	<i>Escherichia coli</i>	<i>Escherichia coli</i>
	<i>Citrobacter freundii</i>	2	<i>Citrobacter braakii</i>	<i>Citrobacter braakii</i>	<i>Citrobacter braakii</i>
	<i>Clostridium difficile</i>	1	<i>Bacteroides vulgatus</i>	<i>Bacteroides vulgatus</i>	<i>Bacteroides vulgatus</i>
	<i>Enterobacter cloacae</i>	1	<i>Enterobacter aerogenes</i>	<i>Enterobacter aerogenes</i>	<i>Enterobacter aerogenes</i>
			<i>Enterobacter bugandensis</i>	<i>Enterobacter bugandensis</i>	<i>Enterobacter bugandensis</i>
			<i>Enterobacter cloacae</i> (n = 6)	<i>Enterobacter hormaechei</i>	<i>Enterobacter hormaechei</i>
	<i>Enterobacter xiangfangensis</i> (n = 2)				
	<i>Enterococcus faecalis</i>	1	<i>Enterococcus faecium</i>	<i>Enterococcus faecium</i>	<i>Enterococcus faecium</i>
			<i>Staphylococcus haemolyticus</i>	<i>Staphylococcus haemolyticus</i>	<i>Staphylococcus haemolyticus</i>
	<i>Escherichia coli</i>	1	<i>Klebsiella oxytoca</i>	<i>Klebsiella michiganensis</i>	<i>Klebsiella michiganensis</i>
	<i>Haemophilus parainfluenzae</i>	1	<i>Moraxella catarrhalis</i>	<i>Moraxella catarrhalis</i>	<i>Moraxella catarrhalis</i>
	<i>Klebsiella oxytoca</i>	4	<i>Klebsiella oxytoca</i>	<i>Klebsiella michiganensis</i>	<i>Klebsiella michiganensis</i>
	<i>Klebsiella pneumoniae</i>	1	<i>Escherichia coli</i>	<i>Escherichia coli</i>	<i>Escherichia coli</i>
			<i>Klebsiella oxytoca</i>	<i>Klebsiella oxytoca</i>	<i>Klebsiella oxytoca</i>
			<i>Klebsiella pneumoniae</i> (n = 7)	<i>Klebsiella quasipneumoniae</i>	<i>Klebsiella quasipneumoniae</i>
	7	<i>Klebsiella variicola</i>	<i>Klebsiella variicola</i>	<i>Klebsiella variicola</i>	
	<i>Neisseria gonorrhoeae</i>	1	<i>Aerococcus sanguinicola</i>	<i>Aerococcus christensenii</i>	<i>Aerococcus christensenii</i>
	<i>Pasteurella multocida</i>	1	<i>Pasteurella dagmatis</i>	<i>Pasteurella dagmatis</i>	<i>Pasteurella dagmatis</i>
	<i>Pseudomonas stutzeri</i>	1	<i>Moraxella catarrhalis</i>	<i>Moraxella catarrhalis</i>	<i>Moraxella catarrhalis</i>
	<i>Raoultella planticola</i>	1	<i>Raoultella ornithinolytica</i>	<i>Raoultella ornithinolytica</i>	<i>Raoultella ornithinolytica</i>
	<i>Rothia mucilaginosa</i>	1	<i>Haemophilus influenzae</i>	<i>Haemophilus influenzae</i>	<i>Haemophilus influenzae</i>
	<i>Staphylococcus aureus</i>	1	No organism ID possible	<i>Halomonas hydrothermalis</i>	<i>Halomonas hydrothermalis</i>
			<i>Pseudomonas aeruginosa</i>	<i>Pseudomonas aeruginosa</i>	<i>Pseudomonas aeruginosa</i>
			<i>Staphylococcus argenteus</i>	<i>Staphylococcus argenteus</i>	<i>Staphylococcus argenteus</i>
	<i>Staphylococcus hominis</i>	1	<i>Staphylococcus saprophyticus</i>	<i>Staphylococcus saprophyticus</i>	<i>Staphylococcus saprophyticus</i>
	<i>Staphylococcus lugdunensis</i>	1	<i>Corynebacterium striatum</i>	<i>Corynebacterium striatum</i>	<i>Corynebacterium striatum</i>
	<i>Streptococcus anginosus</i>	1	<i>Streptococcus constellatus</i>	<i>Streptococcus agalactiae</i>	<i>Streptococcus agalactiae</i>
			<i>Streptococcus constellatus</i>	<i>Streptococcus constellatus</i>	<i>Streptococcus constellatus</i>
	<i>Streptococcus dysgalactiae</i>	1	<i>Enterococcus faecalis</i>	<i>Enterococcus faecalis</i>	<i>Enterococcus faecalis</i>
	<i>Streptococcus</i> sp. Lancefield group B	2	No organism ID possible (n = 1)	<i>Streptococcus dysgalactiae</i>	<i>Streptococcus dysgalactiae</i>
			<i>Streptococcus canis</i> (n = 1)		
	<i>Streptococcus mitis</i>	1	<i>Streptococcus oralis</i>	<i>Streptococcus oralis</i>	<i>Streptococcus oralis</i>
	<i>Streptococcus pneumoniae</i>	2	<i>Streptococcus mitis</i>	<i>Streptococcus mitis</i>	<i>Streptococcus mitis</i>
			<i>Streptococcus pseudopneumoniae</i>	<i>Streptococcus pseudopneumoniae</i>	<i>Streptococcus pseudopneumoniae</i>
	<i>Streptococcus pyogenes</i>	1	<i>Corynebacterium striatum</i>	<i>Corynebacterium striatum</i>	<i>Corynebacterium striatum</i>
<i>Streptococcus sanguinis</i>	3	<i>Streptococcus gordonii</i>	<i>Streptococcus gordonii</i>	<i>Streptococcus gordonii</i>	

(Continued)

Table 2. (Continued)

Discordance type	Initial DCM species identification	Number of isolates	MALDI-TOF re-test results	KmerFinder species identification	rMLST species identification
One bioinformatics tool in concordance with DCM identification and the other tool discordant	<i>Citrobacter freundii</i>	3	Not re-tested	<i>Citrobacter freundii</i>	<i>Citrobacter portucalensis</i>
	<i>Enterobacter cloacae</i>	14	Not re-tested*	<i>Enterobacter cloacae</i>	<i>Enterobacter hormaechei</i>
		1	Not re-tested		<i>Enterobacter kobei</i>
		3	Not re-tested		<i>Enterobacter ludwigii</i>
	<i>Klebsiella oxytoca</i>	5	Not re-tested	<i>Klebsiella oxytoca</i>	<i>Klebsiella grimontii</i>
		3	Not re-tested		<i>Klebsiella michiganensis</i>
	<i>Proteus vulgaris</i>	1	Not re-tested	<i>Proteus vulgaris</i>	<i>Proteus cibarius</i>
		1	Not re-tested		<i>Proteus genomsp. 4</i>
		1	Not re-tested		<i>Proteus genomsp. 6</i>
	<i>Enterococcus casseliflavus</i>	1	Not re-tested	<i>Enterococcus gallinarum</i>	<i>Enterococcus casseliflavus</i>
	<i>Enterococcus gallinarum</i>	1	Not re-tested	<i>Enterococcus sp.</i>	<i>Enterococcus gallinarum</i>
	<i>Escherichia coli</i>	2	Not re-tested	<i>Escherichia marmotae</i>	<i>Escherichia coli</i>
	<i>Pasteurella canis</i>	1	Not re-tested	<i>Pasteurella dagmatis</i>	<i>Pasteurella canis</i>
	<i>Prevotella baroniae</i>	1	Not re-tested	<i>Prevotella denticola</i>	<i>Prevotella baroniae</i>
	<i>Shigella sonnei</i>	1	Not re-tested	<i>Escherichia coli</i>	<i>Shigella sonnei</i>
	<i>Peptoniphilus harei</i>	1	Not re-tested	<i>Anaerococcus sp.</i>	<i>Peptoniphilus harei</i>
No agreement between species identification	<i>Anaerococcus sp.</i>	1	Not re-tested	<i>Anaerococcus sp.</i>	<i>Anaerococcus lactolyticus</i>
	<i>Corynebacterium macginleyi</i>	1	No organism ID possible	<i>Corynebacterium simulans</i>	<i>Corynebacterium accolens</i>
	<i>Corynebacterium tuberculostearicum</i>	2	No organism ID possible	<i>Corynebacterium striatum</i>	<i>Corynebacterium pseudogenitalium</i>
	<i>Enterobacter asburiae</i>	1	<i>Enterobacter cloacae</i>	<i>Enterobacter cloacae</i>	<i>Enterobacter rogenkampii</i>
	<i>Enterobacter cloacae</i>	1	<i>Enterobacter xiangfangensis</i>	<i>Enterobacter xiangfangensis</i>	<i>Enterobacter hormaechei</i>
	<i>Escherichia coli</i>	1	<i>Acinetobacter baumannii</i>	<i>Acinetobacter calcoaceticus</i>	<i>Acinetobacter baumannii</i>
	<i>Klebsiella oxytoca</i>	1	<i>Raoultella planticola</i>	<i>Raoultella planticola</i>	<i>Raoultella ornithinolytica</i>
	<i>Neisseria gonorrhoeae</i>	1	<i>Actinomyces odontolyticus</i>	<i>Streptococcus equinus</i>	<i>Granulicatella adiacens</i>
	<i>Pasteurella sp.</i>	1	Not re-tested	<i>Pasteurella multocida</i>	<i>Pasteurella sp.</i>
	<i>Prevotella sp.</i>	1	<i>Anaerococcus hydrogenalis</i>	<i>Anaerococcus sp.</i>	<i>Anaerococcus hydrogenalis</i>
	<i>Propionibacterium sp.</i>	1	Not re-tested	<i>Propionimicrobium sp.</i>	<i>Propionimicrobium lymphophilum</i>
	<i>Proteus hauseri</i>	1	<i>Proteus vulgaris</i>	<i>Proteus vulgaris</i>	<i>Proteus genomsp. 4</i>
	<i>Pseudomonas aeruginosa</i>	1	<i>Pseudomonas sp.</i>	<i>Pseudomonas alkylphenolica</i>	<i>Pseudomonas sp.</i>
	<i>Raoultella planticola</i>	1	<i>Klebsiella oxytoca</i>	<i>Klebsiella oxytoca</i>	<i>Klebsiella grimontii</i>
	<i>Streptococcus sp.</i>	1	<i>Streptococcus mitis</i>	<i>Streptococcus mitis</i>	<i>Streptococcus infantis</i>
	Undetermined	<i>Actinomyces turicensis</i>	1	No peaks found	No output
Undetermined		1	<i>Enterococcus faecium</i>	<i>Enterococcus faecium</i>	<i>Enterococcus faecium</i>
		1	<i>Erwinia persicina</i>	<i>Erwinia persicina</i>	<i>Erwinia persicina</i>
		1	<i>Escherichia coli</i>	<i>Escherichia coli</i>	<i>Escherichia coli</i>
		1	<i>Morganella morganii</i>	<i>Morganella morganii</i>	<i>Morganella morganii</i>
		1	<i>Pseudomonas aeruginosa</i>	<i>Pseudomonas aeruginosa</i>	<i>Pseudomonas aeruginosa</i>
		1	<i>Actinotignum sanguinis</i>	<i>Actinotignum schaalii</i>	<i>Actinotignum timonense</i>
		1	No organism ID possible	<i>Luteipulveratus mongoliensis</i>	<i>Barrientosimonas humi</i>

\*One of these isolates was re-tested using MALDI-TOF for mislabelling verification. The species *E. cloacae* was confirmed.

Cells highlighted in orange show discordances that persist after MALDI-TOF re-testing or inability to obtain species identification by the tool(s). Cells highlighted in blue show discordances for isolates not subjected to MALDI-TOF re-testing or inability to obtain species identification by the tool(s).

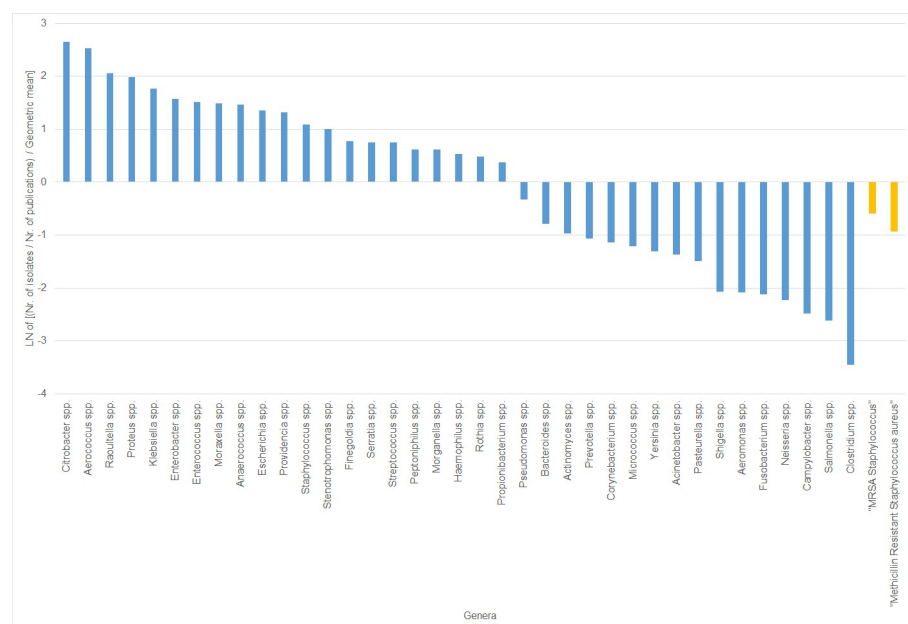
<https://doi.org/10.1371/journal.pone.0261999.t002>

but also congeneric species. An additional case (0.05%) corresponded to inability of KmerFinder to perform species identification, whereas both the DCM method and rMLST identified the same species. MALDI-TOF repetition of this isolate was unable to provide an identification. In the remaining case (0.05%), there was no agreement at the genus level by any of the methods and new MALDI-TOF experiments were unable to characterize the isolate.

The quality of bioinformatics tool outputs varied. KmerFinder identified 1,771 isolates (88.3%) with query genome coverages higher than 90%, 140 isolates (7%) with coverages between 80% and 90%, 19 isolates (0.9%) with coverages between 70% and 80% and 19 isolates (0.9%) between 60% and 70%. Seven isolates (0.3%) were identified with 50% to 60% and an additional seven isolates were identified with 40% to 50% of query coverage. The remaining 43 (2.1%) were identified with coverages lower than 40%. rMLST identified 1,858 isolates (92.6%) with 100% of support, while identifying 114 isolates (5.7%) with support between 95% and 99%. The remaining 34 isolates (1.7%) were identified with less than 95% of support. A full description of the quality outputs and their relationship to species identification concordances and discordances can be found in the [S4 Table](#).

### Estimation of research focus

Minor differences were observed when comparing the ratios calculated with the number of publications worldwide and with the number of publications in Denmark, while comparing the numbers of publications in the last five years and for all time, and with the inclusion of the key-word “clinical” in the search terms ([S3 Table](#)). For 21 genera present in the collection it appeared that the number of isolates observed was disproportionately higher than the number of publications focusing on that genus ([Fig 1](#)). This difference was particularly notable for the genera *Citrobacter*, *Aerococcus*, *Raoultella*, *Proteus* and *Klebsiella*. Of the 21 genera, nine are included in national surveillance systems, ECDC surveillance systems and/or the WHO priority pathogens list for R&D of new antibiotics ([S5 Table](#)). Conversely, for 16 genera it appeared that the number of isolates present in the collection was lower than expected when compared



**Fig 1. Comparison of prevalence and research attention focus according to bacterial genera.**

<https://doi.org/10.1371/journal.pone.0261999.g001>

with the number of publications, used as a measure of research focus. Of these, ten are included in national, ECDC and/or WHO lists of priority pathogens or pathogens under surveillance systems. The genera for which this difference was more accentuated were *Clostridium*, *Salmonella*, *Campylobacter*, *Neisseria* and *Fusobacterium*. Of note, the ratio for *Neisseria* genus (-2.22) could have been less accentuated if six additional presumptive *N. gonorrhoeae* isolates had not been excluded. Regarding MRSA isolates results seem to indicate that the number of publications focusing on this pathogen is higher than expected when compared with the number of potential MRSA isolates in national clinical settings.

Logarithms of the proportions “number of bacterial isolates of each genera in the collection, per number of publications concerning those genera”, relative to the mean number of bacterial isolates per publications. Results for “all time”, at national (Danish) level, including the search term “clinical”. Positive numbers correspond to genera over-represented in the bacterial collection and negative numbers correspond to under-represented genera. Yellow bars correspond to search terms not corresponding to bacterial genera but instead antimicrobial-resistant bacteria. Number of publications extracted from Web of Science All Databases collection. December 2019.

## Discussion

In this study we presented a point-prevalence survey of all bacterial isolates processed in one day in all clinical microbiology departments across Denmark, and thus encompassing isolates from hospitals and clinical practices. We clearly observe that *E. coli* was the predominant bacterial species identified, followed by *S. aureus*, *Streptococcus* spp. and *Klebsiella* spp. isolates. Urinary tract infections had the highest prevalence, likely because community samples were also included in this study. Skin and soft tissue infections, respiratory infections and isolates originating from blood culture also represented a high percentage of the observed bacteria. We observed a heterogeneous relation between sample sources and bacterial species, with the most notable exceptions being skin or soft tissue infections mainly associated with *Staphylococcus* and *Streptococcus* genera, stool samples primarily related with *E. coli* isolates, eye samples associated with *Staphylococcus* spp. and *Haemophilus* spp. isolates, ear samples related with *Staphylococcus* spp. isolates and reproductive system samples mainly connected with *Streptococcus* spp. isolates. It should be noted that the use of point-prevalence data obtained from only one single day is likely not representative of the seasonal or annual distribution of bacterial species; examples of genera known to be subjected to marked seasonal variations are *Campylobacter* and *Salmonella* which present a much higher national prevalence during summer months ([www.ssi.dk](http://www.ssi.dk)). Further work to increase the confidence of point-prevalence results of bacterial isolates in Denmark can include new surveys performed in different seasons or collection of epidemiological data from governmental databases. Our findings corroborate previous research which was however focused on aetiology of hospital-associated infections only. The SENTRY program analyzed bloodstream infection isolates throughout a 20-year study incorporating over 200 medical centers worldwide and showed that the most common etiological agent observed was *S. aureus*, followed by *E. coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Enterococcus faecalis*, *Staphylococcus epidermidis*, *Enterobacter cloacae*, *Streptococcus pneumoniae*, *Enterococcus faecium*, and *Acinetobacter baumannii*-*Acinetobacter calcoaceticus* species complex [9]. Additionally, the ECDC report on the prevalence of healthcare-associated infections revealed a somewhat similar profile of causative agents in hospital associated infections in two point-prevalence surveys from 2016 to 2017 in the EU/EEA, with the most frequently isolated bacterial species being *E. coli*, followed by *S. aureus*, *Klebsiella* spp., *Enterococcus* spp., *P. aeruginosa*, *Clostridium difficile*, coagulase negative staphylococci,

*Enterobacter* spp. and *Proteus* spp. [11]. The most prevalent types of infection were respiratory tract infections, urinary tract infections, surgical site infections, bloodstream infections and gastro-intestinal infections, corroborating observations already established in previous surveys undertaken in 2011 and 2012 [10]. In long-term healthcare facilities skin infections also proved prevalent. One study reporting on the relative prevalence of all bacterial species observed nationwide corresponds to a five-year analysis of clinical samples at one Ethiopian hospital, revealing that *S. aureus* was the most prevalent bacterial species, followed by *Salmonella* spp., *E. coli*, *P. aeruginosa*, *Shigella* spp., *K. pneumoniae* and *Streptococcus pyogenes*, with other species presenting lower prevalence. The main infection sources were stool cultures, urine samples, ear discharges, wound swabs and blood [12]. This distinct profile at its respective setting indicates that geographical location, and potentially national health-care resources, can strongly affect the epidemiological findings for each particular country.

We observed a moderate degree of homogeneity across the whole country of Denmark, with few exceptions. F3 provided very high percentages of *Haemophilus* spp. (34%), *Aerococcus* spp. (48%) and *Moraxella* spp. (36%) isolates while being responsible for only 10.6% of all isolates collected. These higher numbers of isolates are believed to be due to regional variation in the testing and reporting practices. F4 –responsible for 17.5% of the whole collection–provided 34% and 38% of *Klebsiella* spp. and *Proteus* spp. isolates, respectively, and 28% of all *Escherichia* spp. isolates. Conversely, it only contributed 0.4% (n = 1) of *Streptococcus* spp. isolates, and 0% of all *Aerococcus* spp. and *Moraxella* spp. DCM F4 does not usually perform antimicrobial susceptibility testing for *Streptococcus* spp., *Staphylococcus* spp., *Aerococcus* spp. and *Moraxella* spp. isolates, hence the absence of these genera from the samples processed there. The higher number of isolates belonging to other bacterial genera observed in that laboratory can be attributed to the presence of isolates from the previous day than that of the plate collection, which can be considered as a limitation regarding the point-prevalence results reported here. F7 contributed 1% and 0% of *Escherichia* spp. and *Aerococcus* spp. isolates, respectively, but 5% of the overall collection. F7 does not perform analyses of samples originating from general clinical practice, which likely explains the low prevalence of UTI-associated genera in their laboratory. We expect these findings to be useful both in guiding further research efforts and in making future institutional or governmental decisions. It is important to note that Denmark is a small northern-European country and these observations are likely very dependent on development stage and implementation of clinical microbiology, population size, geographical location and even seasonal variability. Furthermore, we believe that these results demonstrate that regional results of point-prevalence, prevalence or incidence studies should not be applied to national contexts unless properly validated, given the clear variation of species distribution in each laboratory.

Our study showed that there was very high concordance between WGS and the current routine methods for species identification on a collection of bacterial species that were not *a priori* selected. This indicates that WGS has the potential to be routinely used in clinical diagnostics. Currently, MALDI-TOF is the main method used routinely for species identification at DCM in Denmark, but despite its widespread use it has known weaknesses, such as several genera can only be identified at species group or complex level, as in the case of the *Streptococcus mitis* group or the *Enterobacter cloacae* complex. Although MALDI-TOF libraries can be improved by adding own mass spectrum entries as shown in this work for *Staphylococcus argenteus* (Table 2), bioinformatics libraries have the advantage of being transversal to several different sequencing platforms thus allowing for their application in various settings. Kmer-Finder performed very well but there were limitations in the percentages of query coverage, which were nevertheless above 80% for 95% of cases. Lower coverages were not surprising and are explainable according to the mechanisms of the alignment program, as it performs

alignment against full reference genomes. In practice, low percentages of query coverages will be observed for species which present few publicly available reference genomes, few complete genomes or genomes with low quality, or for species which present a naturally high genomic plasticity or variability. rMLST performed better than KmerFinder in terms of confidence as it systematically returned higher percentages of support, mainly since it considers a limited number of conserved genes (approximately 53 genes) for species identification. These are used to calculate a percentage of support corresponding to the percentage of detected species-specific exact-matching alleles in regard to the total number of alleles that support any reported taxa found in the genome (including any unknown taxa). However, rMLST seemed to lead to a slightly higher number of discordances in species identification when compared to KmerFinder, notably for isolates belonging to *Enterobacter* genus ( $n = 18$ , or 56.3% of the 32 isolates that rMLST could not identify as the same species as the DCM and KmerFinder). We obtained concordance at genus identification level for all tools in 1,999 isolates (99.7%). In a clinical context identification at this level might prove sufficient as often intrinsic and acquired resistance mechanisms are similar for the different congeneric species, meaning that empiric therapeutic approaches and control efforts might not change even if misidentifying congeneric species. Furthermore, additional information provided by bioinformatics pipelines will complement this identification and allow for more directed clinical and/or therapeutic interventions, such as *in silico* antibiograms [30]. We only observed discordances at genera level in five cases (0.2%), and two cases of undetermined concordance (0.1%) due to absence of either DCM ( $n = 1$ ) or KmerFinder ( $n = 1$ ) identifications. These situations might correspond to true failures of one or more of the methods tested or can also be the results of bacterial contaminations or undetected mislabelling which would lead to an incorrect bacterial isolate being analysed. Our findings expand on previous research focused on selected bacterial species. Long et al., 2013 analysed 130 clinical samples through BLAST alignment against the NCBI Nucleotide database, being unable to identify only 13 samples (10%) from the WGS datasets [20]. Similarly, Roach et al., 2015 performed genomic analysis of 1,229 bacterial isolates collected over one year at an intensive care unit; definitive genera identification through classical microbiological diagnostics methods was achieved for 1,054 samples, with 70 of these cases (6.6%) yielding discordances at genus identification level between the laboratory classification and the genomic identification [21]. Hasman et al., 2014 analysed the potential of applying WGS methods directly to 19 urine samples; while the metagenomics analyses yielded complex results, WGS and bioinformatics analyses of 18 of the purified isolates identified those as belonging to the same genera as predicted through conventional identification methods [22]. Likewise, Schmidt et al., 2017 were able to effectively identify the species of six human pathogens from urine samples using WGS and bioinformatics tools, both at biological sample and cultivated bacteria levels [23].

Our comparison between the point prevalence of different pathogens and the research focus suggests that currently there is a discrepancy between the prevalence of some bacteria and research focus. We believe that this is a highly important point and suggest that public health attention as well as funding for control programs and basic research should take into consideration the current prevalence of the pathogens, as an addition to other relevant metrics such as historical epidemiological data from surveillance programs, antimicrobial resistance emergence trends and national and international burdens of disease. It should be noted, however, that the number of research publications is not a flawless representation of research attention paid to certain species, as the number of publications can be high due to the infectiveness of that species or due to severe health impact of the infection, and possibly not directly related to funding attributed to the control of those infections. Additionally, a lower prevalence of certain species highly targeted in scientific research might be a reflection of seasonal

variation in prevalence or even demonstrate that the surveillance and research efforts were successful in reducing the burden of that pathogen. Other limitations of the method used to calculate the research focus are a difficulty in estimating the total research focus attributed to bacterial infections and the possibility that some species might be represented in other publications not considered in this study due to those being indexed with broader terms such as “*Enterobacteriaceae*”. Furthermore, many surveillance systems focus on isolates with particular antimicrobial resistance (AMR) profiles. As we did not investigate all AMR determinants it was difficult to estimate the adequacy of such systems, with the example of MRSA having been the only one analysed. It can be noted that a discrepancy in a national context was observed for this particular pathogen, which appeared to be the focus of a high number of research publications and is known to be the main focus of two national research groups while having a low prevalence in the bacterial collection. This difference highlights the success of national surveillance programs and control efforts.

MALDI-TOF is a fast and accurate identification method with very low analysis cost per sample and the potential to be improved, and will likely remain in use throughout clinical microbiology laboratories. However, it is plausible that in the future WGS could be the main diagnostics tools used in clinical microbiology settings, particularly integrating the advance of bioinformatics algorithms and databases and the increase of validation studies performed. User-friendly interfaces that perform complete bioinformatics analyses of genomes are available making this technology functional for healthcare workers that have no previous bioinformatics experience. These data can be stored and shared, and analyses can be performed in the exact same way across different healthcare settings in a reproducible way that registers all analyses parameters and allows for retrospective investigations. Other possible and important applications of this technology in these settings include antimicrobial susceptibility profiling, determination of sequence types or serotypes and timely outbreak detection and control, which are currently not achievable by one single diagnostics technology in place in the DCM. Furthermore, the advance of metagenomics sequencing techniques might in the future allow for application of these bioinformatics methods directly in biological samples.

In conclusion, we present a point-prevalence analysis of all bacterial isolates processed in Denmark in one day from clinical microbiology settings. We compared classical microbiology diagnostics methods (MALDI-TOF) with Illumina NextSeq technology and two bioinformatics tools (KmerFinder and rMLST) and our results show that WGS can be applied in a clinical setting for species identification purposes. We have compared the research attention given to bacterial genera and their effective national prevalence and suggest that regional data should not be applied to national contexts without a previous analysis of potential biases or local variations.

## Supporting information

**S1 Fig. Comparison of different publication metrics extracted from Web of Science, December 2019, and relation to the point-prevalence of bacterial isolates.**

(XLSX)

**S1 Table. Bacterial isolates collected in the “One Day in Denmark” collection and whole-genome sequence quality parameters.**

(XLSX)

**S2 Table. Distribution of bacterial isolates according to their genera and the providing DCM, in absolute frequency and in percentages relative to the total number of isolates of those genera.**

(XLSX)

**S3 Table. Publication metrics extracted from Web of Science, "All Databases", December 2019 and respective analysis.**

(XLSX)

**S4 Table. Quality of KmerFinder and rMLST outputs and description of species identification concordances and discordances.**

(XLSX)

**S5 Table. Comparison of the prevalence of bacterial genera studied in this project and their presence or absence from two surveillance systems (SSI at national level and ECDC at international level) and the WHO priority list for research and development.**

(XLSX)

## Acknowledgments

We thank Marie Stengaard Jensen for her assistance in collecting and organizing the samples. We thank Birthe S. Rosenqvist Lund, Christina Aaby Svendsen, Gunhild Larsen, Hanne Mordhorst, Hanne Nørgaard Nielsen, Inge Marianne Hansen, Jacob Dyring Jensen and Kate Vina Vibefeldt for their help with all laboratory work included in this project. We thank the staff of the Danish Departments of Clinical Microbiology for enabling the project.

## Author Contributions

**Conceptualization:** Ana Rita Rebelo, Tobias Ibfelt, Valeria Bortolaia, Pimlapas Leekitcharoenphon, Dennis Schröder Hansen, Hans Linde Nielsen, Svend Ellermann-Eriksen, Michael Kemp, Bent Løwe Røder, Niels Frimodt-Møller, Turid Snekløth Søndergaard, John Eugenio Coia, Claus Østergaard, Henrik Westh, Frank Møller Aarestrup.

**Data curation:** Ana Rita Rebelo, Tobias Ibfelt.

**Formal analysis:** Ana Rita Rebelo, Tobias Ibfelt, Michael Pedersen.

**Funding acquisition:** Frank Møller Aarestrup.

**Investigation:** Ana Rita Rebelo, Tobias Ibfelt.

**Methodology:** Ana Rita Rebelo, Valeria Bortolaia, Pimlapas Leekitcharoenphon, Michael Pedersen, Frank Møller Aarestrup.

**Project administration:** Ana Rita Rebelo, Valeria Bortolaia, Pimlapas Leekitcharoenphon, Henrik Westh, Frank Møller Aarestrup.

**Resources:** Valeria Bortolaia, Pimlapas Leekitcharoenphon, Dennis Schröder Hansen, Hans Linde Nielsen, Svend Ellermann-Eriksen, Michael Kemp, Bent Løwe Røder, Niels Frimodt-Møller, Turid Snekløth Søndergaard, John Eugenio Coia, Claus Østergaard, Henrik Westh, Frank Møller Aarestrup.

**Supervision:** Valeria Bortolaia, Frank Møller Aarestrup.

**Validation:** Ana Rita Rebelo.

**Visualization:** Ana Rita Rebelo.

**Writing – original draft:** Ana Rita Rebelo.

**Writing – review & editing:** Ana Rita Rebelo, Tobias Ibfelt, Valeria Bortolaia, Pimlapas Leekitcharoenphon, Dennis Schröder Hansen, Hans Linde Nielsen, Svend Ellermann-Eriksen, Michael Kemp, Bent Løwe Røder, Niels Frimodt-Møller, Turid Snekløth Søndergaard,



John Eugenio Coia, Claus Østergaard, Michael Pedersen, Henrik Westh, Frank Møller Aarestrup.

## References

1. World Health Organization. Global Health Estimates 2016 summary tables: global deaths by cause, age and sex, 2000–2016. Geneva, Switzerland: 2018.
2. World Health Organization. Report on the Burden of Endemic Health Care-Associated Infection Worldwide. Geneva, Switzerland: 2011.
3. Cassini A, Högberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis 2019; 19. [https://doi.org/10.1016/S1473-3099\(18\)30605-4](https://doi.org/10.1016/S1473-3099(18)30605-4) PMID: 30409683
4. Baker MG, Easther S, Wilson N. A surveillance sector review applied to infectious diseases at a country level. *BMC Public Health* 2010; 10:332. <https://doi.org/10.1186/1471-2458-10-332> PMID: 20540772
5. Cassini A, Colzani E, Pini A, Mangen MJ, Plass D, McDonald SA, et al. Impact of infectious diseases on population health using incidence-based disability-adjusted life years (DALYs): results from the Burden of Communicable Diseases in Europe study, European Union and European Economic Area countries, 2009 to 2013. *Eurosurveillance* 2018; 23:pii = 17–00454. <https://doi.org/10.2807/1560-7917.ES.2018.23.16.17-00454> PMID: 29692315
6. World Health Organization. Prioritization of pathogens to guide discovery, research and development of new antibiotics for drug-resistant bacterial infections, including tuberculosis. Geneva, Switzerland: 2017.
7. Furuse Y. Analysis of research intensity on infectious disease by disease burden reveals which infectious diseases are neglected by researchers 2019; 116. <https://doi.org/10.1073/pnas.1814484116> PMID: 30598444
8. Perry MR, McClean D, Simonet C, Woolhouse M, McNally L. Focussing on resistance to front-line drugs is the most effective way to combat the antimicrobial resistance crisis. *BioRxiv* 2018:498329. <https://doi.org/10.1101/498329>.
9. Diekema DJ, Hsueh P, Mendes RE, Pfaller MA, Rolston K V, Sader HS, et al. The Microbiology of Bloodstream Infection: 20-Year Trends from the SENTRY Antimicrobial Surveillance Program. *Antimicrob Agents Chemother* 2019; 63:e00355–19. <https://doi.org/10.1128/AAC.00355-19> PMID: 31010862
10. European Centre for Disease Prevention and Control. Point prevalence survey of healthcare-associated infections and antimicrobial use in European acute care hospitals, 2011–2012. Stockholm, Sweden: 2013.
11. Suetens C, Latour K, Kärki T, Ricchizzi E, Kinross P, Moro ML, et al. Prevalence of healthcare-associated infections, estimated incidence and composite antimicrobial resistance index in acute care hospitals and long-term care facilities: results from two European point prevalence surveys, 2016 to 2017. *Eurosurveillance* 2018; 23:pii = 1800516. <https://doi.org/10.2807/1560-7917.ES.2018.23.46.1800516> PMID: 30458912
12. Abebe M, Tadesse S, Meseret G, Derbie A. Type of bacterial isolates and antimicrobial resistance profile from different clinical samples at a Referral Hospital, Northwest Ethiopia: five years data analysis. *BMC Res Notes* 2019; 12:568. <https://doi.org/10.1186/s13104-019-4604-6> PMID: 31511033
13. Statens Serum Institut, National Food Institute. Danmap 2018. Denmark: 2018.
14. Mogensen CB, Skjöt-arkil H, Lassen AT, Johansen IS, Chen M, Petersen P, et al. Cross sectional study of multiresistant bacteria in Danish emergency departments: prevalence, patterns and risk factors for colonization (AB-RED project). *BMC Emerg Med* 2018; 18:25. <https://doi.org/10.1186/s12873-018-0178-1> PMID: 30126361
15. Skjöt-arkil H, Mogensen CB, Lassen AT, Johansen IS, Chen M, Petersen P, et al. Carrier prevalence and risk factors for colonisation of multiresistant bacteria in Danish emergency departments: a cross-sectional survey. *BMJ Open* 2019; 9:e029000. <https://doi.org/http://dx.doi.org/10.1136/bmjopen-2019-029000>. PMID: 31253624
16. Mogensen CB, Kjældgaard P, Jensen C, Chen M. MRSA screening in emergency department detects a minority of MRSA carriers. *Danish Medicin J* 2015; 62:A5151. PMID: 26522478
17. Rossen JWA, Friedrich AW, Moran-Gilad J. Practical issues in implementing whole-genome-sequencing in routine diagnostic microbiology. *Clin Microbiol Infect* 2018; 24:355–60. <https://doi.org/10.1016/j.cmi.2017.11.001> PMID: 29117578

18. Ellington MJ, Ekelund O, Aarestrup FM, Canton R, Doumith M, Giske C, et al. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. *Clin Microbiol Infect* 2017; 23:2–22. <https://doi.org/10.1016/j.cmi.2016.11.012> PMID: 27890457
19. Quainoo S, Coolen JPM, van Hijum SAFT, Huynen MA, Melchers WJG, van Schaik W, et al. Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis. *Clin Microbiol Rev* 2017; 30:1015–63. <https://doi.org/10.1128/CMR.00016-17> PMID: 28855266
20. Long SW, Williams D, Valson C, Cantu CC, Cernoch P, Musser JM, et al. A Genomic Day in the Life of a Clinical Microbiology Laboratory. *J Clin Microbiol* 2013; 51:1272–7. <https://doi.org/10.1128/JCM.03237-12> PMID: 23345298
21. Roach DJ, Burton JN, Lee C, Stackhouse B, Butler-Wu SM, Cookson BT, et al. A Year of Infection in the Intensive Care Unit: Prospective Whole Genome Sequencing of Bacterial Clinical Isolates Reveals Cryptic Transmissions and Novel Microbiota. *PLoS Genet* 2015; 11:e1005413. <https://doi.org/10.1371/journal.pgen.1005413> PMID: 26230489
22. Hasman H, Saputra D, Sicheritz-Ponten T, Lund O, Svendsen CA, Frimodt-Moller N, et al. Rapid Whole-Genome Sequencing for Detection and Characterization of Microorganisms Directly from Clinical Samples. *J Clin Microbiol* 2014; 52:139–46. <https://doi.org/10.1128/JCM.02452-13> PMID: 24172157
23. Schmidt K, Mwaigwisya S, Crossman LC, Doumith M, Munroe D, Pires C, et al. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J Antimicrob Chemother* 2017; 72:104–14. <https://doi.org/10.1093/jac/dkw397> PMID: 27667325
24. Larsen M V., Cosentino S, Lukjancenko O, Saputra D, Rasmussen S, Hasman H, et al. Benchmarking of Methods for Genomic Taxonomy. *J Clin Microbiol* 2014; 52:1529–39. <https://doi.org/10.1128/JCM.02981-13> PMID: 24574292
25. Clausen PTLC, Aarestrup FM, Lund O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics* 2018; 19:307. <https://doi.org/10.1186/s12859-018-2336-6> PMID: 30157759
26. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, et al. Ribosomal multilocus sequence typing: Universal characterization of bacteria from domain to strain. *Microbiology* 2012; 158:1005–15. <https://doi.org/10.1099/mic.0.055459-0> PMID: 22282518
27. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012; 67:2640–4. <https://doi.org/10.1093/jac/dks261> PMID: 22782487
28. Tindall BJ, Sutton G, Garrity GM. *Enterobacter aerogenes* Hormaeche and Edwards 1960 (Approved Lists 1980) and *Klebsiella mobilis* Bascomb et al. 1971 (Approved Lists 1980) share the same nomenclatural type (ATCC 13048) on the Approved Lists and are homotypic synonyms, with consequences for the name *Klebsiella mobilis* Bascomb et al. 1971 (Approved Lists 1980) 2017;1960:502–4. <https://doi.org/10.1099/ijsem.0.001572>.
29. Scholz CFP, Kilian M. The natural history of cutaneous propionibacteria, and reclassification of selected species within the genus *Propionibacterium* to the proposed novel genera *Acidipropionibacterium* gen. nov., *Cutibacterium* gen. nov. and *Pseudopropionibacterium* gen. nov. *Int J Syst Evol Microbiol* 2016; 66:4422–32. <https://doi.org/10.1099/ijsem.0.001367> PMID: 27488827
30. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattor V, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 2020; 75(12):3491–3500. <https://doi.org/10.1093/jac/dkaa345> PMID: 32780112