





Impact of Contaminating DNA in Whole-Genome Amplification Kits Used for Metagenomic Shotgun Sequencing for Infection Diagnosis

Matthew Thoendel,^a  Patricio Jeraldo,^b Kerryl E. Greenwood-Quaintance,^c Janet Yao,^b Nicholas Chia,^b Arlen D. Hanssen,^d Matthew P. Abdel,^d  Robin Patel^{a,c}

Division of Infectious Diseases, Department of Medicine, Mayo Clinic, Rochester, Minnesota, USA^a; Department of Surgery, Mayo Clinic, Rochester, Minnesota, USA^b; Division of Clinical Microbiology, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA^c; Department of Orthopedic Surgery, Mayo Clinic, Rochester, Minnesota, USA^d

ABSTRACT Whole-genome amplification (WGA) is a useful tool for amplification of very small quantities of DNA for many uses, including metagenomic shotgun sequencing for infection diagnosis. Depending on the application, background DNA from WGA kits can be problematic. Three WGA kits were tested for their utility in a metagenomics approach to identify the pathogens in sonicate fluid comprised of biofilms and other materials dislodged from the surfaces of explanted prosthetic joints using sonication. The Illustra V2 Genomiphi, Illustra single cell Genomiphi, and Qiagen REPLI-g single cell kits were used to test identical sonicate fluid samples. Variations in the number of background reads, the genera identified in the background, and the number of reads from known pathogens known to be present in the samples were observed between kits. These results were then compared to those obtained with a library preparation without prior WGA using an NEBNext Ultra II paired-end kit, which requires a very small amount of input DNA. This approach also resulted in the presence of contaminant bacterial DNA and yielded fewer reads from the known pathogens. These findings highlight the impact that WGA kit selection can have on metagenomic analysis of low-biomass samples and the importance of the careful selection and consideration of the implications of using these tools.

KEYWORDS metagenomics, prosthetic joint infection, whole-genome amplification

Metagenomic shotgun sequencing is a rapidly developing powerful tool to examine microbial communities. It is also gaining interest in the field of clinical microbiology as a way to identify pathogens in normally sterile specimens (1–7). Not only does this approach allow for pathogen identification, but the gene content information can also be used for other analyses, such as antibiotic resistance prediction (8, 9), typing for tracking of outbreaks or epidemiological studies (10), or assessment for other relevant genes, such as those encoding virulence factors (11, 12).

One significant barrier to metagenomic shotgun sequencing approaches is the amount of genetic material necessary to perform these experiments. This is not an obstacle in settings such as fecal microbiome studies, where abundant organisms are present, but in other scenarios (e.g., in scenarios with low-biomass environmental samples and uncultured bacteria or in single cell analysis), the number of cells and the amount of genetic material can be limited (13–17). It is in these situations that whole-genome amplification (WGA) has the potential to play an important role in pathogen detection.

Received 8 December 2016 Returned for modification 10 January 2017 Accepted 17 March 2017

Accepted manuscript posted online 29 March 2017

Citation Thoendel M, Jeraldo P, Greenwood-Quaintance KE, Yao J, Chia N, Hanssen AD, Abdel MP, Patel R. 2017. Impact of contaminating DNA in whole-genome amplification kits used for metagenomic shotgun sequencing for infection diagnosis. *J Clin Microbiol* 55:1789–1801. <https://doi.org/10.1128/JCM.02402-16>.

Editor Nathan A. Ledebauer, Medical College of Wisconsin

Copyright © 2017 American Society for Microbiology. All Rights Reserved.

Address correspondence to Robin Patel, patel.robin@mayo.edu.

Whole-genome amplification refers to a collection of methods to amplify the entire DNA content of an organism, ideally in an unbiased manner. One common form of this approach is multiple displacement amplification (MDA), in which a high-fidelity phi29 polymerase amplifies DNA in an isothermic reaction with the aid of random hexamer primers (18). When the polymerase encounters a previously synthesized strand, it displaces the strand to continue polymerization while creating a branching single-stranded template that new primers can bind to and initiate polymerization. The results are large branching DNA molecules that can then be fragmented and used for next-generation sequencing.

The approach is not without its limitations, as there is observed bias in the amplification of templates (19–22). Biases against high-GC-content DNA (20–22), low-abundance templates (23), and smaller DNA fragments (21) have all been reported. Additionally, if little or no template DNA is involved, amplification products are still produced. This raises the question as to whether these amplification products are the result of contaminating DNA in WGA reagents or an artifact of the extension of random hexamers. The presence of contaminating DNA in some WGA reagents has been suggested by other investigators (24). This contamination can confound metagenomic shotgun sequencing analysis of samples from sterile sites, where the presence of any microorganism could be considered significant.

We sought to use WGA as a method to amplify DNA for metagenomic analysis of materials dislodged from the surfaces of explanted prosthetic joint components using sonication (i.e., sonicate fluid). These prostheses were removed due to mechanical failure or prosthetic joint infection (PJI). PJIs occur in up to 2% of joint arthroplasties, with devastating consequences for individuals, as they often require multiple surgeries and prolonged intravenous antibiotic treatment to achieve a cure (25). Treatment varies depending on the pathogen identified; however, in approximately 20% of cases, no pathogen is detected using conventional techniques (25). Metagenomic shotgun sequencing is a promising tool to aid in these cases. The overwhelming ratio of human to pathogen DNA in PJI samples makes it necessary to enrich for microbial DNA, for example, using MolYsis kit reagents (Molzym, Bremen, Germany) (26). The depletion of host DNA, however, results in insufficient DNA quantities for next-generation sequencing in most cases. WGA is one tool available to amplify DNA to sufficient amounts for next-generation sequencing.

During previous studies, we encountered the problem of contaminating DNA complicating the analysis of sequencing results (26), which is emerging as a common problem encountered when attempting to do metagenomic analysis of clinical samples (3, 27–29). Here, we evaluated three commercially available WGA kits to characterize the relative amount and composition of contaminating DNA associated with the use of each of them to inform selection of the appropriate WGA kit for further metagenomics studies of PJI.

RESULTS

To allow an even comparison, prior to WGA all samples were prepared in an identical manner, with the same DNA extraction undergoing WGA with the different kits prior to next-generation sequencing library preparation and metagenomic shotgun sequencing. Two methods were used to determine the bacterial DNA content. The Livermore metagenomics analysis tool kit (LMAT) was used for taxonomic assignment of individual reads, in which a k-mer-based approach was utilized to determine the lowest common ancestor for a sequence based on a database containing a wide range of genomes, including the genomes of human, bacterial, viral, and fungal species (30). The MetaPhlan2 tool was also used for identification of pathogens in samples (31). This tool searches for marker genes that are unique to different genera, species, subspecies, etc., and if sufficient markers are present, it identifies the organism as being present and predicts the relative genome frequency if multiple species are detected.

Both analysis tools identified patterns of species unique to the individual WGA kits. The Illustra single cell Genomiphi kit consistently had the fewest total number of bacterial background reads (Table 1). This was evident in samples from infected joints

TABLE 1 Sonicate fluid samples tested and relative read counts^a

Sample source, sample no.	Site	Microbiology Organism in sonicate fluid (concn)	No. of positive tissue cultures	No. of reads								
				Qiagen REPL-g single cell kit			Illustra single cell kit			Illustra V2 kit		
				Total	Pathogen reads	Other microbial reads	Total	Pathogen reads	Other microbial reads	Total	Pathogen reads	Other microbial reads
Culture-positive PJI												
980	H	Group C <i>Streptococcus</i> (<20)	1 of 3	27,244,012	2,694,906	103	36,576,986	286,577	556	25,754,214	3,075,181	566
982	K	<i>Staphylococcus</i> <i>epidemicus</i> (>100)	3 of 4	28,320,234	3,272,895	591	30,284,116	102,934	203	26,508,366	9,539,147	17,625
986	K	<i>Staphylococcus aureus</i> (<20)	1 of 5	26,835,306	26,397	77,581	28,991,289	424	308	27,391,148	168,046	240,703
996	H	<i>Bacteroides fragilis</i> (>100)	3 of 3	28,388,841	155,761	4,853	28,745,844	11,134	1,346	28,904,638	96,154	8,140
1002	K	<i>Corynebacterium striatum</i> (51–100)	3 of 3	32,165,206	536,155	2,594	27,925,551	5,503	422	32,338,186	1,000,810	53,019
Culture-negative PJI, 984												
	K	Culture negative	0 of 4	31,616,449	NA	2,821	26,240,307	NA	305	36,808,230	NA	16,974
Aseptic failure												
983	K	Culture negative	0 of 3	32,924,419	NA	278	29,076,492	NA	522	32,128,178	NA	45,539
987	K	Anaerobic organism (<20)	1 of 3, <i>Bacillus</i> species	30,316,155	NA	7,628	30,587,175	NA	265	27,955,363	NA	655,417
Controls												
<i>C. glutamicum</i>	NA	NA	NA	28,121,822	26,912,706	2,145	28,973,351	22,276,446	856	26,986,400	26,149,302	1,311
Ringer's solution	NA	NA	NA	27,488,254	NA	24,140	29,531,289	NA	3,560	27,810,322	NA	835,014
WGA with no template	NA	NA	NA	25,669,747	NA	6,695	53,626,034	NA	46,759	24,512,662	NA	7,818,383

^aSamples are grouped by classification (PJI, prosthetic joint infection). Site indicates the anatomical surgical site (H, hip; K, knee). Microbiology findings include culture results for sonicate fluid as well as accompanying results for intraoperative tissue specimens collected. Sonicate fluid indicates the identified organism as well as the concentration (in number of CFU/10 ml). Total reads were the total number of sequencing reads prior to preferential of human reads. Pathogen reads were determined by identifying the number of reads from the identified pathogen's genus, as categorized by LMAT. Other microbial reads were calculated by adding all reads assigned to a genus other than the known pathogen and includes bacterial, viral, fungal, and protozoal reads. NA, not applicable.

TABLE 2 Taxonomic identification of reads^a

Sample source, sample no.	LMAT identification(s) (no. of reads)	Qiagen REPLI-g single cell kit	Illustra single cell kit	Illustra V2 kit
Culture-positive PJI 980	Streptococcus (2,694,906)		Streptococcus (286,577) , <i>Toxoplasma</i> (489)	Streptococcus (3,075,181) , <i>Pseudomonas</i> (232), <i>Anaerococcus</i> (135)
982	Staphylococcus (3,272,895) , <i>Propionibacterium</i> (195), <i>Mupapillomavirus</i> (137), <i>Finegoldia</i> (115)		Staphylococcus (102,934) , <i>Toxoplasma</i> (131)	Staphylococcus (9,539,147) , <i>Pseudomonas</i> (15,292), <i>Acinetobacter</i> (739), <i>Pseudoperonospora</i> (346), <i>Streptococcus</i> (303), <i>Elizabethkingia</i> (242), <i>Malassezia</i> (176), <i>Propionibacterium</i> (142)
986	<i>Streptococcus</i> (37,268), Staphylococcus (26,397) , <i>Prevotella</i> (13,472), <i>Haemophilus</i> (11,513), <i>Campylobacter</i> (3,344), <i>Propionibacterium</i> (2,298), <i>Gemella</i> (2,045), <i>Alloprevotella</i> (1,756), <i>Chryseobacterium</i> (1,531), <i>Rothia</i> (1,458), <i>Neisseria</i> (913), <i>Capnocytophaga</i> (911), <i>Acinetobacter</i> (297), <i>Corynebacterium</i> (248), <i>Alloiococcus</i> (156)		Staphylococcus (424) , <i>Toxoplasma</i> (118)	Staphylococcus (168,046) , <i>Haemophilus</i> (69,045), <i>Streptococcus</i> (66,370), <i>Pseudomonas</i> (58,244), <i>Corynebacterium</i> (18,344), <i>Elizabethkingia</i> (12,471), <i>Propionibacterium</i> (7,678), <i>Gemella</i> (2,298), <i>Neisseria</i> (2,092), <i>Rothia</i> (1,186), <i>Empedobacter</i> (835), <i>Acinetobacter</i> (479), <i>Fusobacterium</i> (434), <i>Lactobacillus</i> (237), <i>Prevotella</i> (185), <i>Granulicatella</i> (144), <i>Delftia</i> (130), <i>Achromobacter</i> (126)
996	Bacteroides (155,761) , <i>Clostridium</i> (4,115), <i>Parabacteroides</i> (191), <i>Corynebacterium</i> (168), <i>Blautia</i> (143)		Bacteroides (11,134) , <i>Toxoplasma</i> (669), <i>Clostridium</i> (454)	Bacteroides (96,154) , <i>Clostridium</i> (4,052), <i>Pseudomonas</i> (1,483), <i>Yarrowia</i> (563), <i>Corynebacterium</i> (364), <i>Propionibacterium</i> (331), <i>Staphylococcus</i> (345), <i>Parabacteroides</i> (239), <i>Streptococcus</i> (116), <i>Anaerotruncus</i> (104)
1002	Corynebacterium (536,155) , <i>Lactococcus</i> (2,195), <i>Acinetobacter</i> (187), <i>Propionibacterium</i> (108)		Corynebacterium (5,503) , <i>Toxoplasma</i> (346)	Corynebacterium (1,000,810) , <i>Pseudomonas</i> (32,865), <i>Staphylococcus</i> (5,704), <i>Propionibacterium</i> (5,396), <i>Lactococcus</i> (3,008), <i>Streptococcus</i> (2,121), <i>Acinetobacter</i> (1,293), <i>Dermabacter</i> (634), <i>Malassezia</i> (424), <i>Elizabethkingia</i> (382), <i>Achromobacter</i> (224), <i>Actinobaculum</i> (123), <i>Delftia</i> (108), <i>Debaryomyces</i> (102), <i>Arcanobacterium</i> (101)
Culture-negative PJI, 984	<i>Streptococcus</i> (2,425), <i>Acinetobacter</i> (167)		<i>Toxoplasma</i> (230)	<i>Pseudomonas</i> (12,472), <i>Acinetobacter</i> (1,395), <i>Elizabethkingia</i> (936), <i>Alloiococcus</i> (706), <i>Streptococcus</i> (395), <i>Staphylococcus</i> (264), <i>Achromobacter</i> (205), <i>Propionibacterium</i> (169), <i>Delftia</i> (134), <i>Malassezia</i> (111)
Aseptic Failure 983	<i>Streptococcus</i> (352), <i>Gloeocapsa</i> (137)		<i>Toxoplasma</i> (442)	<i>Pseudomonas</i> (42,315), <i>Acinetobacter</i> (788), <i>Elizabethkingia</i> (742), <i>Staphylococcus</i> (630), <i>Streptococcus</i> (475), <i>Propionibacterium</i> (289)

(Continued on next page)

TABLE 2 (Continued)

Sample source, sample no.	LMAT identification(s) (no. of reads)	QiaGen REPLI-g single cell kit	Illustra single cell kit	Illustra V2 kit
987	<i>Acinetobacter</i> (5,396), <i>Streptococcus</i> (1,179), <i>Prevotella</i> (433), <i>Propionibacterium</i> (346), <i>Kurthia</i> (122)		<i>Toxoplasma</i> (173)	<i>Pseudomonas</i> (642,812), <i>Elizabethkingia</i> (5,283), <i>Propionibacterium</i> (3,271), <i>Peptoniphilus</i> (1,742), <i>Acinetobacter</i> (560), <i>Epilithonimonas</i> (372), <i>Pseudoperonospora</i> (353), <i>Anaerococcus</i> (255), <i>Streptococcus</i> (188), <i>Capnocytophaga</i> (172), <i>Malassezia</i> (158)
Controls				
<i>C. glutamicum</i>	<i>Corynebacterium</i> (26,912,706) , <i>Rothia</i> (1,070), <i>Streptococcus</i> (545), <i>Neisseria</i> (221), <i>Haemophilus</i> (240)		<i>Corynebacterium</i> (22,276,446) , <i>Streptococcus</i> (333), <i>Toxoplasma</i> (126)	<i>Corynebacterium</i> (26,149,302) , <i>Pseudomonas</i> (627), <i>Streptococcus</i> (289)
Ringer's solution	<i>Neisseria</i> (10,575), <i>Streptococcus</i> (8,441), <i>Dolosigranulum</i> (3,912), <i>Haemophilus</i> (716), <i>Propionibacterium</i> (385)		<i>Streptococcus</i> (1,192), <i>Toxoplasma</i> (969), <i>Staphylococcus</i> (929), <i>Propionibacterium</i> (127), <i>Corynebacterium</i> (123)	<i>Pseudomonas</i> (301,181), <i>Staphylococcus</i> (286,646), <i>Streptococcus</i> (135,065), <i>Yarrowia</i> (40,273), <i>Propionibacterium</i> (27,201), <i>Achromobacter</i> (15,251), <i>Elizabethkingia</i> (8,617), <i>Acinetobacter</i> (4,813), <i>Haemophilus</i> (4,423), <i>Alloicoccus</i> (2,610), <i>Granulicatella</i> (2,381), <i>Malassezia</i> (2,557), <i>Delftia</i> (681), <i>Fusarium</i> (671), <i>Corynebacterium</i> (593), <i>Veillonella</i> (498), <i>Alloprevotella</i> (384), <i>Chryseobacterium</i> (216), <i>Dolosigranulum</i> (234), <i>Neisseria</i> (177), <i>Aspergillus</i> (165)
WGA with no template	<i>Achromobacter</i> (5,072), <i>Rothia</i> (1,031), <i>Micrococcus</i> (451), <i>Delftia</i> (110)		<i>Sphingomonas</i> (35,596), <i>Staphylococcus</i> (6,414), <i>Propionibacterium</i> (3,596), <i>Peptoniphilus</i> (697), <i>Finnegoldia</i> (196), <i>Elizabethkingia</i> (149)	<i>Pseudomonas</i> (5,297,561), <i>Propionibacterium</i> (1,557,410), <i>Staphylococcus</i> (407,691), <i>Lactococcus</i> (297,420), <i>Elizabethkingia</i> (134,333), <i>Malassezia</i> (89,791), <i>Delftia</i> (13,668), <i>Acinetobacter</i> (7,858), <i>Phyllobacterium</i> (2,808), <i>Comamonas</i> (3,175), <i>Chryseobacterium</i> (1,732), <i>Achromobacter</i> (1,734), <i>Corynebacterium</i> (1,061), <i>Rhodococcus</i> (1,231), <i>Micrococcus</i> (452), <i>Polynucleobacter</i> (426), <i>Streptococcus</i> (274), <i>Wolbachia</i> (223), <i>Alternaria</i> (220)

†Taxonomic identification of reads was performed using LMAT, and the reads were grouped by genus. All genera with 100 or more assigned reads are listed. Known pathogens identified by culture are in boldface.

as well as uninfected aseptic failures and negative controls. However, the number of reads from pathogens identified by culture was also notably less (Table 1). The Qiagen REPLI-g single cell kit and Illustra V2 Genomiphi kit resulted in an average of 42 and 141 times more reads, respectively, when the number of pathogens was compared to the number identified by the Illustra single cell kit. This was in large part due to the increased number of reads identified as human in origin in the Illustra single cell kit. For example, in the *Corynebacterium glutamicum* positive-control sample to which no human cells were added, if human reads were not prefiltered prior to LMAT analysis, there were 3,479,254 reads identified as human in the Illustra single cell sample, while the Qiagen and Illustra V2 kits resulted in only 117,798 and 68,198 reads, respectively (data not shown). This resulted in fewer reads identified as *C. glutamicum* (Table 1).

The taxonomic assignments of the reads are listed in Table 2. The reads, including those assigned at lower taxonomy levels, such as species and strains, were grouped at the genus level to simplify analysis. Coverage statistics to evaluate the variability in coverage along the length of a reference genome were also calculated (see Table S1 in the supplemental material). The most common organism identified in the samples analyzed with the Illustra single cell kit was *Toxoplasma gondii*, although this likely represents a misclassification of human reads as toxoplasmal due to database errors. When the full LMAT database was used for classification of these samples, the reads were identified as human (data not shown). Similar findings with this database have been previously reported (32).

The Illustra V2 kit had the most reads attributed to contaminant bacterial DNA (Table 1). Reads were frequently identified by LMAT as being from *Pseudomonas*, *Propionibacterium*, *Acinetobacter*, *Staphylococcus*, *Elizabethkingia*, *Streptococcus*, and *Achromobacter* species (Table 2). A similar pattern was observed when the MetaPhlan2 tool was used to identify the presence of bacteria, where *Pseudomonas* and *Propionibacterium* species were frequently identified, particularly in samples without a predominant known pathogen (Table S2).

The samples analyzed with the Qiagen REPLI-g single cell kit had background read counts between those of the other kits (Table 1). This kit also consistently resulted in moderate amounts of reads from known pathogens relative to the Illustra kits (Table 1). *Streptococcus*, *Propionibacterium*, and *Acinetobacter* species were the most common contaminants observed (Table 2).

To further evaluate the consistency of background DNA in kits, the beta diversity between samples was calculated and plotted. The HUMAnN2 pipeline was used to analyze the gene content of samples after human reads had been computationally subtracted. The QIIME pipeline was then utilized to evaluate the diversity present between samples. Samples were found to cluster together (indicating a similar composition) on the basis of the WGA kit used to amplify the DNA (Fig. 1A) rather than the origin of the samples (Fig. 1B). The exceptions to this pattern were samples containing the positive control *C. glutamicum*, in which the presence of the positive-control bacterium dominated the sample, minimizing the presence of background bacterial reads.

The NEBNext Ultra II kit was also utilized to sequence the samples without WGA to determine whether this approach would limit background bacterial DNA reads, as low-input kits are increasingly being used. In most instances, this approach resulted in fewer known pathogen reads (for all samples except sample 996) but more reads from contaminant species, particularly *Propionibacterium* and *Achromobacter* species (Table 3).

In every sample in which a bacterium was identified using conventional microbiology techniques, the organism was detected by metagenomic analysis using each of the three different WGA kits (Tables 1 and 2). However, when sample 986, a sample from a PJI with a low burden of *Staphylococcus aureus*, was amplified with the Illustra single cell kit, it did not produce sufficient reads to be identified using the MetaPhlan2 tool. The consistent presence of *Clostridium* reads from all WGA preparations of sample 996, as well as the absence of these reads as the background in other samples, suggests the

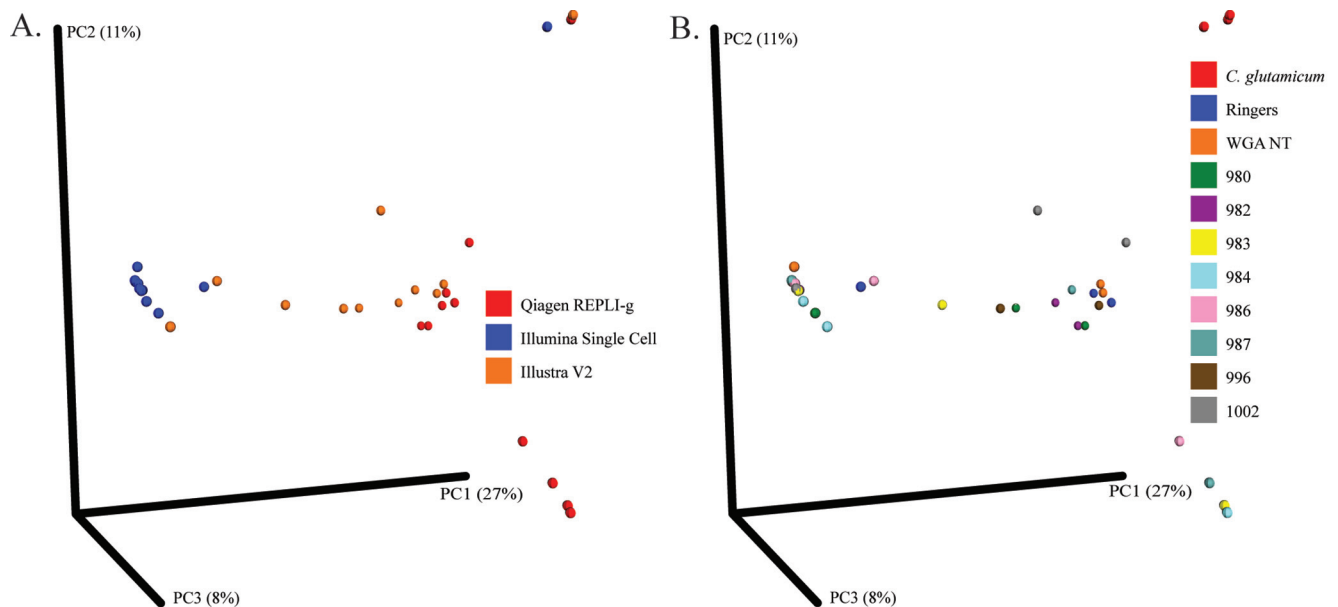


FIG 1 Beta diversity of samples by sample source versus WGA kit. The beta diversity of the gene content of the samples was determined using QIIME and plotted on a principal coordinate analysis plot. Samples are colored according to the WGA kit used for amplification (A) or by template sample source (B). NT, no template.

possible presence of a *Clostridium* species, in addition to the *Bacteroides fragilis* isolate identified by culture and sequencing, though there were insufficient reads for identification by use of the MetaPhlan2 tool. The patient from whom this specimen derived received vancomycin and ertapenem for treatment, which would have also treated the *Clostridium* species, if present. No potential pathogens were detected in the culture-negative PJI or aseptic failure cases with a signal above that for the common background contaminants strong enough to identify them as being present in the samples.

DISCUSSION

Metagenomic whole-genome sequencing is a powerful tool whose utility in diagnosing infections is only beginning to be utilized. For clinical samples with very little DNA available, WGA is a useful tool to amplify genomic material that is present in amounts too small for next-generation sequencing. Our findings highlight the variability in some of the available WGA kits as well as the dramatic impact that this can have on results.

A strength of this study was the type of specimens used to compare the WGA kits. We used sonicate fluid from arthroplasty components resected due to infectious complications or other noninfectious reasons. These samples were well characterized from a microbiological standpoint, with culture results being readily available. They contained different bacteria with a range of infectious burdens, reflective of the variability encountered in clinical samples, as well as uninfected samples, as determined by clinical and microbiological data. This allows for appreciation of the impact of background DNA when identifying a pathogen that is known to be present, as well as what to expect when no bacteria are presumably present.

Perhaps the most striking finding was the wide variation in that amount of contaminant DNA between kits. The Illustra V2 kit generated the largest amount of background reads in regard to both the total number of reads and the number of bacterial species present. The Illustra single cell kit detected the fewest background bacterial species; however, the higher levels of human reads generated limited its utility, as it consistently resulted in fewer reads corresponding to the known pathogens being detected. The larger number of human reads in the *C. glutamicum*-containing control sample suggested increased contaminant human DNA in the Illustra single cell reagents, but this was not directly confirmed. The Qiagen REPLI-g single cell kit

TABLE 3 Read counts and taxonomic identification of reads of non-WGA samples^a

Sample source, sample no.	No. of reads			LMAT identification(s) (no. of reads)
	Total	Pathogen reads	Other microbial reads	
Culture-positive PJI				
980	53,582,421	141,373	414	Streptococcus (141,373)
982	53,690,669	24,419	713	Staphylococcus (24,419) , <i>Propionibacterium</i> (145), <i>Achromobacter</i> (107)
986	55,232,392	718	2,539	<i>Propionibacterium</i> (916), Staphylococcus (718) , <i>Achromobacter</i> (257), <i>Streptococcus</i> (232), <i>Pseudomonas</i> (142)
996	47,452,448	232,328	5,376	Bacteroides (232,328) , <i>Clostridium</i> (3,421), <i>Parabacteroides</i> (1,108), <i>Propionibacterium</i> (136)
1002	43,173,874	60,545	990	Corynebacterium (60,545) , <i>Propionibacterium</i> (265), <i>Achromobacter</i> (127)
Culture-negative PJI, 984	40,410,935	NA	6,575	<i>Propionibacterium</i> (3,097), <i>Achromobacter</i> (884), <i>Micrococcus</i> (507), <i>Malassezia</i> (432), <i>Delftia</i> (258), <i>Pseudomonas</i> (207), <i>Streptococcus</i> (125), <i>Acinetobacter</i> (115), <i>Methylobacterium</i> (101)
Aseptic failure				
983	47,136,446	NA	8,527	<i>Propionibacterium</i> (2,641), <i>Streptococcus</i> (1,548), <i>Achromobacter</i> (666), <i>Rothia</i> (238), <i>Micrococcus</i> (230), <i>Pseudomonas</i> (209), <i>Gemella</i> (209), <i>Corynebacterium</i> (169), <i>Acinetobacter</i> (178), <i>Prevotella</i> (134), <i>Neisseria</i> (119), <i>Delftia</i> (139), <i>Haemophilus</i> (136), <i>Fusobacterium</i> (133), <i>Actinomyces</i> (123), <i>Methylobacterium</i> (118), <i>Staphylococcus</i> (111), <i>Malassezia</i> (109)
987	31,724,272	NA	39,114	<i>Propionibacterium</i> (17,250), <i>Achromobacter</i> (5,555), <i>Pseudomonas</i> (4,282), <i>Corynebacterium</i> (1,720), <i>Delftia</i> (1207), <i>Streptococcus</i> (866), <i>Staphylococcus</i> (910), <i>Micrococcus</i> (778), <i>Methylobacterium</i> (693), <i>Acinetobacter</i> (699), <i>Malassezia</i> (422), <i>Bifidobacterium</i> (355), <i>Stenotrophomonas</i> (445), <i>Rothia</i> (240), <i>Lactobacillus</i> (247), <i>Agrobacterium</i> (262), <i>Dolosigranulum</i> (176), <i>Comamonas</i> (198), <i>Prevotella</i> (149), <i>Alloicoccus</i> (135), <i>Sphingomonas</i> (156), <i>Burkholderia</i> (169), <i>Neisseria</i> (142), <i>Kocuria</i> (146), <i>Aerococcus</i> (129), <i>Bacillus</i> (145), <i>Actinomyces</i> (140), <i>Cupriavidus</i> (105), <i>Geobacillus</i> (101)

^aSequencing libraries were prepared with the NEBNext PE Ultra II kit without prior WGA. Total read counts and reads assigned by LMAT to the known pathogen's genus as well as other nonpathogen genera are listed. All genera with >100 reads are listed with the corresponding read counts under LMAT Identifications. Known pathogens identified by culture are in boldface. NA, not applicable.

provided what we considered the best balance, with fewer background reads than the Illustra V2 kit, yet more reads from pathogens than the Illustra single cell kit. This is also supported by comparing the ratio of known pathogen-to-nonpathogen microbial reads. In four of five samples, the REPLI-g kit provided the highest ratio of pathogen-to-nonpathogen microbial reads. It is also important to note that the species identified from the background DNA are also those that routinely cause a wide range of infections. So while clinical judgment may be able to differentiate identifications such as *Toxoplasma* and *Gloeocapsa* species from true pathogens in certain clinical settings, the presence of genera such as *Pseudomonas*, *Acinetobacter*, or *Streptococcus* would not be so easily differentiated and could lead to overinterpretation or misidentification of causative organisms.

There was a surprising difference in the number of reads originating from known pathogens when comparing the results for samples sequenced by a non-WGA method (NEBNext Ultra II) to those obtained with the various WGA kits. With the exception of sample 996, the non-WGA libraries resulted in far fewer pathogen reads, especially when accounting for the baseline total number of reads obtained for each sample. Additional analysis of the samples by the non-WGA method revealed a large number of small sequencing fragments (average size, 150 to 200 bp) that were human in origin, likely reflecting the degradation of human DNA during pretreatment with the MoLYsis kit reagents and DNA isolation. The difference in relative pathogen counts could therefore reflect either a bias of WGA against these small fragments (21), a bias during library preparation of samples by the non-WGA method against large fragments, or a

combination of both. More aggressive selection of smaller fragments prior to library preparation could potentially lessen this discrepancy.

While there has been debate about the origin of background DNA generated by MDA WGA methods (i.e., amplification of contaminant DNA versus nonspecific extension of random primers), there is evidence for the presence of contaminant DNA in WGA reagents prior to amplification (24). The consistent makeup of the background DNA from the individual kits observed in this study also suggests the presence of specific contaminant DNA in the kits. Because the same DNA preparation was used for each WGA reaction, the contribution of contaminant DNA in steps such as microbial DNA enrichment or DNA extraction is unlikely to account for the differences observed between WGA kits.

The impact of the observed WGA background DNA largely depends on the application for which WGA is used. If one is analyzing only human sequences, then the variable microbial DNA composition has little effect, though any human DNA contamination in WGA kits certainly would. If the goal is to analyze a single cell genome which is aligned to a reference genome, then any background reads would have little impact, as long as the background DNA is not significantly similar to that of the bacterium of interest. However, when one combines a small amount of input DNA with a broad metagenomics analysis approach (a common situation for which metagenomic shotgun sequencing is likely to be used in clinical microbiology), then the presence of contaminant DNA can be greatly magnified, and while the relative amount of contaminating DNA generated can be lessened by increasing the amount of input template DNA, the value of WGA is likely to be high in settings where minimal template DNA is available, so increasing the amount of template DNA is not always possible.

Commercial reagents for the preparation of sequencing libraries continue to improve, and less and less input DNA is required to generate libraries. These methods still require amplification of DNA using PCR-based methods, which raises the potential for contaminating DNA amplification and amplification bias. When comparing the NEBNext Ultra II library preparation method using DNA amplified by a non-WGA method with the NEBNext Ultra library preparation method using DNA amplified by WGA, we continued to observe contaminating DNA reads with a pattern distinct to the NEBNext Ultra II kit. This suggests that simply incorporating this new kit requiring small amounts of DNA will not provide an easy solution to contaminant DNA for metagenomic analysis of low-biomass samples. This is in agreement with the findings of previous studies, in which high levels of contaminating DNA reads were seen when decreasing amounts of template DNA were used for Nextera XT library kit preparation (27).

The range of the number of reads from known pathogens also highlights the challenge of defining cutoffs for the presence or absence of microorganisms based on just the number of reads from an organism. The amount of human DNA in samples can dramatically shift the total number of microbial reads. For example, using the Illustra V2 kit for a sample such as sample 996, in which 96,154 reads were observed for the known pathogen, if one were to apply a cutoff of 90,000 reads as being indicative of a bacterium being present, then for samples such as Ringer's solution or samples undergoing WGA reactions with no template, in which presumably no bacteria (or human DNA) were present, a cutoff of 90,000 would have resulted in numerous contaminants being incorrectly identified as being present. Setting a minimum percentage of all microbial reads could be added; however, this could lead to missed calls in polymicrobial settings. Additional analyses, such as the extent of genome coverage, may be useful to help determine the likelihood that a microorganism is truly present in a sample.

This study does have limitations. Single lots of WGA kits were used for the reported experiments. While this provided consistency within the study, it is possible that there is lot-to-lot variability in contaminant DNA and that this could change the makeup of the background DNA, impacting which kit may be best suited for a method. Our subsequent experience with multiple lots of the Qiagen REPLI-g kits indicates that the findings are not highly different between lots. There are also limitations to the tools

used for the taxonomic assignment of reads. *Toxoplasma* species was frequently identified as a contaminant in the Illustra single cell kit-based analysis. This highlights a limitation of many databases used for assigning a taxonomic assignment of reads. These tools make use of whole-genome sequencing data sets which, in many cases, can include contaminating human DNA mislabeled as microbial DNA (32). Additionally, we performed WGA in the reaction volumes recommended per the manufacturers' instructions. It has been demonstrated that the use of smaller volumes, particularly the very small volumes typically used in sorted single cell applications, result in less contaminant DNA (33). Given that we were investigating WGA as a tool for use with clinical samples, we considered the volumes used to be reflective of potential methods that will be used in the future for these types of samples.

There are many potential sources for contaminant DNA beyond the WGA kit used. There can be background, nonclinically significant DNA in the specimen to begin with, or contaminant DNA can be introduced during specimen collection, prior to arrival in the laboratory. Introduction of DNA through handling in the laboratory is a possibility at every step; mitigation measures, such as the use of personal protective equipment, dedicated and airflow-controlled work areas, and laminar flow hoods and the treatment of surfaces with reagents, such as diluted bleach, to remove environmental DNA, may be helpful (27). It may be that specially constructed laboratories will be needed to ideally perform this type of analysis. Laboratory disposables (e.g., pipettes, tips) are potential sources of contaminant DNA that can be difficult to control. Reagents often not included in commercial kits (e.g., ethanol, Tris-EDTA buffer) must also be considered potential sources of contaminant DNA. Often, these are not ensured to be free of microbial DNA, so measures must be taken to try to remove any DNA present; it is helpful to minimize the number of reagent lots used and, of course, to carefully track which ones are used for each sample, potentially allowing contaminants to be recognized by findings across specimens.

Accounting for possible contaminant DNA from WGA kits or other sources when making clinical interpretations of metagenomic data will be challenging. Identifying common contaminant species in negative controls and excluding them from interpretation should be attempted, but these steps do not always solve the issue, as potentially clinically relevant species may be contaminants. Subtracting reads on the basis of what is detected in negative controls also falls short, given that the relative contribution of contaminant DNA in negative controls may be higher than that in clinical samples where additional DNA is present, so the final number of contaminant reads from negative controls may be higher than that present in patient samples, compromising simple subtraction. Contaminant read numbers also vary between negative controls, making this approach even more problematic.

There are potential methods to account for contaminant DNA during the metagenomics data analysis. We have observed that reads from contaminant DNA often result from the large-scale amplification of small DNA fragments. This may be accounted for by aligning reads to a reference genome to evaluate genome coverage, although this approach has limitations as the number of reads decreases. Identifying cutoffs for what constitutes sufficient coverage needs to be defined but, as discussed above, is challenging. Single nucleotide polymorphism analysis could possibly be employed to differentiate contaminants from new identifications if a particular species is commonly encountered in negative controls. This approach may also prove beneficial if there is concern for cross-contamination between samples being tested. While these and other methods offer promise for discriminating clinically meaningful results from the background findings, much work remains to make this type of testing clinically useful in routine practice; in the end, it is likely that interpretation of results, particularly in gray areas, will require a correlation with other clinical findings.

Overall, our findings highlight not only the impact that WGA as a whole can have on the metagenomic analysis of infections but also that the selection of a specific kit can have a dramatic impact, including not only on the background contaminant DNA observed but also on the relative number of reads observed from known pathogens.

When designing metagenomic experiments incorporating WGA, these factors should be considered as part of the methods selection process.

MATERIALS AND METHODS

Samples. Samples were collected under Mayo Clinic Institutional Review Board protocol 10-005574. Sonicate samples were prepared from resected prosthetic hip and knee components using previously described vortexing/sonication methods (34). Complete microbiological data were available for all samples to allow identification of any bacteria present. Eight samples, including samples from culture-positive PJIs, culture-negative PJIs, and aseptic failures, were selected. *C. glutamicum* ATCC 13032 was used as a positive control. Negative controls included sterile Ringer's solution, which was used for implant sonication and subjected to the same DNA purification procedure as sonicate fluid, as well as WGA reactions without a template.

Sample preparation. All steps were conducted in a laminar flow hood to minimize the risk of DNA contamination. Work areas were thoroughly cleaned with diluted bleach prior to use, and gloves and laboratory coats were used during all steps. Prior to DNA extraction, 1 ml of sonicate fluid underwent microbial DNA enrichment using a MoYsis Basic5 kit (Molzym, Bremen, Germany), which lyses human cells and degrades released DNA, to improve detection of bacterial DNA (26). DNA extraction was performed using a MoBio bacteremia DNA isolation kit. Whole-genome amplification was then carried out using a Qiagen REPLI-g single cell kit (Qiagen, Hilden, Germany), an Illustra V2 Genomiphi kit (GE Healthcare Bio-Sciences, Pittsburgh, PA), or an Illustra single cell Genomiphi kit per the manufacturers' protocols using 1 μ l of input DNA. Amplified DNA was then purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA) per the manufacturer's protocol. Ringer's solution or 10^5 CFU of *C. glutamicum* suspended in Ringer's solution was also purified using the above-described methods. WGA reactions without template DNA added were included as additional controls. For the Illustra V2 and Illustra single cell kits, the amplification steps of the WGA reactions for the no-template and Ringer's solution negative controls were extended to 4 h to obtain sufficient amounts of DNA products.

Sequencing. Paired-end libraries from WGA reactions were prepared by the Mayo Clinic Medical Genome Facility using a NEBNext Ultra DNA library preparation kit (New England BioLabs, Ipswich, MA). Samples were sequenced with an Illumina HiSeq 2500 sequencer in rapid run mode with 2×250 -bp reads. Samples were multiplexed with six samples per lane. Samples that did not undergo WGA underwent library preparation with the NEBNext Ultra II DNA library preparation kit and were sequenced under conditions similar to those described above, except with multiplexing of four samples per lane due to sample number availability.

Data analysis. Illumina adapters were removed using the Trimmomatic trimming tool (35). Human and phiX sequences were then prefiltered using BioBloom tools (36) by aligning reads to a library of human and phiX sequences and computationally subtracting any reads matching these libraries. The Livermore metagenomics analysis tool kit (LMAT) was then used for taxonomic assignment of reads (30), utilizing the kML+Human.v4-14.20.g10.db database (32) and setting a minimum identification score of 1.0, which is considered a high-quality identification. Reads identified as human by LMAT (taxonomy identifier number 9,606) were removed. The Trimmomatic tool was then used to trim low-quality reads (parameters of leading = 3, trailing = 3, MAXINFO = 220:0.1, and minimum length = 70). Trimmed reads were then analyzed using MetaPhlan2 as an additional tool to identify organisms on the basis of the presence of unique marker genes (31).

Beta-diversity analysis and PCoA plot generation. Fastq files depleted of human and phiX sequences by BioBloom tools and LMAT were concatenated, and HUMAN2 (version 0.5.0; Department of Biostatistics, Harvard T. H. Chan School of Public Health [<http://huttenhower.sph.harvard.edu/humann2>]) was used to determine the gene abundances of the organisms present. QIIME was then used to calculate beta diversity and generate principal coordinate analysis (PCoA) plots (37).

Accession number(s). Files containing next-generation sequencing data with human reads removed have been deposited in the NCBI Sequence Read Archive under accession numbers SAMN06604700 through SAMN06604732 and SAMN06678448 through SAMN06678455. They are available under BioProject accession number [PRJNA378504](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA378504).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/JCM.02402-16>.

SUPPLEMENTAL FILE 1, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

The research reported in this publication was supported by the National Institutes of Health under award number R01AR056647. N.C. is supported by award number R01CA179243.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

R. Patel reports grants from BioFire, Check-Points, Curetis, 3M, Merck, Hutchison Biofilm Medical Solutions, Accelerate Diagnostics, Allergan, and The Medicines Com-

pany. R. Patel is a consultant to Curetis, Roche, Qvella, and Diaxonhit. In addition, R. Patel has a patent on a *Bordetella pertussis*/*B. parapertussis* PCR with royalties paid by TIB, a patent on a device/method for sonication with royalties paid by Samsung to Mayo Clinic, and a patent on an antibiofilm substance. R. Patel serves on an Actelion data monitoring board. R. Patel receives travel reimbursement and an editor's stipend from ASM and IDSA and honoraria from the USMLE, Up-to-Date, and the Infectious Diseases Board Review Course.

REFERENCES

- Goldberg B, Sichtig H, Geyer C, Ledebner 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:e01888-15. <https://doi.org/10.1128/mBio.01888-15>.
- Quan PL, Wagner TA, Briese T, Torgerson TR, Hornig M, Tashmukhamedova A, Firth C, Palacios G, Baisre-De-Leon A, Paddock CD, Hutchison SK, Egholm M, Zaki SR, Goldman JE, Ochs HD, Lipkin WI. 2010. Astrovirus encephalitis in boy with X-linked agammaglobulinemia. *Emerg Infect Dis* 16:918–925. <https://doi.org/10.3201/eid1606.091536>.
- 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. <https://doi.org/10.1056/NEJMoa1401268>.
- Hoffmann B, Tappe D, Hoper D, Herden C, Boldt A, Mawrin C, Niederstrasser O, Muller T, Jenckel M, van der Grinten E, Lutter C, Abendroth B, Teifke JP, Cadar D, Schmidt-Chanasit J, Ulrich RG, Beer M. 2015. A variegated squirrel bornavirus associated with fatal human encephalitis. *N Engl J Med* 373:154–162. <https://doi.org/10.1056/NEJMoa1415627>.
- 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. <https://doi.org/10.1093/cid/ciu912>.
- Greninger AL, Messacar K, Dunnebacke T, Naccache SN, Federman S, Bouquet J, Mirsky D, Nomura Y, Yagi S, Glaser C, Vollmer M, Press CA, Kleinschmidt-DeMasters BK, Dominguez SR, Chiu CY. 2015. Clinical metagenomic identification of Balamuthia mandrillaris encephalitis and assembly of the draft genome: the continuing case for reference genome sequencing. *Genome Med* 7:113. <https://doi.org/10.1186/s13073-015-0235-2>.
- Thoendel M, Jeraldo P, Greenwood-Quaintance KE, Chia N, Abdel MP, Steckelberg JM, Osmon DR, Patel R. A possible novel prosthetic joint infection pathogen, *Mycoplasma salivarium*, identified by metagenomic shotgun sequencing. *Clin Infect Dis*, in press.
- Gibson MK, Forsberg KJ, Dantas G. 2015. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J* 9:207–216. <https://doi.org/10.1038/ismej.2014.106>.
- Yang Y, Jiang X, Chai B, Ma L, Li B, Zhang A, Cole JR, Tiedje JM, Zhang T. 2016. ARGs-OAP: online analysis pipeline for antibiotic resistance genes detection from metagenomic data using an integrated structured ARG-database. *Bioinformatics* 32:2346–2351. <https://doi.org/10.1093/bioinformatics/btw136>.
- Scholz M, Ward DV, Pasolli E, Tolio T, Zolfo M, Asnicar F, Truong DT, Tett A, Morrow AL, Segata N. 2016. Strain-level microbial epidemiology and population genomics from shotgun metagenomics. *Nat Methods* 13:435–438. <https://doi.org/10.1038/nmeth.3802>.
- Jeraldo P, Hernandez A, Nielsen HB, Chen X, White BA, Goldenfeld N, Nelson H, Alhquist D, Boardman L, Chia N. 2016. Capturing one of the human gut microbiome's most wanted: reconstructing the genome of a novel butyrate-producing, clostridial scavenger from metagenomic sequence data. *Front Microbiol* 7:783. <https://doi.org/10.3389/fmicb.2016.00783>.
- Yergeau E, Masson L, Elias M, Xiang S, Madey E, Huang H, Brooks B, Beaudette LA. 2016. Comparison of methods to identify pathogens and associated virulence functional genes in biosolids from two different wastewater treatment facilities in Canada. *PLoS One* 11:e0153554. <https://doi.org/10.1371/journal.pone.0153554>.
- Utturkar SM, Cude WN, Robeson MS, Il, Yang ZK, Klingeman DM, Land ML, Allman SL, Lu TS, Brown SD, Schadt CW, Podar M, Doktycz MJ, Pelletier DA. 2016. Enrichment of root endophytic bacteria from Populus deltoides and single-cell genomics analysis. *Appl Environ Microbiol* 82:5698–5708. <https://doi.org/10.1128/AEM.01285-16>.
- Court CM, Ankeny JS, Sho S, Hou S, Li Q, Hsieh C, Song M, Liao X, Rochefort MM, Wainberg Z, Graeber TG, Tseng HR, Tomlinson JS. 2016. Reality of single circulating tumor cell sequencing for molecular diagnostics in pancreatic cancer. *J Mol Diagn* 18:688–696. <https://doi.org/10.1016/j.jmoldx.2016.03.006>.
- Capal P, Blavet N, Vrana J, Kubalaková M, Dolezel J. 2015. Multiple displacement amplification of the DNA from single flow-sorted plant chromosome. *Plant J* 84:838–844. <https://doi.org/10.1111/tpj.13035>.
- Prado-Alvarez M, Couraleau Y, Chollet B, Tourbiez D, Arzul I. 2015. Whole-genome amplification: a useful approach to characterize new genes in unculturable protozoan parasites such as Bonamia exitiosa. *Parasitology* 142:1523–1534. <https://doi.org/10.1017/S0031182015000967>.
- Rinke C, Lee J, Nath N, Goudeau D, Thompson B, Poulton N, Dmitrieff E, Malmstrom R, Stepanauskas R, Woyke T. 2014. Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat Protoc* 9:1038–1048. <https://doi.org/10.1038/nprot.2014.067>.
- Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M, Lasken RS. 2002. Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A* 99:5261–5266. <https://doi.org/10.1073/pnas.082089499>.
- de Bourcy CF, De Vlaminck I, Kanbar JN, Wang J, Gawad C, Quake SR. 2014. A quantitative comparison of single-cell whole genome amplification methods. *PLoS One* 9:e105585. <https://doi.org/10.1371/journal.pone.0105585>.
- Probst AJ, Weinmaier T, DeSantis TZ, Santo Domingo JW, Ashbolt N. 2015. New perspectives on microbial community distortion after whole-genome amplification. *PLoS One* 10:e0124158. <https://doi.org/10.1371/journal.pone.0124158>.
- Direito SO, Zaura E, Little M, Ehrenfreund P, Roling WF. 2014. Systematic evaluation of bias in microbial community profiles induced by whole genome amplification. *Environ Microbiol* 16:643–657. <https://doi.org/10.1111/1462-2920.12365>.
- Yilmaz S, Allgaier M, Hugenholtz P. 2010. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nat Methods* 7:943–944. <https://doi.org/10.1038/nmeth.1210-943>.
- Raghunathan A, Ferguson HR, Jr, Bornarth CJ, Song W, Driscoll M, Lasken RS. 2005. Genomic DNA amplification from a single bacterium. *Appl Environ Microbiol* 71:3342–3347. <https://doi.org/10.1128/AEM.71.6.3342-3347.2005>.
- Blainey PC, Quake SR. 2011. Digital MDA for enumeration of total nucleic acid contamination. *Nucleic Acids Res* 39:e19. <https://doi.org/10.1093/nar/gkq1074>.
- Tande AJ, Patel R. 2014. Prosthetic joint infection. *Clin Microbiol Rev* 27:302–345. <https://doi.org/10.1128/CMR.00111-13>.
- Thoendel M, Jeraldo PR, Greenwood-Quaintance KE, Yao JZ, Chia N, Hanssen AD, Abdel MP, Patel R. 2016. Comparison of microbial DNA enrichment tools for metagenomic whole genome sequencing. *J Microbiol Methods* 127:141–145. <https://doi.org/10.1016/j.mimet.2016.05.022>.
- 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. <https://doi.org/10.1186/s12915-014-0087-z>.
- Bukowska-Osko I, Perlejewski K, Nakamura S, Motooka D, Stokowy T, Kosinska J, Popiel M, Ploski R, Horban A, Lipowski D, Caraballo Cortes K, Pawelczyk A, Demkow U, Stepien A, Radkowski M, Laskus T. 2017. Sensitivity of next-generation sequencing metagenomic analysis for

- detection of RNA and DNA viruses in cerebrospinal fluid: the confounding effect of background contamination. *Adv Exp Med Biol* 944:53–62. https://doi.org/10.1007/5584_2016_138.
29. Perlejewski K, Bukowska-Osko I, Nakamura S, Motooka D, Stokowy T, Ploski R, Rydzanicz M, Zakrzewska-Pniewska B, Podlecka-Pietowska A, Nojszewska M, Gogol A, Caraballo Cortes K, Demkow U, Stepien A, Laskus T, Radkowski M. 2016. Metagenomic analysis of cerebrospinal fluid from patients with multiple sclerosis. *Adv Exp Med Biol* 935:89–98. https://doi.org/10.1007/5584_2016_25.
30. Ames SK, Hysom DA, Gardner SN, Lloyd GS, Gokhale MB, Allen JE. 2013. Scalable metagenomic taxonomy classification using a reference genome database. *Bioinformatics* 29:2253–2260. <https://doi.org/10.1093/bioinformatics/btt389>.
31. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N. 2015. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat Methods* 12:902–903. <https://doi.org/10.1038/nmeth.3589>.
32. Ames SK, Gardner SN, Marti JM, Slezak TR, Gokhale MB, Allen JE. 2015. Using populations of human and microbial genomes for organism detection in metagenomes. *Genome Res* 25:1056–1067. <https://doi.org/10.1101/gr.184879.114>.
33. Marcy Y, Ishoey T, Lasken RS, Stockwell TB, Walenz BP, Halpern AL, Beeson KY, Goldberg SM, Quake SR. 2007. Nanoliter reactors improve multiple displacement amplification of genomes from single cells. *PLoS Genet* 3:1702–1708.
34. Trampuz A, Piper KE, Jacobson MJ, Hanssen AD, Unni KK, Osmon DR, Mandrekar JN, Cockerill FR, Steckelberg JM, Greenleaf JF, Patel R. 2007. Sonication of removed hip and knee prostheses for diagnosis of infection. *N Engl J Med* 357:654–663. <https://doi.org/10.1056/NEJMoa061588>.
35. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
36. Chu J, Sadeghi S, Raymond A, Jackman SD, Nip KM, Mar R, Mohamadi H, Butterfield YS, Robertson AG, Birol I. 2014. BioBloom tools: fast, accurate and memory-efficient host species sequence screening using bloom filters. *Bioinformatics* 30:3402–3404. <https://doi.org/10.1093/bioinformatics/btu558>.
37. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336. <https://doi.org/10.1038/nmeth.f.303>.