bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Molecular diagnosis of orthopaedic device infection direct from sonication fluid by metagenomic sequencing

Teresa L. Street,[a]# Nicholas D. Sanderson,[a] Bridget L. Atkins,[b,c] Andrew J. Brent,[a,b] Kevin Cole,[d,e] Dona Foster,[a] Martin A. McNally,[b] Sarah Oakley,[c] Leon Peto,[a] Adrian Taylor,[b] Tim E. A. Peto,[a,f] Derrick W. Crook,[a,f] David W. Eyre[a,f]

Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK[a]

Bone Infection Unit, Nuffield Orthopaedic Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, UK[b]

Microbiology Laboratory, John Radcliffe Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, UK[c]

Department of Infectious Diseases and Microbiology, Royal Sussex County Hospital, Brighton, UK[d]

Public Health England, Microbiology, Royal Sussex County Hospital, Brighton, UK[e]

National Institute for Health Research Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK[f]

Running title: Metagenomic sequencing for orthopaedic device infection

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

18 # Address correspondence to Teresa L. Street, **teresa.street@ndm.ox.ac.uk**

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Abstract

Culture of multiple periprosthetic tissue samples is the current gold-standard for microbiological diagnosis of prosthetic joint infections (PJI). Additional diagnostic information may be obtained through sonication fluid culture of explants. However, current techniques can have relatively low sensitivity, with prior antimicrobial therapy and infection by fastidious organisms influencing results. We assessed if metagenomic sequencing of complete bacterial DNA extracts obtained direct from sonication fluid can provide an alternative rapid and sensitive tool for diagnosis of PJI.

We compared metagenomic sequencing with standard aerobic and anaerobic culture in 97 sonication fluid samples from prosthetic joint and other orthopaedic device infections. Reads from Illumina MiSeq sequencing were taxonomically classified using Kraken. Using 50 samples (derivation set), we determined optimal thresholds for the number and proportion of bacterial reads required to identify an infection and validated our findings in 47 independent samples.

Compared to sonication fluid culture, the species-level sensitivity of metagenomic sequencing was 61/69(88%,95%CI 77-94%) (derivation samples 35/38[92%,79-98%]; validation 26/31[84%,66-95%]), and genus-level sensitivity was 64/69(93%,84-98%). Species-level specificity, adjusting for plausible fastidious causes of infection, species found in concurrently obtained tissue samples, and prior antibiotics, was 85/97(88%,79-93%) (derivation 43/50[86%,73-94%], validation 42/47[89%,77-96%]). High levels of human

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

DNA contamination were seen despite use of laboratory methods to remove it. Rigorous laboratory good practice was required to prevent bacterial DNA contamination.

We demonstrate metagenomic sequencing can provide accurate diagnostic information in PJI. Our findings combined with increasing availability of portable, random-access sequencing technology offers the potential to translate metagenomic sequencing into a rapid diagnostic tool in PJI.

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

Prosthetic joint infections (PJI) are a devastating and difficult to treat complication of joint replacement surgery. Although the relative incidence of PJI is low (0.8% knee and 1.2% hip replacements across Europe)(1), given the increasing numbers of arthroplasties performed worldwide it is a significant healthcare burden and cause of expense. For individual patients, PJI often requires multiple surgeries, intensive, long-term antimicrobial therapy, and a prolonged period of rehabilitation. Fast, accurate and reliable diagnosis of PJI is necessary to inform treatment choices, particularly for antibiotic resistant organisms. Culture of multiple periprosthetic tissue (PPT) samples remains the gold-standard method of microbial detection (2-4). However, culture can be relatively insensitive with only 65% of causative bacteria detected in infections even when multiple PPT samples are collected (2, 5). Infections with fastidious organisms or where a patient has received prior antimicrobial treatment are often culture-negative.

Culture of sonication fluid from explanted prostheses may improve microbiological yield in PJI, by disrupting bacterial biofilm. Since sonication was first applied to explanted hip prostheses in 1998 (6) several clinical studies have reported improved sensitivity of sonication fluid culture over PPT culture for the diagnosis of hip, knee and shoulder PJI (7, 8), and sonication has been adopted by many centers, either alone or in combination with PPT culture. Additionally, several molecular assays have been investigated to improve the sensitivity of PJI diagnosis. PCR assays using DNA extracted from sonication fluid (9-12) have reported sensitivity ranging from 70% to 96%. However, this approach can only identify pathogens in a pre-defined multiplex panel, thus may miss atypical or rare

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

pathogens not targeted in the assay design. Other studies identify pathogens by amplification and sequencing of the universal bacterial 16S ribosomal RNA gene (13, 14). A drawback of these methods is the potential for generating false-positive results from contaminating bacterial DNA.

The potential of high-throughput sequencing as a diagnostic tool for infectious diseases is widely recognized (15-17). Metagenomic sequencing offers the possibility to detect all DNA in a clinical sample, which can be compared to reference genome databases to identify pathogens. Additionally, a profile of common laboratory and kit contaminants can be generated from negative controls sequenced concurrently and accounted for (18, 19). In addition to diagnostic data, whole-genome sequencing can also simultaneously provide characterization of infection outbreaks (20, 21), tracking of transmission (22-24) and prediction of antimicrobial resistance (25-28). An advantage offered by sequencing is the speed at which it can deliver genetic information (29) compared to traditional microbiological culture and antimicrobial susceptibility testing, which can take days to weeks depending on the pathogen. By removing a culture step and sequencing directly from clinical samples the time taken to diagnosis can be reduced further (30) and pathogens not identified by conventional methods can be detected (31-33). Here, we investigated if metagenomic sequencing of complete bacterial DNA extracts obtained direct from sonication fluid can provide an alternative rapid and sensitive tool for diagnosis of PJI, without the need for a culture step.

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Materials and Methods

Sample collection and processing. Intra-operative samples from the Nuffield Orthopaedic Centre (NOC) in Oxford University Hospitals (OUH), UK, between June 2013 and January 2017 were investigated. The NOC is a tertiary level specialist musculoskeletal hospital, including a dedicated Bone Infection Unit, undertaking approximately 200 revision arthroplasties annually. A subset of samples submitted were chosen at random following culture to provide a ratio of approximately 2:1 bacterial culture-positive samples to culture-negative samples.

Prosthetic joint implants and metalwork, received into the OUH microbiology laboratory following revision arthroplasty and operative management of other orthopaedic device related infection, were placed directly into single-use sterile polypropylene containers (Lock & Lock brand) and covered with between 10ml and 400ml of sterile 0.9% saline solution (Oxoid Ltd, Basingstoke, UK) depending on the size of the prosthesis/device, with sufficient fluid to cover at least 90% the prosthesis/device, up to a maximum of 400ml. Sonication was performed as described previously (7) with minor modifications. Briefly, the implant was vortexed for 30 seconds, subjected to sonication for 1 minute followed by additional vortexing for 30 seconds. Sonication was performed in a Bransonic 5510 ultrasonic water bath (Branson, Danbury, CT, USA) at a frequency of 40kHz. The resulting sonication fluid was plated in 0.1ml aliquots onto Columbia blood (CBA) and chocolate agar plates for aerobic incubation and CBA plates for anaerobic incubation. Aerobic incubation was performed at 35-37°C with 5% CO_2 for up to 5 days. Anaerobic incubation was performed at 35-37°C for 10 days. All cultured microorganisms were identified by MALDI-

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

TOF on a Microflex LT using Biotyper v.3.1 (Bruker Daltonics, Billerica, MA, USA). Samples were recorded as culture-positive where ≥50 CFU/ml were observed, and additionally at <50 CFU/ml for organisms considered highly pathogenic (including *Staphylococcus aure*us and Enterobacteriaceae.).

Periprosthetic tissue samples were also collected during surgery and processed by the microbiology laboratory. Briefly, BACTEC bottles were inoculated with 0.5ml of an inoculum generated by vortexing each tissue sample in 3ml 0.9% saline with Ballotini balls for 15 seconds. Bottles were incubated under aerobic (Plus Aerobic/F culture vials) and anaerobic (Lytic/10 Anaerobic/F culture vials) conditions in a BD BACTEC FX system (BD Biosciences, Sparks, MD, USA) for up to 10 days. Any bottles that flagged positive were sub-cultured onto agar plates and processed as described above to determine species.

Bacterial DNA extraction from sonication fluid. Prior to DNA extraction sonication fluids were concentrated by centrifugation. 40ml of fluid was transferred to a sterile, disposable 50ml polypropylene tube and centrifuged at 15,000 × g in a Sorvall RC5C Plus centrifuge (SLA-1500 rotor with custom made inserts) for 1 hour at 16°C. Samples with <40ml starting volume of sonication fluid were made up to 40ml with the same saline used for sonication. All but approximately 1ml of the supernatant was discarded, and the pellet resuspended in this volume of fluid before being passed through a 5µm syringe filter to deplete the number of human cells present, and therefore the amount of human DNA in the final extract. Bacterial cells passing through the filter were pelleted, washed with 1ml 0.9% saline and resuspended in 500µl molecular biology grade water before being mechanically lysed in Pathogen Lysis tubes (S) (Qiagen, Hilden, Germany) with a Fastprep-24 tissue

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

133 homogeniser (MP Biomedicals, Santa Ana, CA, USA) (3 x 40 seconds at 6.5 m/s). DNA was
134 extracted by ethanol precipitation, using GlycoBlue (Life Technologies, Paisley, UK) as a co-
135 precipitant, and resuspended in 50µl 1 x Tris-EDTA (TE) buffer. DNA was purified using
136 AMPure XP solid phase reversible immobilisation (SPRI) beads (Beckman Coulter, High
137 Wycombe, UK) and eluted in 26µl TE buffer. DNA concentration was measured using a
138 Qubit 2.0 fluorometer (Life Technologies, Paisley, UK). A subset of samples were treated
139 with the NEBNext microbiome DNA enrichment kit (New England Biolabs, Ipswich, MA,
140 USA) for human DNA removal, before an additional purification using AMPure XP SPRI
141 beads and final elution in 15µl TE buffer. Samples were extracted in batches, with a
142 negative control of sterile 0.9% saline prepared alongside each batch using this same
143 protocol.

144 **Library preparation and Illumina MiSeq sequencing.** DNA extracts quantified as
145 ≥0.2ng/µl were sequenced on a MiSeq desktop sequencer (Illumina, San Diego, CA, USA).
146 Libraries were prepared as previously described, using a variation of the Illumina Nextera
147 XT protocol (34). Briefly, 1ng of DNA was prepared for sequencing following the Illumina
148 Nextera XT protocol, with the modification of 15 cycles during the index PCR. Libraries
149 were quantified using a Qubit 2.0 fluorometer and their average size determined with an
150 Agilent 2200 Tapestation (Agilent Technologies, Santa Clara, CA, USA) before being
151 manually normalised. Libraries were prepared and sequenced together in the same batch.
152 Paired-end sequencing was performed using a 600-cycle MiSeq reagent kit v3 and samples
153 were sequenced in batches of between 1 and 13 on a single flow cell.

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Bioinformatics analysis. Raw sequencing reads were adapter trimmed using BBDuk (https://sourceforge.net/projects/bbmap/) with the following parameters: minlength=36, k=19, ktrim=r, hdist=1, mink=12 and the adapter sequence file provided within the bbmap package. Taxonomic classification of trimmed reads was performed using Kraken (35) and a bespoke database constructed from all bacterial genomes deposited in the NCBI RefSeq database as of January 2015 (updated January 2017 for validation set, see below), with default parameters and no K-mer removals. Where no refseq genome was available for an organism cultured from a PJI at OUH since June 2013, available whole-genome assemblies were also added to the database where available in NCBI. Additionally, the human genome GRCh38 was included in the database to allow detection of host DNA. An optimum filtration threshold, using Kraken-filter, that balanced false-positive removal and sensitivity was determined using simulated datasets of reference genomes. Reference genomes representative of common pathogenic species were used to generate simulated Illumina MiSeq datasets and analysed with Kraken using different filtration thresholds. A threshold value of 0.15 provided optimum read classification sensitivity whilst minimizing spurious results. Kraken output was visualised using Krona (36).

Statistical analysis.

In order to correct for samples which may contain small numbers of contaminating and non-specific bacterial reads, a threshold was determined to identify the presence of true infection, using the first 50 samples sequenced as a derivation set. Two thresholds (1 and 2), and three parameters (a-c), were used to determine true infection: 1) samples with more reads from a given species than an upper read cut-off (a) were included; 2) samples

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

with more species-specific reads than a lower read cut-off (b) and with the percentage of species-specific reads as a proportion of all bacterial reads present above a percentage cut-off (c) were also included. Parameter values were selected by numerical optimisation, using R version 3.3.2, comparing sequencing results to sonication fluid culture results, and maximising the value of the Youden Index (37) (sensitivity + specificity - 1). Sensitivity was calculated taking each species identified from each culture-positive sonication sample as a separate data point. Specificity was calculated using the total number of sonication samples as the denominator; as such samples contaminated by more than one species were counted as one false-positive.

To ensure that read cut-off parameters were chosen without a penalty for potentially difficult to culture anaerobic species, the specificity value optimised was adjusted. Potential 'false-positive' sequencing results with plausible fastidious anaerobic causes of infection (including *Fusobacterium nucleatum*, *Propionibacterium acnes* and *Veillonella parvula*) in culture-negative samples were excluded when calculating the specificity value used for parameter optimisation.

Where bacterial reads were detected over the thresholds described above in a negative control, that sample was deemed to be contaminated. In the derivation set, in order to maximise the number of sequences available for analysis, only samples with evidence of the same contaminating organisms were excluded from each contaminated batch, rather than discarding the whole batch. During the derivation phase of the study, several batches of samples were found to be contaminated with DNA from other studies performed concurrently in the same research laboratory. Six of eight saline negative control extracts

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

displayed contamination with a single or multiple species at read numbers exceeding the determined diagnostic thresholds. All samples within these batches that displayed similar contamination were excluded from subsequent analysis if Kraken classification resulted in >100 reads corresponding to the majority of the contaminating species. A total of 22 samples (in addition to the 50 successfully sequenced) were excluded on this basis (Figure 1). In batches 4 and 5 the negative controls were contaminated with *Staphylococcus aureus*, *Escherichia coli*, and *P. acnes*, and 15 samples were excluded with >100 reads from ≥2/3 species; in batch 6 the negative control was contaminated with *Serratia marcescens*, *Klebsiella pneumoniae, E. coli*, and *P. acnes*, and 2 samples with >100 reads from ≥3/4 species were excluded; in batches 2, 9 and 10 the negative control was contaminated with *P. acnes*, and 5 samples were excluded with >100 *P. acnes* reads. To address this issue, prior to the validation phase of the study, all pipettes, laminar flow and PCR hoods and laboratory benches used for DNA extraction and library preparation were deep-cleaned with Virkon disinfectant and RNase AWAY surface decontaminant (Thermo Fisher Scientific, Waltham, MA, USA) in order to remove any possible sources of microbial or DNA contamination. All DNA extraction and library preparation reagents were replaced, and used in pre-prepared per-batch aliquots used exclusively for this study. Sonication fluid samples were handled one at a time in the laminar flow hood, which was cleaned as above between each sample. Fresh gloves were worn each time a new sample was handled during the DNA extraction phase of the protocol. Having implemented these changes, for the validation phase, a more stringent quality control standard was applied, requiring the negative control to be contamination-free for any of the samples in a batch to be analyzed.

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Technical replicates. To ensure sequencing reproducibility one DNA sample was sequenced twice and biological replicates (DNA extraction process repeated) were sequenced for six samples (four in duplicate and two in triplicate). Samples extracted and sequenced as replicates showed good reproducibility. In four duplicate and one triplicate culture-positive samples the same species was recovered by WGS on all occasions (samples 164, 171, 182, 183, and 193). A single replicate, 182a, had an additional, likely contaminating, species identified (not found in sonication fluid or PPT culture). A single culture-negative sample (176) was processed in triplicate. One of the three replicates (176a) had an apparent contaminating species identified (also not found in sonication fluid or PPT culture).

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Results

A total of 131 sonication fluids from patients undergoing revision arthroplasty or removal of other orthopaedic devices, were aerobically and anaerobically cultured and underwent metagenomic sequencing (Figure 1). Additionally, a median (IQR) [range] 5 (4-5) [1-8] PPT samples were cultured from each patient. *S. aureus*, isolated from 22% of sonication fluids and 29% PPTs, and *Staphylococcus epidermidis*, from 16% sonication fluids and 25% of PPTs, were the 2 most frequently cultured species (Table 1).

From the first 72 sonication fluid samples sequenced 22 samples from six batches were excluded, as these samples and negative controls from the same batches showed similar contamination (see Methods, Figure 1). The remaining 50 samples, the derivation set, were used to determine optimal sequence thresholds for identifying true infection. Of 59 subsequently sequenced validation samples, 12 from a single batch were excluded as the negative control was contaminated with *P. acnes*, leaving 47 validation samples sequenced in batches with uncontaminated negative controls.

The 97 sonication fluid samples passing sequencing quality control checks were obtained predominantly from knee (42/97, 43%) and hip (32, 33%) PJI, with other samples from ankle (6, 6%), and shoulder (3, 3%) PJI, and other orthopaedic device infection (14, 14%) (Supplementary Table 2) . The median (IQR) [range] sonication fluid volume was 200ml (100-400ml) [15-400ml] (Supplementary Table 2). On culture, 35 (36%) sonication fluid samples had no growth, or less than <50 CFU of an organism not considered to be highly pathogenic (skin and oral flora), 55 (57%) had a single organism isolated, and 7 (7%) two

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

252 organisms isolated. Greater than 10^6 reads were achieved in 91/97 (94%) samples.
253 Taxonomic classification by Kraken identified a median (IQR) [range] 0.72% (0.01-0.41%)
254 [<0.01% - 24.0%] of reads as bacterial, with <1% bacterial reads in 84/97 (87%) samples.
255 Human reads accounted for >90% of reads in 94/97 (97%) of samples. Six test samples
256 were processed with and without the NEBNext microbiome DNA enrichment kit. Use of the
257 kit did not reduce the amount of human DNA sequenced. The mean proportion of reads
258 classified as human with the enrichment kit was 98.4%, and without it 98.2% (p=0.06,
259 Supplementary Table 1).

260 Optimal thresholds for determining if samples contain low-level contamination or true
261 infection were determined by numerical optimisation, choosing thresholds that maximised
262 the sensitivity and specificity of sequencing (Figure 2). The final thresholds chosen to
263 determine the presence of true infection were ≥1150 reads from a single species, or ≥125
264 reads from a single species if ≥15% of the total bacterial reads also belong to that same
265 species.

266 Table 2 compares sonication culture results with metagenomic sequencing findings
267 applying these thresholds. PPT culture results, and the consensus microbiology diagnosis
268 based on both sonication and PPT samples are also given for comparison. Compared to
269 sonication fluid culture, metagenomic sequencing had an overall species-level sensitivity of
270 61/69 (88%, 95%CI, 77-94%). Sensitivity in the derivation samples was 35/38 (92%,79-
271 98%), and in the validation samples 26/31 (84%, 66-95%). Three samples were identified
272 to the genus level only. Hence overall genus-level sensitivity was 64/69 (93%,84-98%). Of
273 the other five samples where the species cultured was not identified on sequencing, two

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

274 samples cultured a coagulase-negative *Staphylococcus*, not identified on tissue culture, one
275 sample was polymicrobial (where several species found in sonication fluid or tissue were
276 identified, but not all) and the remaining two samples failed to detect a pathogen found in
277 sonication and tissue fluid.

278 Overall species level specificity was 78/97 (80%,71-88%). However, of 19 samples where
279 additional species were identified on sequencing compared to sonication culture, three had
280 the same species found in tissue culture (but not in sonication fluid, or at <50 CFU; samples
281 400, 414, 502). Four samples had plausible anaerobic causes of infection (*Fusobacterium*
282 *nucleatum*, *Veillonella parvula*, *Finegoldia magna*, *Parvimonas micra* [identified alongside
283 *Streptococcus anginosus*]) (samples 354, 369, 400, 485). Samples 341 and 475 contained *S.*
284 *aureus* and *Streptococcus dysgalactiae* DNA respectively, both in patients who had received
285 prior flucloxacillin and no microbiological diagnosis was reached based on culture.
286 However, 12 samples (including sample 485) had other species found on sequencing not
287 otherwise identified. In some cases these were clearly laboratory contaminants, e.g. sample
288 219 contained *Achromobacter xylosoxidans* reads and an *A. xylosoxidans* culture-positive
289 sample was sequenced in the same batch from a concurrent study. Notably *P. acnes* was a
290 common contaminant occurring in 7/97 (7%) samples overall. Adjusting for plausible
291 fastidious causes of infection, species found in concurrently obtained PPT samples, and
292 prior antibiotics, i.e. assuming these samples were actually genuinely positive for the
293 species found on sequencing, overall species-level specificity was 85/97 (88%,79-93%),
294 43/50 (86%,73-94%) in the derivation samples and 42/47 (89%,77-96%) in the validation
295 samples.

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 3 shows the relationship between the proportion of sequence reads obtained that were classified as bacterial, the sonication fluid culture CFU counts, and the concordance between sonication fluid culture and sequencing. Sequencing false positive results were more likely when cultures were negative.

More simplistic thresholds for determining true infection performed less well. Within the derivation set, using a single cut-off for the proportion of bacterial reads from a given species, irrespective of the absolute numbers of bacterial reads present, the optimal cut-off value was 25%. Using this threshold, species-level sensitivity was 30/38 (79%) and adjusted specificity 44/50 (88%). Similarly, if only a single absolute read number cut-off is used the optimal value is 410 reads from a single species, sensitivity 30/38 (79%) and adjusted specificity 45/50 (90%).

Sequencing results were also compared to a consensus microbiology diagnosis based on IDSA guidelines (4), considering any species isolated twice or any virulent species isolated as a cause of infection, combining sonication and PPT culture results (Supplementary Table 2). 65/97 (67%) samples showed complete agreement between the consensus species list from culture and sequencing, 15/97 (15%) a partial match with at least one species found on culture also found on sequencing, 15/97 (15%) had none of the species cultured found on sequencing, and 2/97 (2%) had a plausible additional species found on sequencing not found on culture.

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Discussion

Diagnosis of PJI by culture of sonication fluid and PPT is not always conclusive, and may take up to 10-14 days for slow-growing organisms. Here we assess, for the first time, the use of metagenomic sequencing of complete bacterial DNA extracts obtained direct from sonication fluid in the diagnosis of PJI. We develop a novel filtering strategy to ensure that low-level contaminating DNA is successfully ignored, while infections are detected accurately. Compared to sonication fluid culture, metagenomic sequencing achieved a species-level sensitivity of 88% and specificity of 87% after adjusting for plausible fastidious causes of infection, species found in concurrently obtained PPT samples, and prior antibiotics. Importantly we demonstrate similar performance of our method and filtering algorithm in the subset of samples that formed an independent validation set, sensitivity 84%, adjusted specificity 89%.

Sequencing failed to identify an organism cultured from sonication fluid for eight samples. For two samples a coagulase-negative *Staphylococcus* was cultured, but only from sonication fluid and not from tissue samples. These isolates therefore could plausibly have been plate contaminants, and not present in the DNA sequenced. For three other samples identification to the genus-level was possible. One sample contained *Staphylococcus condimenti* which was not included in our custom Kraken database, highlighting that despite including 2786 bacterial genomes this approach is only as good as the database that is used. Another sample was identified as a *Bacillus spp.* both on culture and by sequencing, and the third by sequencing as *Staphylococcus spp.* in the context of a mixed

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

337 *Staphylococcus* infection. For the three remaining samples, sequencing failed to identify a
338 pathogen found on culture.

339 Sequencing was also able to detect potential pathogens not identified by sonication fluid
340 culture. For three samples we identified additional species from sequencing that were
341 supported by the tissue culture findings, suggesting in some settings sequencing may be
342 more sensitive that sonication fluid culture alone without PPT culture. We also identified
343 four examples of probable anaerobic pathogens not identified by routine anaerobic culture
344 of sonication fluid or PPT: *Fusobacterium nucleatum*, *Veillonella parvula*, *Finegoldia magna*,
345 *Parvimonas micra*, and we were also able to identify a plausible pathogen in two patients
346 who had received prior antibiotics where the routine microbiology was uninformative.

347 Controlling for contamination during sampling and culture is a major challenge in
348 investigating PJI, and underlies why multiple independent PPT samples remains the gold
349 standard for diagnosis. Contamination is an even greater concern in molecular diagnostics
350 given the additional potential for DNA contamination. There are published examples
351 demonstrating the potential for contamination leading to misinterpretation of sequencing
352 data from clinical specimens (38, 39). In our laboratory samples were handled in laminar
353 flow hoods and extracted in a dedicated pre-PCR extraction laboratory. DNA was handled
354 in a PCR hood and sequencing libraries were manipulated in a dedicated post-PCR
355 sequencing laboratory. Despite these measures, we still observed contamination in some of
356 our samples. During the derivation phase of our study it is likely that one or more of the
357 reagents used became contaminated with DNA from other sequencing projects in our
358 laboratory. Although we were able to account for this in our analysis, and then validate our

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

findings in a separate set of samples having resolved this issue, it demonstrates that rigorous laboratory practice would be key to deploying our method. There may also be a role for sealed systems that perform DNA extraction and sequencing in a separated environment. Our experience also re-enforces the requirement that negative controls are included in each sequencing batch, as is routine in molecular microbiology diagnostic assays, to ensure contamination is detected if it does occur. A limitation of our study is that the saline used for sonication was not PCR-grade, and this could be considered in future work.

Despite addressing the major contamination issues in the derivation phase, contamination with *P. acnes* remained an issue in one of our validation batches. Overall, false positive results for *P. acnes* were found in 7% of samples. Species-specific filtering may be required to address this, our one true-positive sample with *P. acnes* present on culture had $>10^5$ *P. acnes* reads. However, larger datasets are required than ours to address this definitively. In the mean-time, even with molecular diagnostics, the value of multiple samples per patient remains.

Sonication fluid can be a large volume sample, typically 50-400ml. As a result, the microbial cells released from the orthopaedic device during sonication are likely to be heavily diluted. This, coupled with the simultaneous release of any human cells from the prosthesis and transfer of blood along with the device, results in a sonication fluid sample that is both low in bacterial cells and high in contaminating host cells. An effective microbial DNA extraction protocol is necessary to isolate as much bacterial DNA as possible, while limiting the amount of host DNA in the final extract. Our results demonstrate that despite efforts to

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

filter out human cells, or remove human DNA post-extraction, host DNA accounted for >90% of reads in the majority of samples sequenced. Use of a specialist microbiome enrichment kit did not improve bacterial DNA yield. However, if the efficiency of human DNA removal can be improved in future this might significantly add to the precision of metagenomic sequencing as more sequencing effort would be appropriately directed towards potential pathogens.

In addition to the issues around contamination with bacterial and human DNA, a further limitation of our study as designed is that it undertakes a laboratory-level comparison of sonication fluid culture and metagenomics sequencing. We did not have ethical approval to review patient notes and so were unable to compare sonication fluid sequencing to the presence of a final overall diagnosis of infection. Future studies should consider how sequencing might contribute to the overall diagnosis of PJI, as part of assessment that jointly considers clinical, histological, and microbiological data.

This study demonstrates as a proof of principle that metagenomic sequencing can be used in the culture-free diagnosis of PJI directly from sonication fluid. Improvements to the method of human DNA removal from direct samples before sequencing are ongoing, and if these are successful, this is likely to greatly improve the efficiency, and therefore accuracy, of metagenomic sequencing. Generating greater numbers of bacterial reads direct from clinical specimens may make prediction of antimicrobial susceptibilities direct from samples possible, as has been achieved from whole-genome sequencing of cultured organisms (25-28). If this can be achieved reliably, it is possible that sequencing can offer a complete microbiology diagnosis without the need for culture. The increasing availability

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

of portable, rapid, random-access strand sequencing technology offers the potential that in future sequencing may become a same-day diagnostic tool. Applications of rapid sequencing in PJI might include perioperative microbiological diagnosis to guide local intraoperative antimicrobials, for example in cement or beads. Earlier diagnosis may also ensure post-operative antimicrobials are more focused, improving antimicrobial stewardship, while treating resistant organisms effectively. Earlier diagnosis may also reduce hospital stays and therefore reduce costs. Sequencing is also likely to be helpful in situations where multiple samples containing the same commensal species are identified. Sequencing will be able to determine whether these are clonal, suggesting true infection rather than contamination, instead of having to rely on current proxies such as antibiograms, which only imperfectly distinguish non-clonal isolates. Ultimately, same-day sequencing may significantly improve the precision, efficiency and cost of PJI care. This study provides a foundation for further development towards this goal.

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Funding Information

This work is supported by National Institute for Health Research Oxford Biomedical Research Centre. DWC and TEAP are NIHR senior investigators. DWE is a NIHR clinical lecturer.

Acknowledgements

The authors thank the microbiology laboratory staff of the John Radcliffe Hospital, Oxford University Hospitals NHS Foundation Trust, for providing assistance with sample collection and processing.

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

1. **Lamagni T.** 2014. Epidemiology and burden of prosthetic joint infections. J Antimicrob Chemother **69 Suppl 1:**i5-10.

2. **Atkins BL, Athanasou N, Deeks JJ, Crook DW, Simpson H, Peto TE, McLardy-Smith P, Berendt AR.** 1998. Prospective evaluation of criteria for microbiological diagnosis of prosthetic-joint infection at revision arthroplasty. The OSIRIS Collaborative Study Group. J Clin Microbiol **36:**2932-2939.

3. **Parvizi J, Zmistowski B, Berbari EF, Bauer TW, Springer BD, Della Valle CJ, Garvin KL, Mont MA, Wongworawat MD, Zalavras CG.** 2011. New definition for periprosthetic joint infection: from the Workgroup of the Musculoskeletal Infection Society. Clin Orthop Relat Res **469:**2992-2994.

4. **Osmon DR, Berbari EF, Berendt AR, Lew D, Zimmerli W, Steckelberg JM, Rao N, Hanssen A, Wilson WR.** 2013. Diagnosis and Management of Prosthetic Joint Infection: Clinical Practice Guidelines by the Infectious Diseases Society of America. Clinical Infectious Diseases **56**.

5. **Bejon P, Berendt A, Atkins BL, Green N, Parry H, Masters S, McLardy-Smith P, Gundle R, Byren I.** 2010. Two-stage revision for prosthetic joint infection: predictors of outcome and the role of reimplantation microbiology. J Antimicrob Chemother **65:**569-575.

6. **Tunney MM, Patrick S, Gorman SP, Nixon JR, Anderson N, Davis RI, Hanna D, Ramage G.** 1998. Improved detection of infection in hip replacements. A currently underestimated problem. J Bone Joint Surg Br **80:**568-572.

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

449 7. **Trampuz A, Piper KE, Jacobson MJ, Hanssen AD, Unni KK, Osmon DR,**
450 **Mandrekar JN, Cockerill FR, Steckelberg JM, Greenleaf JF, Patel R.** 2007.
451 Sonication of removed hip and knee prostheses for diagnosis of infection. N Engl J
452 Med **357:**654-663.
453 8. **Piper KE, Jacobson MJ, Cofield RH, Sperling JW, Sanchez-Sotelo J, Osmon DR,**
454 **McDowell A, Patrick S, Steckelberg JM, Mandrekar JN, Sampedro MF, Patel R.**
455 2009. Microbiologic Diagnosis of Prosthetic Shoulder Infection by Use of Implant
456 Sonication. Journal of Clinical Microbiology **47:**1878-1884.
457 9. **Achermann Y, Vogt M, Leunig M, Wust J, Trampuz A.** 2010. Improved Diagnosis
458 of Periprosthetic Joint Infection by Multiplex PCR of Sonication Fluid from Removed
459 Implants. Journal of Clinical Microbiology **48:**1208-1214.
460 10. **Gomez E, Cazanave C, Cunningham SA, Greenwood-Quaintance KE, Steckelberg**
461 **JM, Uhl JR, Hanssen AD, Karau MJ, Schmidt SM, Osmon DR, Berbari EF,**
462 **Mandrekar J, Patel R.** 2012. Prosthetic joint infection diagnosis using broad-range
463 PCR of biofilms dislodged from knee and hip arthroplasty surfaces using sonication.
464 J Clin Microbiol **50:**3501-3508.
465 11. **Cazanave C, Greenwood-Quaintance KE, Hanssen AD, Karau MJ, Schmidt SM,**
466 **Gomez Urena EO, Mandrekar JN, Osmon DR, Lough LE, Pritt BS, Steckelberg JM,**
467 **Patel R.** 2013. Rapid molecular microbiologic diagnosis of prosthetic joint infection.
468 J Clin Microbiol **51:**2280-2287.
469 12. **Ryu SY, Greenwood-Quaintance KE, Hanssen AD, Mandrekar JN, Patel R.** 2014.
470 Low sensitivity of periprosthetic tissue PCR for prosthetic knee infection diagnosis.
471 Diagnostic Microbiology and Infectious Disease **79:**448-453.

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

13. **Fenollar F, Roux V, Stein A, Drancourt M, Raoult D.** 2006. Analysis of 525 samples to determine the usefulness of PCR amplification and sequencing of the 16S rRNA gene for diagnosis of bone and joint infections. J Clin Microbiol **44:**1018-1028.

14. **Tunney MM, Patrick S, Curran MD, Ramage G, Hanna D, Nixon JR, Gorman SP, Davis RI, Anderson N.** 1999. Detection of prosthetic hip infection at revision arthroplasty by immunofluorescence microscopy and PCR amplification of the bacterial 16S rRNA gene. Journal of Clinical Microbiology **37:**3281-3290.

15. **Didelot X, Bowden R, Wilson DJ, Peto TE, Crook DW.** 2012. Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet **13:**601-612.

16. **Lefterova MI, Suarez CJ, Banaei N, Pinsky BA.** 2015. Next-Generation Sequencing for Infectious Disease Diagnosis and Management A Report of the Association for Molecular Pathology. Journal of Molecular Diagnostics **17:**623-634.

17. **Goldberg B, Sichtig H, Geyer C, Ledeboer N, Weinstock GM.** 2015. Making the Leap from Research Laboratory to Clinic: Challenges and Opportunities for Next-Generation Sequencing in Infectious Disease Diagnostics. Mbio **6**.

18. **Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW.** 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol **12:**87.

19. **Laurence M, Hatzis C, Brash DE.** 2014. Common Contaminants in Next-Generation Sequencing That Hinder Discovery of Low-Abundance Microbes. Plos One **9**.

20. **Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, Ip CLC, Wilson DJ, Didelot X, O'Connor L, Lay R, Buck D, Kearns AM, Shaw A, Paul J,**

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Wilcox MH, Donnelly PJ, Peto TEA, Walker AS, Crook DW. 2012. A pilot study of rapid benchtop sequencing of Staphylococcus aureus and Clostridium difficile for outbreak detection and surveillance. Bmj Open **2**.

21. **Reuter S, Ellington MJ, Cartwright EJ, Koser CU, Torok ME, Gouliouris T, Harris SR, Brown NM, Holden MT, Quail M, Parkhill J, Smith GP, Bentley SD, Peacock SJ.** 2013. Rapid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology. JAMA Intern Med **173:**1397-1404.

22. **Didelot X, Eyre DW, Cule M, Ip CLC, Ansari MA, Griffiths D, Vaughan A, O'Connor L, Golubchik T, Batty EM, Piazza P, Wilson DJ, Bowden R, Donnelly PJ, Dingle KE, Wilcox M, Walker AS, Crook DW, Peto TEA, Harding RM.** 2012. Microevolutionary analysis of Clostridium difficile genomes to investigate transmission. Genome Biology **13**.

23. **Eyre DW, Fawley WN, Best EL, Griffiths D, Stoesser NE, Crook DW, Peto TE, Walker AS, Wilcox MH.** 2013. Comparison of multilocus variable-number tandem-repeat analysis and whole-genome sequencing for investigation of Clostridium difficile transmission. J Clin Microbiol **51:**4141-4149.

24. **Walker TM, Lalor MK, Broda A, Ortega LS, Morgan M, Parker L, Churchill S, Bennett K, Golubchik T, Giess AP, Elias CD, Jeffery KJ, Bowler ICJW, Laurenson IF, Barrett A, Drobniewski F, McCarthy ND, Anderson LF, Abubakar I, Thomas HL, Monk P, Smith EG, Walker AS, Crook DW, Peto TEA, Conlon CP.** 2014. Assessment of Mycobacterium tuberculosis transmission in Oxfordshire, UK, 2007-12, with whole pathogen genome sequences: an observational study. Lancet Respiratory Medicine **2:**285-292.

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

517 25. **Stoesser N, Batty EM, Eyre DW, Morgan M, Wyllie DH, Del Ojo Elias C, Johnson**
518 **JR, Walker AS, Peto TE, Crook DW.** 2013. Predicting antimicrobial susceptibilities
519 for Escherichia coli and Klebsiella pneumoniae isolates using whole genomic
520 sequence data. J Antimicrob Chemother **68:**2234-2244.
521 26. **Gordon NC, Price JR, Cole K, Everitt R, Morgan M, Finney J, Kearns AM, Pichon**
522 **B, Young B, Wilson DJ, Llewelyn MJ, Paul J, Peto TEA, Crook DW, Walker AS,**
523 **Golubchik T.** 2014. Prediction of Staphylococcus aureus Antimicrobial Resistance
524 by Whole-Genome Sequencing. Journal of Clinical Microbiology **52:**1182-1191.
525 27. **Walker TM, Kohl TA, Omar SV, Hedge J, Elias CDO, Bradley P, Iqbal Z,**
526 **Feverriegel S, Niehaus KE, Wilson DJ, Clifton DA, Kapatai G, Ip CLC, Bowden R,**
527 **Drobniewski FA, Allix-Beguec C, Gaudin C, Parkhill J, Diet R, Supply P, Crook**
528 **DW, Smith EG, Walker AS, Ismail N, Niemann S, Petot TEA, MMM MMM.** 2015.
529 Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug
530 susceptibility and resistance: a retrospective cohort study. Lancet Infectious
531 Diseases **15:**1193-1202.
532 28. **Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE, Jr., Walker H, Hawkins PA, Tran T,**
533 **Whitney CG, McGee L, Beall BW.** 2016. Penicillin-Binding Protein Transpeptidase
534 Signatures for Tracking and Predicting beta-Lactam Resistance Levels in
535 Streptococcus pneumoniae. MBio **7**.
536 29. **Pankhurst LJ, Del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J,**
537 **Fermont JM, Gascoyne-Binzi DM, Kohl TA, Kong C, Lemaitre N, Niemann S, Paul**
538 **J, Rogers TR, Roycroft E, Smith EG, Supply P, Tang P, Wilcox MH, Wordsworth S,**
539 **Wyllie D, Xu L, Crook DW, Group C-TS.** 2016. Rapid, comprehensive, and

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. Lancet Respir Med **4:**49-58.

30. **Hasman H, Saputra D, Sicheritz-Ponten T, Lund O, Svendsen CA, Frimodt-Moller N, Aarestrup FM.** 2014. Rapid Whole-Genome Sequencing for Detection and Characterization of Microorganisms Directly from Clinical Samples. Journal of Clinical Microbiology **52:**139-146.

31. **Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G, Salamat SM, Somasekar S, Federman S, Miller S, Sokolic R, Garabedian E, Candotti F, Buckley RH, Reed KD, Meyer TL, Seroogy CM, Galloway R, Henderson SL, Gern JE, DeRisi JL, Chiu CY.** 2014. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. N Engl J Med **370:**2408-2417.

32. **Naccache SN, Peggs KS, Mattes FM, Phadke R, Garson JA, Grant P, Samayoa E, Federman S, Miller S, Lunn MP, Gant V, Chiu CY.** 2015. Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next-generation sequencing. Clin Infect Dis **60:**919-923.

33. **Wilson MR, Shanbhag NM, Reid MJ, Singhal NS, Gelfand JM, Sample HA, Benkli B, O'Donovan BD, Ali IK, Keating MK, Dunnebacke TH, Wood MD, Bollen A, DeRisi JL.** 2015. Diagnosing Balamuthia mandrillaris Encephalitis With Metagenomic Deep Sequencing. Ann Neurol **78:**722-730.

34. **Votintseva AA, Pankhurst LJ, Anson LW, Morgan MR, Gascoyne-Binzi D, Walker TM, Quan TP, Wyllie DH, Del Ojo Elias C, Wilcox M, Walker AS, Peto TE, Crook DW.** 2015. Mycobacterial DNA extraction for whole-genome sequencing from early positive liquid (MGIT) cultures. J Clin Microbiol **53:**1137-1143.

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

563 35. **Wood DE, Salzberg SL.** 2014. Kraken: ultrafast metagenomic sequence
564 classification using exact alignments. Genome Biol **15:**R46.
565 36. **Ondov BD, Bergman NH, Phillippy AM.** 2011. Interactive metagenomic
566 visualization in a Web browser. BMC Bioinformatics **12:**385.
567 37. **Youden WJ.** 1950. Index for rating diagnostic tests. Cancer **3:**32-35.
568 38. **Kearney MF, Spindler J, Wiegand A, Shao W, Anderson EM, Maldarelli F,**
569 **Ruscetti FW, Mellors JW, Hughes SH, Le Grice SF, Coffin JM.** 2012. Multiple
570 sources of contamination in samples from patients reported to have XMRV infection.
571 PLoS One **7:**e30889.
572 39. **Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, Aronsohn**
573 **A, Hackett J, Jr., Delwart EL, Chiu CY.** 2013. The perils of pathogen discovery:
574 origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction
575 spin columns. J Virol **87:**11966-11977.

576

577

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure legends

FIGURE 1. Study samples and quality control. **Staphylococcus epidermidis*, other coagulase-negative Staphylococci, viridans group Streptococci, *Propionibacterium acnes*

FIGURE 2. Sequencing data filtering calibration heatmaps. Two thresholds (1 and 2), and three parameters (a-c), were used to determine true infection: 1) samples with more reads from a given species than an upper read cut-off (a, x-axis) were included; 2) samples with more species-specific reads than a lower read cut-off (b, panels) and with the percentage of species-specific reads as a proportion of all bacterial reads present above a percentage cut-off (c, y-axis) were also included. Youden-Index = (sensitivity + specificity) - 1.

FIGURE 3. Sonication culture and sequencing comparison. The proportion of sequencing reads classified as bacterial is shown on the y-axis on a log scale, and the number of CFU from sonication fluid culture on the x-axis. Markers are coloured by the concordance of sonication fluid culture and sequencing. A single marker is shown per patient sample, where only one of several species isolated was found by sequencing this is shown as a false negative marker, similarly any sample with one or more false positive species identified by sequencing is shown as false positive. False negative results where a coagulase-negative *Staphylococcus* was cultured from sonication fluid, but not found in tissue samples or on sequencing are shown separately, as are samples only identified to the genus level by sequencing. Results were very similar if absolute numbers of bacterial reads were plotted on the y-axis instead.

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

599 **TABLE 1:** Summary of species observed in microbiology culture for sonication fluid and
600 periprosthetic tissue (PPT), presented by joint/implant type. Results reported as number
601 observed in sonication fluid (number observed in PPT). [a]Metalwork comprises plates
602 and/or screws from tibia (*n*=3), femur (*n*=4), spine (*n*=2), foot (*n*=2), humerus (*n*=1), ankle
603 (*n*=1) and ulna (*n*=1).

Species	**Ankle** ***n*=6**	**Hip** ***n*=32**	**Knee** ***n*=42**	**Metalwork[a]** ***n*=14**	**Shoulder** ***n*=3**	**Total** ***n*=97**
Staphylococci						
Staphylococcus aureus	0 (1)	5 (9)	10 (11)	5 (6)	1 (1)	21 (28)
Staphylococcus condimenti			1 (0)			1 (0)
Staphylococcus epidermidis	1 (1)	6 (10)	9 (12)	0 (1)	0 (1)	16 (25)
Staphylococcus lugdunensis	0 (1)	1 (1)		1 (1)		2 (3)
Coagulase-negative *Staphylococcus*	1 (0)	0 (3)	1 (2)	0 (1)	1 (0)	3 (6)
Streptococci						
Streptococcus agalactiae			2 (2)			2 (2)
Streptococcus dysgalactiae		0 (1)				0 (1)
Streptococcus oralis		0 (2)				0 (2)
Streptococcus pneumoniae		1 (1)				1 (1)
Streptococcus vestibularis				1 (0)		1 (0)
Enterococci						
Enterococcus faecalis		3 (4)	2 (2)	0 (1)		5 (7)
Enterococcus faecium		0 (1)	3 (3)			3 (4)
Enterobacteriacae						
Citrobacter koseri		1 (1)				1 (1)
Citrobacter species			1 (1)			1 (1)
Enterobacter cloacae	1 (1)		1 (1)			2 (2)
Escherichia coli		1 (2)				1 (2)
Klebsiella oxytoca				0 (1)		0 (1)
Klebsiella pneumoniae			1 (1)			1 (1)
Morganella morganii			2(2)			2 (2)
Proteus mirabilis		0 (1)	1 (1)			1 (2)
Serratia marcescens			1 (1)			1 (1)
Corynebacteria						
Corynebacterium amycolatum		0 (1)				0 (1)
Corynebacterium aurimucosum		0 (1)				0 (1)
Corynebacterium propinquum				0 (1)		0 (1)
Corynebacterium striatum		0 (3)		0 (1)		0 (4)
Other						

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Aeromonas species				0 (1)		0 (1)
Aeromonas hydrophila				1 (0)		1 (0)
Arcanobacterium haemolyticum		1 (0)				1 (0)
Bacillus species	0 (1)	1 (2)				1 (3)
Finegoldia magna				1 (0)		1 (0)
Gemella morbillorum		1 (1)	1 (1)			2 (2)
Granulicatella adiacens				0 (1)		0 (1)
Micrococcus luteus				0 (1)		0 (1)
Mycobacterium fortuitum			0 (1)			0 (1)
Propionibacterium acnes		1 (1)	2 (0)			3 (1)
Propionibacterium spp.		0 (1)				0 (1)
Pseudomonas aeruginosa	1 (2)	0 (2)		1 (2)		2 (6)
No growth	3 (3)	12 (7)	7 (6)	5 (4)	1 (1)	28 (21)
Total	**7 (10)**	**34 (55)**	**45 (47)**	**15 (22)**	**3 (3)**	**104 (137)**

604

TABLE 2: Comparison of species identified from sonication fluid and PPT culture with species identified from metagenomics sequencing reads for all samples passing thresholds for analysis in the derivation (*n*=50) and validation (*n*=49) data sets. CoNS, coagulase-negative *Staphylococcus* species; FN, false-negative result; FP, false-positive result. See Supplementary Table 2 for genus details.

Sample	Sonication species	Sonication cfu	Tissue culture species	Tissue samples (positive/total)	Sequencing species	Reads	% Bacterial reads	False Negative	False Positive
Derivation set									
164	*S. epidermidis*	>490	*S. epidermidis*	3/5	*S. epidermidis*	2716	81%		
171	*S. epidermidis*	>490	*S. epidermidis*	2/5	*S. epidermidis*	3154	79%		
182	*E. faecium*	100-240	*E. faecium*	6/6	*E. faecium*	144	43%		
183	*S. epidermidis*	>490	*S. epidermidis*	4/5	*S. epidermidis*	3362	87%		
193	*S. aureus*	>490	*S. aureus*	5/5	*S. aureus*	360718	97%		
	S. condimenti	>490						Not in database, genus only	
198	*S. epidermidis*	>490	*S. epidermidis*	3/5	*S. epidermidis*	228	52%		
208	*E. faecalis*	>490	*E. faecalis*	5/5	*E. faecalis*	14486	31%		
	E. coli	250-490	*E. coli*	4/5	*E. coli*	6503	14%		
213	*S. aureus*	>490	*S. aureus*	5/5	*S. aureus*	167	80%		
219	*S. lugdunensis*	>490	*S. lugdunensis*	3/4	*S. lugdunensis*	411	27%		
			C. propinquum	4/4	*A. xylosoxidans*	722	47%		FP
			S. epidermidis	1/4					
223	*S. aureus*	>490	*S. aureus*	4/5	*S. aureus*	7504	95%		
229	*S. aureus*	>490	*S. aureus*	1/2	*S. aureus*	6038	98%		
249	*P. acnes*	>490	*P. acnes*	4/5	*P. acnes*	108940	100%		

259	*S. epidermidis*	>490	*S. epidermidis*	3/4	*S. epidermidis*	749	86%		
289	*S. aureus*	250-490	*S. aureus*	5/5	*S. aureus*	2105	94%		
296	*S. marcescens*	250-490	*S. marcescens*	4/4	*S. marcescens*	590	60%		
312	*C. koseri*	>490	*C. koseri*	4/5	*C. koseri*	221516	95%		
329	*M. morganii*	>490	*M. morganii*	6/6	*M. morganii*	18553	95%		
335	*M. morganii*	100-240	*M. morganii*	3/5	*M. morganii*	3555	94%		
352	*Bacillus spp.*	100-240	Bacillus spp.	2/5	*Bacillus spp.*	1109		Genus only	
354	*A. haemolyticum*	>490	*S. aureus*	2/6	*A. haemolyticum*	11182	72%		
	E. faecalis	>490	*E. faecalis*	4/6	*E. faecalis*	1173	8%		
			S. oralis	1/6	*F. nucleatum*	1156	7%		Plausible anaerobe
			CoNS	5/6					
			P. aeruginosa	1/6					
			C. striatum	1/6					
			S. epidermidis	2/6					
361	*F. magna*	>490	No growth	0/6	*F. magna*	3674	95%		
362	*P. acnes*	<50	No growth	0/1					
366	*K. pneumoniae*	>490	*K. pneumoniae*	4/5	*K. pneumoniae*	8981	25%		
369	*E. cloacae*	>490	*E. cloacae*	4/5	*E. cloacae*	2502	11%		
	P. aeruginosa	100-240	*P. aeruginosa*	5/5	*P. aeruginosa*	1192	5%		
			S. epidermidis	4/5	*V. parvula*	14801	65%		Plausible anaerobe
			S. lugdunensis	1/5					
370	*P. acnes*	<50	No growth	0/4					
371	*S. epidermidis*	>490	*S. epidermidis*	3/3	*S. epidermidis*	4998	87%		
			CoNS	1/3					
373	*E. faecalis*	>490	*E. faecalis*	1/5	*E. faecalis*	1234	38%		
	S. epidermidis	100-240	*S. epidermidis*	3/5	*S. epidermidis*	616	19%		
376	*E. cloacae*	>490	*E. cloacae*	4/4	*E. cloacae*	122622	95%		

	CoNS	>490						Probable plate contaminant
382	*S. aureus*	<50	*S. aureus*	4/4	*S. aureus*	440	50%	
			S. dysgalactiae	2/4				
384	*S. epidermidis*	>490	*S. epidermidis*	2/4	*S. epidermidis*	1751	85%	
399	*S. aureus*	Not recorded	*S. aureus*	2/5	*S. aureus*	1955	97%	
404	*S. aureus*	>490	*S. aureus*	4/6	*S. aureus*	2257	39%	
			C. striatum	5/6				
			E. coli	2/6				
408	*S. aureus*	>490	*S. aureus*	4/4	*S. aureus*	368	87%	
410	*S. aureus*	100-240	*S. aureus*	4/4	*S. aureus*	235	27%	
			CoNS	1/4	*C. jeikeium*	401	46%	FP
176	No growth		*S. aureus*	1/4				
			P. aeruginosa	3/4				
346	No growth		*S. aureus*	3/5				
			M. fortuitum	1/5				
359	No growth		*S. epidermidis*	1/4	*P. acnes*	464	24%	FP - *P. acnes*
372	No growth		*S. aureus*	4/4	*P. acnes*	3874	51%	FP - *P. acnes*
			G. adiacens	1/4				
375	No growth		*S. epidermidis*	1/5	*P. acnes*	5686	75%	FP - *P. acnes*
379	No growth		*S. aureus*	3/5				
			CoNS	1/5				
389	No growth		*S. epidermidis*	2/5				
			Bacillus spp.	1/5				
341	No growth		No growth	0/3	*S. aureus*	153	42%	Prior flucloxacillin, plausible pathogen
358	No growth		No growth	0/3				

364	No growth		No growth	0/4				
365	No growth		No growth	0/1	*P. acnes*	318	23%	FP - *P. acnes*
368	No growth		No growth	0/4	*R. pickettii*	3146	40%	FP
					E. cloacae	2629	33%	FP
374	No growth		No growth	0/4				
383	No growth		No growth	0/4				
388	No growth		No growth	0/3				
391	No growth		No growth	0/4				
Validation Set								
256	*G. morbillorum*	>490	*G. morbillorum*	6/6	*G. morbillorum*	784	72%	
397	*S. epidermidis*	>490	*S. epidermidis*	5/5	*S. epidermidis*	6717	94%	
400	*A. hydrophila*	>490	Aeromonas spp.	3/4				FN
	S. aureus	100-240	*S. aureus*	4/4	*S. aureus*	6547	5%	
			P. aeruginosa	2/4	*P. aeruginosa*	86920	68%	In tissue
			K. oxytoca	1/4	*K. oxytoca*	1238	1%	In tissue
					F. magna	15606	12%	Plausible anaerobe
			E. faecalis	1/4	*E. faecalis*	1303	1%	In tissue
405	*S. lugdunensis*	>490	*S. lugdunensis*	6/6	*S. lugdunensis*	311	96%	
406	*E. faecium*	250-490	*E. faecium*	2/3				FN
409	*S. agalactiae*	>490	*S. agalactiae*	5/5	*S. agalactiae*	2556	93%	
423	*S. aureus*	>490	*S. aureus*	4/4	*S. aureus*	15479	98%	
426	*S. aureus*	250-490	*S. aureus*	2/4	*S. aureus*	11981	89%	
430	*S. pneumoniae*	>490	*S. pneumoniae*	5/5	*S. pneumoniae*	5697	82%	
442	*E. faecium*	>490	*E. faecium*	5/5	*E. faecium*	1689	68%	
450	*S. aureus*	>490	*S. aureus*	5/6	*S. aureus*	2584	98%	
			Propionibacterium spp.	1/6				
459	*S. agalactiae*	>490	*S. agalactiae*	5/5	*S. agalactiae*	114212	93%	

465	*S. aureus*	>490	*S. aureus*	4/4	*S. aureus*	1171	97%		
468	*S. aureus*	>490	*S. aureus*	3/3	*S. aureus*	676	93%		
473	*E. faecalis*	250-490	*E. faecalis*	4/4	*E. faecalis*	228	73%		
474	*S. epidermidis*	250-490	*C. striatum*	2/5				Genus only	
			S. aureus	1/5					
			S. epidermidis	3/5					
480	*S. epidermidis*	250-490	*S. epidermidis*	5/5	*S. epidermidis*	557	80%		
482	*S. epidermidis*	>490	*S. epidermidis*	5/5	*S. epidermidis*	1327	88%		
483	*S. aureus*	100-240	No growth	0/5	*S. aureus*	444	85%		
485	*G. morbillorum*	>490	*G. morbillorum*	3/4	*G. morbillorum*	123300	18%		
			S. oralis	1/4	*P. micra*	508822	76%		Plausible anaerobe
			S. aureus	1/4	*S. equi*	16580	2%		FP
			C. amycolatum	1/4	*S. anginosus*	8019	1%		Plausible anaerobe
			P. mirabilis	1/4					
486	*E. faecalis*	>490	*E. faecalis*	5/5	*E. faecalis*	3904	43%		
			S. epidermidis	4/5					
487	*S. aureus*	<50	*S. aureus*	2/4	*S. aureus*	121284	98%		
489	*S. aureus*	100-240	*S. aureus*	2/4	*S. aureus*	858	95%		
			S. epidermidis	1/4					
498	*S. aureus*	<50	*S. aureus*	4/5	*S. aureus*	135	88%		
504	*S. aureus*	>490	*S. aureus*	7/7	*S. aureus*	3229	97%		
507	*P. mirabilis*	<50	*P. mirabilis*	2/5	*P. mirabilis*	184	15%		
					M. morganii	981	83%		FP
511	*P. aeruginosa*	Not recorded	*P. aeruginosa*	3/6				FN	
					P. acnes	1377	69%		FP - *P. acnes*
513	*Citrobacter spp.*	<50	Citrobacter spp.	2/5	*C. koseri*	1133	87%		
514	*S. epidermidis*	>490	*S. epidermidis*	5/5	*S. epidermidis*	11803	91%		

516	CoNS	100-240	No growth	0/4				Probable plate contaminant	
414	*S. epidermidis*	<50	*S. epidermidis*	5/5	*S. epidermidis*	1194	91%		Low sonication count + in tissue
490	*S. epidermidis*	<50	No growth	0/5					
497	*S. vestibularis*	<50	*C. striatum*	4/4					
503	CoNS	<50	No growth	0/5					
512	*S. epidermidis*	<50	*S. epidermidis*	5/5					
475	No growth		*M. luteus*	1/4	*S. dysgalactiae*	156	37%		Prior flucloxacillin, plausible pathogen
476	No growth		*P. aeruginosa*	4/4					
			S. aureus	3/4					
478	No growth		CoNS	1/4					
496	No growth		Bacillus spp.	1/4					
502	No growth		*C. aurimucosum*	2/4	*C. aurimucosum*	2379	42%		In tissue
			S. epidermidis	4/4	*S. epidermidis*	1336	24%		In tissue
			E. faecium	3/4					
			CoNS	2/4					
510	No growth		*S. epidermidis*	2/4	*S. epidermidis*	290	26%		
					P. acnes	232	21%		FP - *P. acnes*
515	No growth		*E. faecalis*	1/7	*P. acnes*	873	34%		FP - *P. acnes*
472	No growth		No growth	0/4					
505	No growth		No growth	0/5					
506	No growth		No growth	0/6					
508	No growth		No growth	0/8					
509	No growth		No growth	0/5					

609

610

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

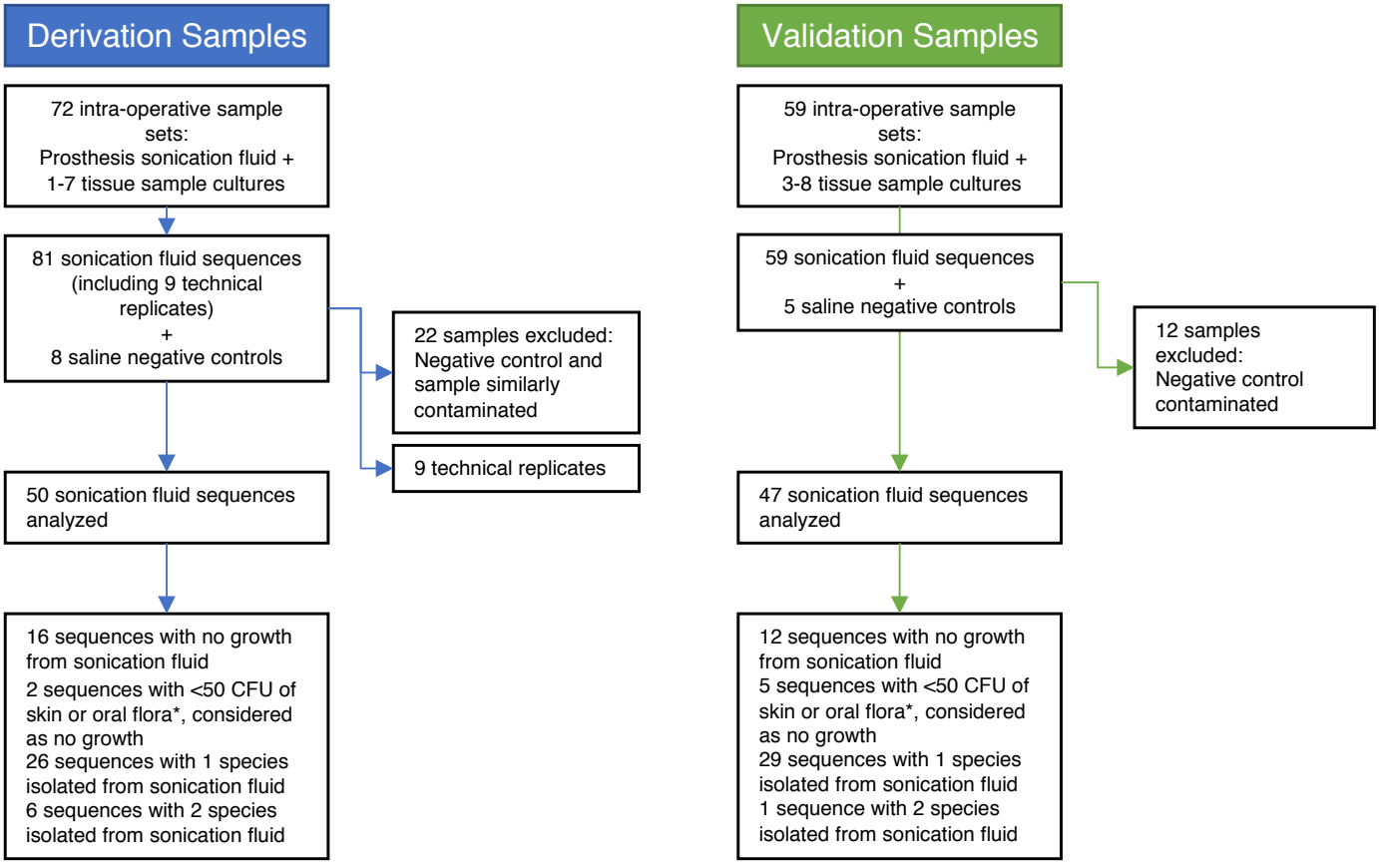

FIGURE 1. Study samples and quality control. **Staphylococcus epidermidis*, other coagulase-negative Staphylococci, viridans group Streptococci, *Propionibacterium acnes.*

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

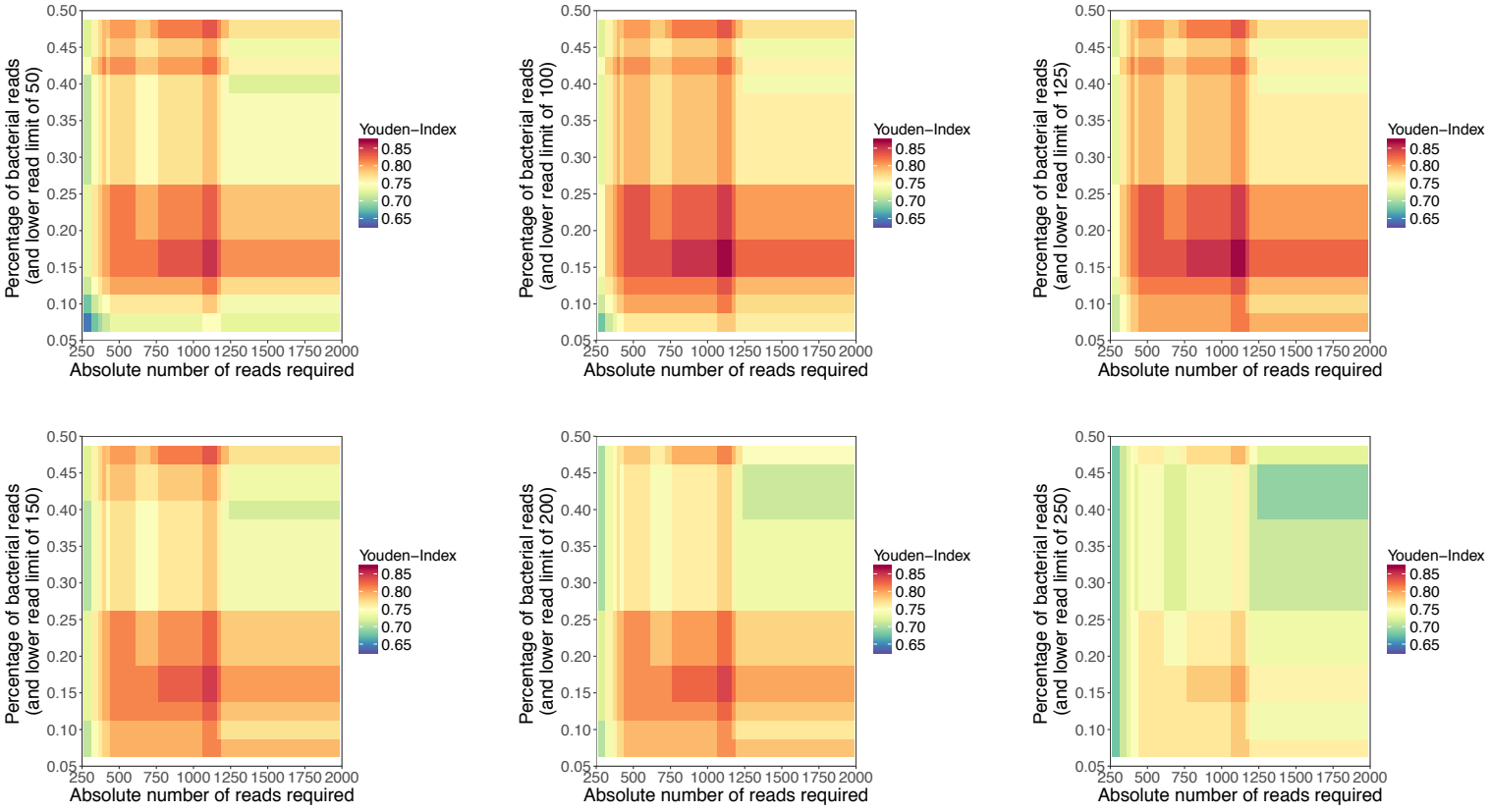

FIGURE 2. Sequencing data filtering calibration heatmaps. Two thresholds (1 and 2), and three parameters (a-c), were used to determine true infection: 1) samples with more reads from a given species than an upper read cut-off (a, x-axis) were included; 2) samples with more species-specific reads than a lower read cut-off (b, panels) and with the percentage of species-specific reads as a proportion of all bacterial reads present above a percentage cut-off (c, y-axis) were also included. Youden-Index = (sensitivity + specificity) -1.

bioRxiv preprint doi: https://doi.org/10.1101/118026; this version posted March 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

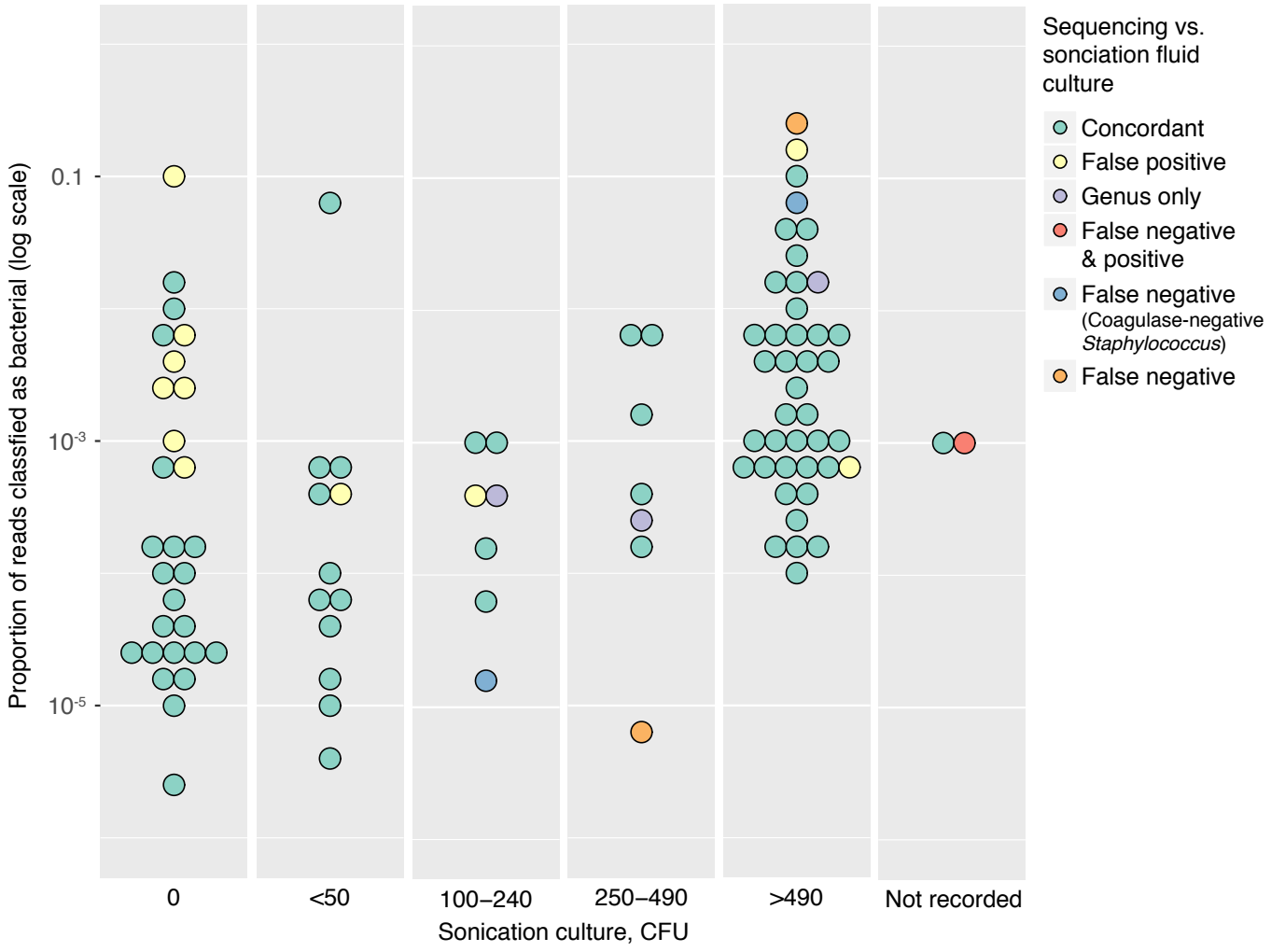

FIGURE 3. Sonication culture and sequencing comparison. The proportion of sequencing reads classified as bacterial is shown on the y-axis on a log scale, and the number of CFU from sonication fluid culture on the x-axis. Markers are coloured by the concordance of sonication fluid culture and sequencing. A single marker is shown per patient sample, where only one of several species isolated was found by sequencing this is shown as a false negative marker, similarly any sample with one or more false positive species identified by sequencing is shown as false positive. False negative results where a coagulase-negative *Staphylococcus* was cultured from sonication fluid, but not found in tissue samples or on sequencing are shown separately, as are samples only identified to the genus level by sequencing. Results were very similar if absolute numbers of bacterial reads were plotted on the y-axis instead.